New Ideas in Psychology 31 (2013) 43–53

Contents lists available at ScienceDirect

New Ideas in Psychology

journal homepage: www.elsevier.com/locate/newideapsych

Deconstructing the construct: A network perspective on psychological phenomena

Verena D. Schmittmann, Angélique O.J. Cramer, Lourens J. Waldorp, Sacha Epskamp, Rogier A. Kievit, Denny Borsboom*

Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands

Keywords:
Psychological construct
Dynamical systems
Latent variable model
Network
Formative model
Reflective model

A B S T R A C T

In psychological measurement, two interpretations of measurement systems have been developed: the reflective interpretation, in which the measured attribute is conceptualized as the common cause of the observables, and the formative interpretation, in which the measured attribute is seen as the common effect of the observables. We advocate a third interpretation, in which attributes are conceptualized as systems of causally coupled (observable) variables. In such a view, a construct like 'depression' is not seen as a latent variable that underlies symptoms like 'lack of sleep' or 'fatigue', and neither as a composite constructed out of these symptoms, but as a system of causal relations between the symptoms themselves (e.g., lack of sleep → fatigue, etc.). We discuss methodological strategies to investigate such systems as well as theoretical consequences that bear on the question in which sense such a construct could be interpreted as real.

Current theorizing and research in psychology is dominated by two conceptualizations of the relationship between psychological attributes (e.g., 'neuroticism') and observable variables (e.g., 'worries about things going wrong'; Edwards & Bagozzi, 2000). In the first of these conceptualizations - the *reflective* model - the attribute is seen as the common cause of observed scores: neuroticism causes worrying about things going wrong. In the second conceptualization - the *formative* model - observed scores define or determine the attribute. The classic example of such a model involves socio-economic status (SES), which is viewed as the joint effect of variables like education, job, salary and neighborhood.

In the present paper, we argue that the dichotomy of reflective/formative models does not exhaust the possibilities that can be used to connect psychological attributes and observable variables. We advocate an alternative conceptualization, in which psychological attributes are conceptualized as *networks* of directly related observables. We discuss the possibilities that this addition to the psychometric arsenal offers, the inferential techniques that it allows for, and the consequences it has for the ontology of psychopathological constructs and the epistemic status of validation strategies.

The structure of this paper is as follows. First, we discuss the ideas that underlie reflective and formative models. Second, we highlight important problems that the models face. Third, we discuss the network approach. Fourth, we touch on the ramifications that this approach has in the context of validity theory.

1. Reflective and formative models

1.1. Reflective models

In reflective models, observed indicators (e.g., item or subtest scores) are modeled as a function of a common latent variable (i.e., unobserved) and item-specific error

* Corresponding author.
E-mail address: d.borsboom@uva.nl (D. Borsboom).

0732-118X/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.newideapsych.2011.02.007

variance. Reflective models are commonly presented as 'measurement models' in modern test theory (Mellenbergh, 1994). Examples are the IRT models of Rasch (1960), Birnbaum (1968) and Samejima (1969), common factor models (Jöreskog, 1971; Lawley & Maxwell, 1963), latent class models (Lazarsfeld, 1959), and latent profile models (Bartholomew, 1987; McLachlan & Peel, 2000). In these models, a latent variable is introduced to account for the covariance between indicators. In a nontrivial sense, it explains this covariance; in most models it is assumed that conditioning on the latent variable makes the covariance vanish (this is an implication of local independence). The latent variable then functions analogously to an unobserved common cause (Pearl, 2000).

This model matches the way in which many theorists in psychology think about the relation between psychological attributes and observations. For instance, in clinical psychology, the conceptual idea of reflective models is often used as a blueprint for a realistic picture of a mental disorder and its symptoms; that is, a mental disorder is thought to be a reflective construct that causes its observable symptoms (e.g., depression causes fatigue and thoughts of suicide). Likewise, personality variables like neuroticism may be considered as the common cause of observable neurotic behaviors, such as feeling jittery and worrying about things going wrong. A set of indicators that measure the observable consequences of such attributes can then be used to make inferences about individual differences in the underlying attributes: Alice has a higher total score on a neuroticism questionnaire than John *because* Alice is more neurotic than John. Fig. 1 presents a reflective model for the items of the Big Five Neuroticism scale of the Dutch NEO-PI-R (Hoekstra, Ormel, & De Fruyt, 2003), obtained from 500 first-year Psychology students (Dolan, Oort, Stoel, & Wicherts, 2009).

In reflective models, indicators are regarded as *exchangeable* save for measurement parameters like reliability (Bollen, 1989). That is, although the indicators in Fig. 1 may differ in their factor loadings (as indicated by the thickness of arrows between N and items) and residual variances, the relation they bear to neuroticism is qualitatively the same. Therefore, it does not make a qualitative difference which neuroticism items one uses.

Second, the observed correlations between the indicators are *spurious* in the reflective model (as indicated by the absence of edges between individual items). That is, observed indicators *should* correlate; but they only do so *because* they share a cause, namely neuroticism. Such thinking makes perfect sense in the case of the reflective construct 'temperature' that is measured with three different thermometers: any correlation between the thermometers is caused by the fact that they measure the same thing, namely temperature. There is no direct causal relation between the thermometers, in that the functioning of thermometer A does not directly cause the temperature reading on thermometer B. Thus, it is feasible to regard the correlations between the thermometers as essentially spurious, and this is indeed a sensible assumption of models that aspire to capture the idea that the relation between indicators and a particular construct is one of measurement.

1.2. Formative models

In formative models, possibly latent composite variables are modeled as a function of indicators. Without residual variance on the composite, models like principal components analysis and clustering techniques serve to construct an optimal composite out of observed indicators. However, one can turn the composite into a latent composite if one introduces residual variance on it. This happens, for instance, if model parameters are chosen in a way that optimizes a criterion variable. Conditional independence of observed indicators given the composite variable is not assumed. Rather, the independence relation is reversed: in formative models, conditioning on the composite variable induces covariance among the observables even if they were unconditionally independent; hence the composite variable functions analogously to a common effect (Pearl, 2000). Fig. 2 presents a formative model for the same dataset as above.

Formative models differ from reflective models in many aspects. Indicators are *not* exchangeable because indicators are hypothesized to capture different aspects of the construct. In the neuroticism example, this implies that 'feeling jittery' and 'worrying about things going wrong' represent different aspects of the construct "neuroticism." As such, removing an indicator potentially alters the formative construct (Bollen & Lennox, 1991; Edwards & Bagozzi, 2000). Also, contrary to reflective models, there is no a priori assumption about whether indicators of a formative construct should correlate positively, negatively or not at all.

2. Problems with the reflective and formative conceptualizations

The status and nature of reflective and formative measurement models have been the source of various discussions (Bagozzi, 2007; Bollen, 2007; Howell, Breivik, & Wilcox, 2007a; Howell, Breivik, & Wilcox, 2007b; see also a special issue of the Journal of Business Research, vol. 16, issue 12, 2008). These have centered on desirable properties of indicators in formative and reflective models (Bollen, 1984; Jarvis, Mackenzie, & Podsakoff, 2003; Wilcox, Howell, & Breivik, 2008), the status of the error term in formative models (Coltman, Devinney, Midgley, & Venaik, 2008; Diamantopoulos, 2006; Edwards, 2011), model selection (e.g. Baxter, 2009; Diamantopoulos & Siguaw, 2006; Jarvis et al., 2003), referential (in)stability (e.g. Burt, 1976; Franke, Preacher, & Rigdon, 2008), and (causal) interpretations of the relation between indicators and latent variables (e.g. Blalock, 1964; Bollen & Lennox, 1991; Borsboom, Mellenbergh, & Van Heerden, 2003; Diamantopoulos, Riefler, & Roth, 2008; Edwards, 2011; Edwards & Bagozzi, 2000).

Such debates have often focused on the question whether there are general reasons to favor one or the other model. Both ends of the spectrum have been defended in this respect, from questioning whether formative models are *ever* appropriate on the one hand (e.g. Edwards, 2011; Wilcox et al., 2008) to arguing that reflective models are adopted too readily, and that formative models may usually

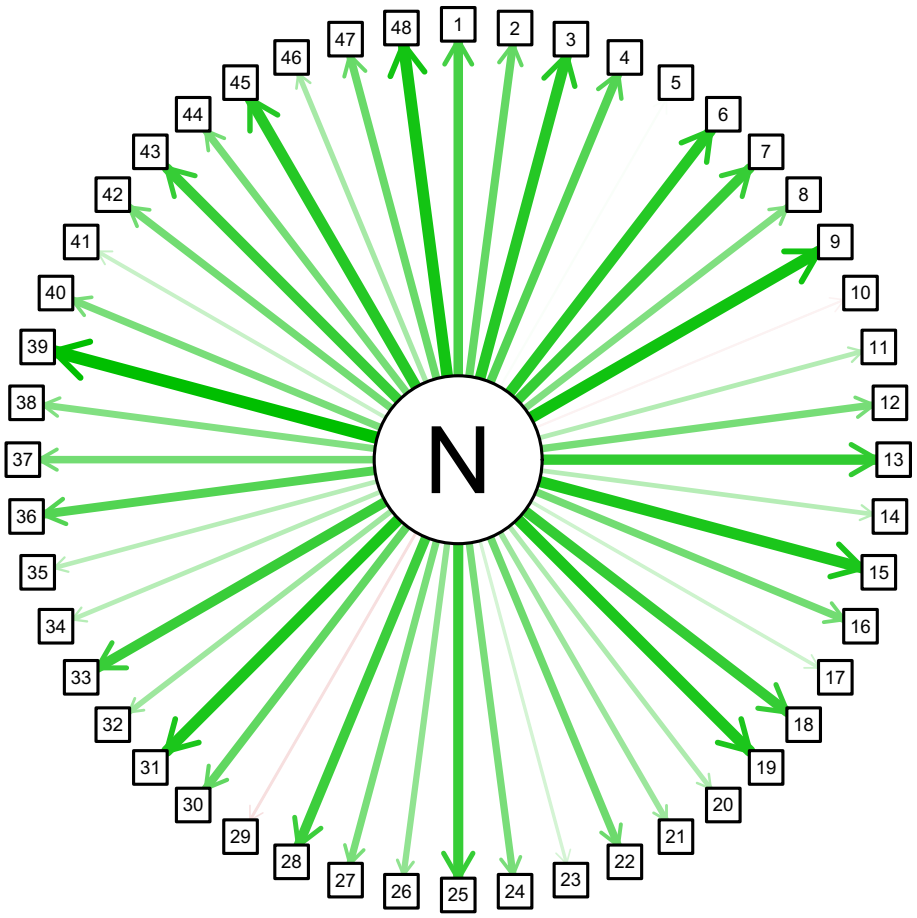

Fig. 1. A reflective model of Neuroticism items of the NEO-PI-R questionnaire. One underlying Neuroticism factor, depicted as a circle, determines the variation in the items, depicted as rectangles. The thicker and darker green an arrow from the factor to an item, the higher the factor loading. Residual variances are not represented. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

be more appropriate (e.g. Coltman et al., 2008; Diamantopoulos et al., 2008; Jarvis et al., 2003). It seems clear that these debates are far from settled. However, they largely rage within the confines of a dichotomous choice between formative and reflective models. An intriguing possibility is that more fundamental issues plague both models and that these issues may be the source of at least some of the debates discussed above.

For instance, although the resemblance of reflective and formative models to common cause and effect models is indeed striking, in many instances of psychological testing the causal relations suggested by these conceptualizations are extremely problematic. Three particular problems concern the role of time, the inability to articulate causal relations between construct and observables in terms of processes, and the subordinate treatment of relations between observables.

2.1. The role of time

In most conceptions of causality, causes are required to precede their effects in time. However, in psychometric models like the reflective and formative models, time is generally not explicitly represented. That is, the dynamics of the system are not explicated. It is therefore unclear whether the latent variables relate to the observables in whatever dynamical process generated the observations; in fact, it is unclear whether the latent variables in question would figure in a dynamic account at all (Borsboom et al., 2003; Molenaar, 2004). This puts the causal interpretation of latent variable models, as for instance fitted to data gathered at a single time point, in a difficult position. For instance, in a reflective model, are we supposed to consider the latent differences to exist 'before' the observed differences? Or, in a formative model with a latent variable, is the latent variable to be considered a 'consequence' of the observables?

Edwards and Bagozzi (2000) suggested that the causal order between latent and observed variables is to be fixed through a thought experiment. In such a thought experiment, the researcher considers whether it is more plausible that, say, SES causes a raise in salary, or that a raise in salary causes a higher SES. In this example, intuition tends to the latter possibility; therefore SES should be modeled as a formative construct. Thought experiments and intuition, however, are dubious guides in determining causality; and even in the simple SES case it could be questioned whether

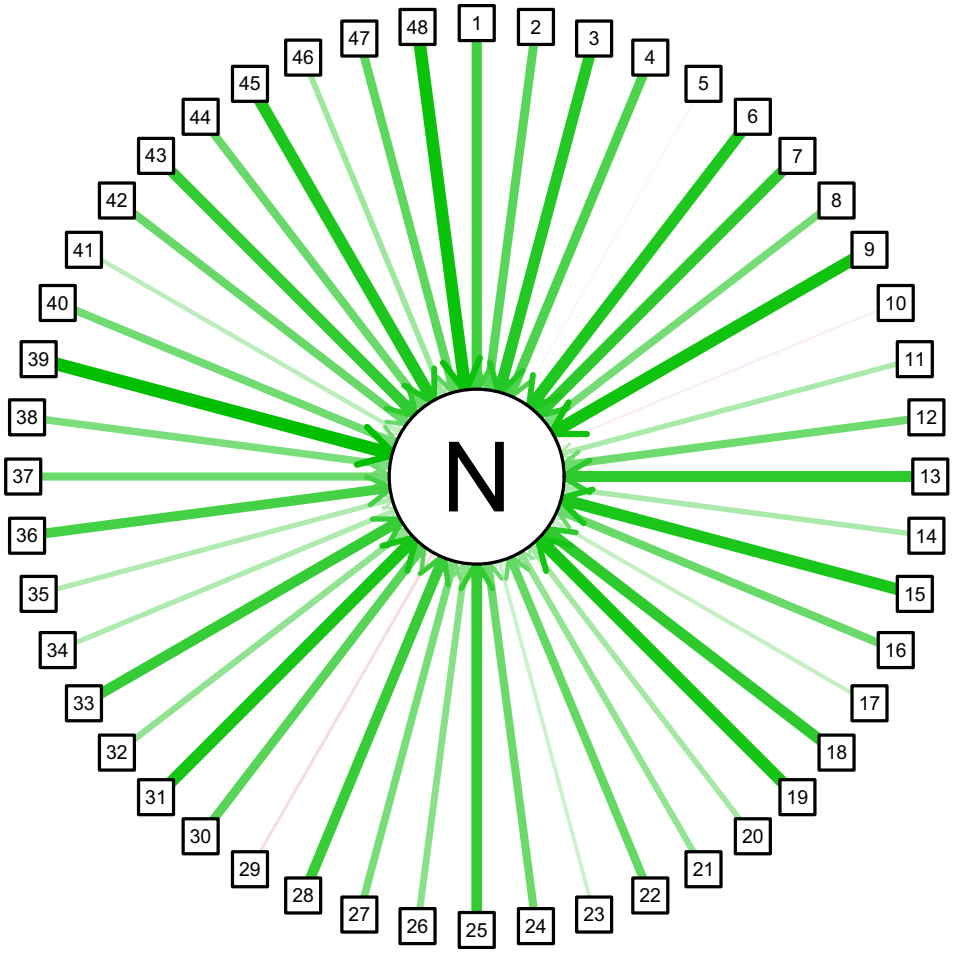

Fig. 2. A formative model of Neuroticism items of the NEO-PI-R questionnaire. Arrows point from the items (rectangles) to the composite variable Neuroticism (N). The thicker and darker green an arrow, the higher the contribution of the item to the composite score. Correlations between items are not represented. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

they are adequate. From a temporal perspective, it is easy to imagine that when John gets a raise that increases his salary, he is able to buy a house in a better neighborhood; this may increase his social status, as a result of which he gets invited to fancy parties, where he becomes friendly with the CEO of a big firm, who offers him a better job; the outcome of which is that he gets a better salary, etc (i.e., a cycle). Given the plausibility of such cyclic developmental trajectories it appears naive to consider the relation between indicators like salary and theoretical variables like SES to be one way traffic.

2.2. Inability to articulate processes

The identification of causal relations is arguably an essential ingredient of the scientific enterprise. Typically, after a causal relation is discovered, it is broken down into constituent processes to illustrate the precise mechanism(s) that realize(s) that relation. For instance, after the general causal relation between smoking and lung cancer was discovered through the standard routes of scientific research researchers endeavored to find out what processes made the causal relation work. That is, they studied processes that lead elements of the causal factor (constituents of tobacco smoke) to trigger mediational processes (tar build up in lung cells) that result in the effect (lung cancer).

Such progress, or even the ambition to realize it, is mostly lacking in psychological measurement. There is rarely a progressive research program that identifies *how*, say, neuroticism causes someone to worry about things going wrong or that identifies the mechanisms that embody the effect of general intelligence on IQ-scores. In fact, it is in many cases quite hard to imagine how such effects could be realized at all. A plausible cause of this problem is that most constructs in psychology are not empirically identifiable apart from the measurement system under validation; no one has been able to identify 'general intelligence' in the brain, for example. This hampers causal research; one may imagine how hard it would be to investigate the effect of smoking on lung cancer if the only measure of smoking were the observation of lung cancer itself (i.e., when smoking would be structurally 'latent').

2.3. Relations between observables

A third important issue in both reflective and formative models is the neglect or subordinate treatment of causal relations *between the observed indicators themselves*. The reflective model relies on the assumption that no direct causal relations exist between observables, and in the formative model relations between observables that are not accounted for by the latent variables are typically treated as a nuisance. However, causal relations between observables are likely to exist in many psychological constructs. Moreover, such causal relations between observables may be the reason why a phenomenon is perceived or interpreted as an entity. Consider again SES. This concept is commonly operationalized as summary statistic over a group of variables that clearly do not measure the same attribute. However, the system is so coherent that researchers discuss topics such as "the relation between SES and intelligence" as if SES in fact did denote a single measured attribute. Why is this? In our view, the perception of SES as a single theoretical entity may arise precisely because its constituents are causally interrelated (education may influence job choice, which may constrain income, which in turn constrain the neighborhood one chooses to live in).

Possibly, these three problems (among others) are at least partly to blame for the intensity and breadth of the debates centering on the use and interpretation of formative versus reflective models. For this reason, a different conceptualization of the relationship between indicators may be appropriate.

3. The network perspective: constructs as dynamical systems

We propose that the variables that are typically taken to be *indicators* of latent variables should be taken to be *autonomous causal entities* in a network of dynamical systems. Instead of positing a latent variable, one assumes a network of directly related causal entities as a result of which one avoids the three problems discussed above.

First, consider criteria for a major depressive episode (MDE; American Psychiatric Association, 1994). These criteria involve symptoms like “lack of sleep,” “fatigue,” and “concentration problems.” In empirical research, scores on these criteria are usually added to form a total score which then functions as a measure of depression. This practice ignores the likely presence of direct relations between symptoms (e.g., lack of sleep → fatigue → concentration problems). Similarly, in personality psychology, one finds items that relate to the ability to get organized, to the tendency of finishing once initiated projects, and to the tendency to adopt a clear set of goals, which are taken to reflect the “conscientiousness” factor of the Big Five model (Hoekstra, Ormel, & De Fruyt, 2003); in a dynamic scheme, however, it appears to be reasonable that, say, having a clear set of goals is an important determinant of getting organized, and that getting organized facilitates finishing projects. In all these cases, the ‘indicators’ function autonomously in the system, rather than being passive indicators of a common construct. These elements are connected causally. We argue that such causal relations *can* be articulated in terms of processes. For some relations, these processes are already known, for example homeostatic processes that are involved in mediating the relation between “lack of sleep” and “fatigue” (both symptoms of depression: Achermann, 2004; Finelli, Baumann, Borbély, & Achermann, 2000).

Fig. 3 gives a flavor of what such a *network* of autonomous causal entities may look like. It represents Neuroticism items as nodes, and the empirical correlations between them as edges. After constructing such a network, one can use techniques from network analysis to visualize the system (Boccaletti, Latora, Moreno, Chavez & Hwang, 2006). For Fig. 3, we used an algorithm for the placement of the nodes, which causes strongly correlated sets of items to cluster together (Fruchterman & Reingold, 1991). For example, item 6 (content relates to feelings of helplessness and the wish for help), is located in the center of the figure, because that item is strongly correlated with other Neuroticism items.

From a network perspective, a construct is seen as a network of variables. These variables are coupled in the sense that they have dependent developmental pathways, because a change in one variable causes a change in another. Studying the construct means studying the network; and such investigation would naturally focus on a) network structure and b) network dynamics. The relation between observables and the construct should not be interpreted as one of measurement, but as one of mereology: the observables do not measure the construct, but are part of it. Therefore, studying the relation between observables and the construct means studying the function of the observables in the network (e.g., which observables are dominant in a network in terms of the strength of relations with other observables?). In the following sections, we will outline a few key concepts of a general framework to investigate such networks of variables, and discuss related methodological procedures.

3.1. Dynamical systems

A general framework to formalize and study the behavior of a network of interconnected variables over time is dynamical systems theory. In psychology, it has for instance been applied to cognitive processes (Van Gelder, 1998), to the construct of intelligence (Van der Maas et al., 2006), and in the area of developmental psychology (Van der Maas & Molenaar, 1992). Put briefly, a dynamical system changes its state (which is represented by a set of interrelated variables) according to equations that describe how the previous state determines the present state, i.e., how the variables influence each other. Given an initial state, the system will move through a trajectory of states over time.

Particularly relevant are *attractor* states of the system. If the system is close to an attractor state, it will converge to it, and remain in there in equilibrium. For example, a depression network may have two attractor states: a disordered, depressed state and a healthy state. A sufficiently large perturbation to the system, such as stressful life events, may propel the person from the healthy state towards the depressed state. In dynamical systems,

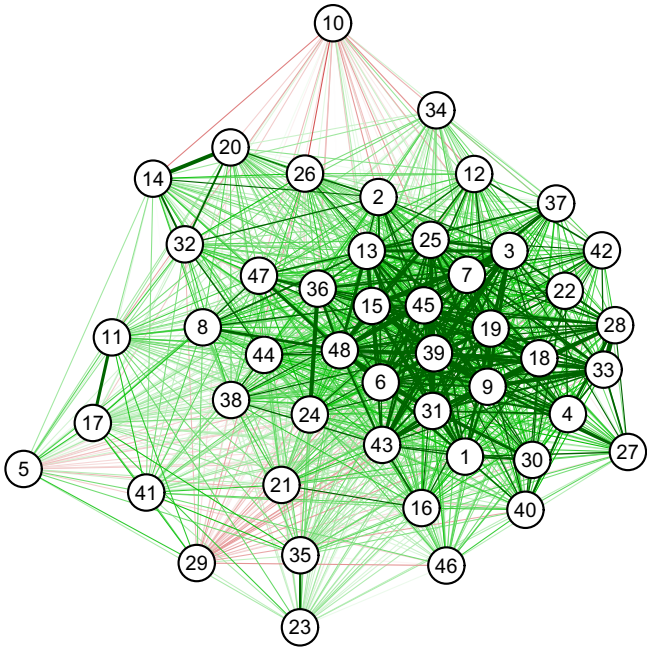

Fig. 3. A network of Neuroticism items of the NEO-PI-R questionnaire. Nodes represent items; edges the empirical correlation between items. Numbers in nodes refer to the order of appearance in the questionnaire. A stronger correlation (positive green; negative red) results in a thicker and darker edge. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.) Graph generated with R-package qgraph, Available from Graph generated with R-package qgraph (Epskamp, Cramer, Waldorp, Schmittmann, Borsboom, 2011)

parameters in the state-transition function determine the number and the type of the equilibrium points. Therefore, if we allow these parameters to change, the system may show qualitative changes in its structure (e.g., a new attractor emerges). Thus the line between depression and health may be continuous for some people, while it is categorical for others.

It would be extremely helpful if we possessed a theory on state-transition-functions that govern psychological systems like those of depression, but these are generally unknown. Thus, we currently lack rules that determine system behavior: who will get depressed and who will not, and in what circumstances? Therefore, in order to study networks of variables, simplifying assumptions will often be made, and simulation methods can be helpful in investigating the plausibility of network properties. For instance, Van der Maas et al. (2006) showed that a mutualistic model of intelligence could produce the positive manifold of intelligence tests; and Borsboom, Cramer, Schmittmann, Epskamp & Waldorp (submitted for publication) were able to show that a network of depression and anxiety symptoms could plausibly reproduce comorbidity statistics regarding these problems.

3.2. Causal inference

A problem with formal theories of dynamical systems is that almost all of the known mathematical results concern deterministic systems. In psychology, we typically deal with probabilistic systems and data characterized by high levels of noise. The difficulty is then to derive, from a statistical pattern, that changes in A are structurally related to changes in B. One way to arrive at a viable method for inferring such relationships between variables is to adopt the assumption of linearity and normality. This assumption gives one access to well-developed causal inference methods (Pearl, 2000; Spirtes, Glymour, & Scheines, 2000). The construction of causal systems through such inference methods is a statistical route that may be used to get a grip on the architecture of networks.

These methods typically work through the detection of conditional independence relations. For instance, consider the graph in Fig. 4(a), which is held to be a representation of the causal relation in the population. Because in the population there is a connection from A to B through C, it is likely that A and B are correlated in the sample. However, A and B are not directly connected, and the explanation for the observed correlation is that C is in between A and B, or put differently, that C separates A and B. So, an easy test to see whether A and B are directly causally connected is to test for a correlation between A and B with C taken out, that is, a partial correlation between A and B. If there is no partial correlation between A and B, then this is taken as evidence that there is no direct causal connection between A and B. Under the assumptions of normality and acyclicity, this result implies that A and B are independent conditional on C.

Knowing this, however, we still do not know the causal relations between A and C and between C and B. In this simple case with three variables, one can distinguish only the graphs of Fig. 4(a)–(c) versus Fig. 4(d). Conditioning on C renders A and B independent in Fig. 4(a)–(c), but yields dependence between A and B in Fig. 4(d). In the latter case,

C is called a collider. A and B become correlated when C is conditioned on, because knowing that, for instance, A is not the cause of C increases the likelihood of B being the cause of C. For instance, suppose that possible causes of a burglar alarm (C) are a dog (A) or a burglar (B); then, given the alarm, learning the presence of a dog decreases the likelihood of a burglar being present (Neapolitan, 2004). To distinguish between cases 4(a)-4(c), one needs additional variables that either cause A or C (creating a collider at A or C) to infer the causal direction between A and C. Generally, variables must be included that could create colliders on the set of variables of interest.

As an illustrative example of inferring causal relations as described above, four participants reported five constituents of depression (tiredness, concentration difficulties, self-content, sad mood, and pleasure in the current activity) on a continuous scale hourly in the daytime on five consecutive days.

Fig. 5 shows the development of those constituents in time for one participant. To these data, we fitted seven confirmatory models, in which we formalized different assumptions about the causal relations between the variables. The best fitting model (as judged by the AIC) was the one shown in Fig. 6.[1] In general, the relations appear to conform to common sense (e.g. if you are tired, you will experience concentration difficulties). The results are illustrative only but can serve to demonstrate how to investigate developmental trajectories of psychological systems.

3.3. Network analysis

Once the network structure has been inferred in one of the aforementioned ways, the network may be subject to further analysis. Many network structure analysis methods are implemented in free software such as the R-package iGraph (Csárdi & Nepusz, 2006). Such methods allow one for instance to examine whether a network has small-world properties (e.g., high clustering of items within one Big Five factor combined with relatively few separating nodes between different Big Five factors; Watts & Strogatz, 1998). The resilience of networks to the removal of their nodes could be of particular importance in psychopathology, where the "removal" of a node might correspond to the administration of medication that directly remedies a symptom. In addition, one can analyze properties of individual nodes, such as their centrality, that is, how strongly a particular node is connected to all the other nodes in a network. Studying network and node properties may help to find meaningful individual differences with respect to the construct.

[1] Besides the best fitting model, the other models, in ascending AIC value, were: 1) tired → concentration AND self-content → sad mood → pleasure (AIC: -1004885); 2) tired → concentration → self-content → sad mood AND pleasure → self-content (AIC: -1004813); 3) tired → concentration → self-content → sad mood AND self-content → pleasure (AIC: -1004813); 4) tired → concentration → self-content → pleasure → self-content (AIC: -1003106); tired → concentration → self-content ← sad mood AND self-content ← pleasure (AIC: -997422).

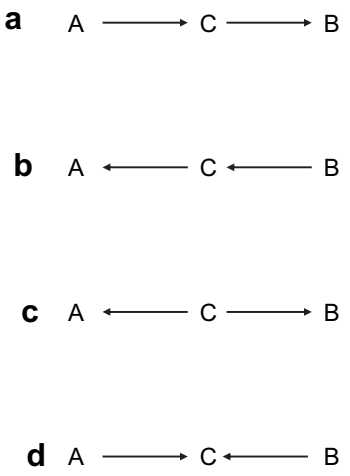

Fig. 4. Four possible relations between three fictitious random variables: A, B, and C. (a): A causes C and C causes B. C is the middle node in a chain and therefore, A and B are independent given C. (b): B causes C and C causes A. C is the middle node in a chain and therefore, A and B are independent given C. (c): C causes both A and B. C functions as the common cause of A and B and therefore, A and B are independent given C. (d): A and B both cause C. C is the middle node of a collider and therefore, A and B are dependent given C.

4. Constructs and their interrelations

The ontological status of psychological constructs as well as the epistemic question of how to measure them has been the topic of considerable controversy in psychology. Borsboom, Mellenbergh, and Van Heerden (2004) have argued that, in order to be plausible candidates for measurement, constructs should in fact refer to structures in reality; structures that play a causal role in determining individual differences in test scores. Maraun and Peters (2005) have suggested that the entire idea of constructs being "unobservable constituents" of "natural reality" is intrinsically misguided. Cronbach and Meehl (1955) famously espoused an agnostic position with respect to this question, arguing that in some cases constructs would refer to causally active latent variables, and in some cases would merely be 'inductive summaries', much like the composites typically formed in formative modeling.

In the network view, a construct label (e.g., the word "depression") does not refer to a latent variable or inductive summary, but to a system. Since there is no latent variable that requires causal relevance, no difficult questions concerning its reality arise. Naturally, the components of the system have to be capable of causal action, but this is typically not much of problem (e.g., returning to the depression example, the reality and causal relevance of sleep loss can hardly stand in doubt). Although psychological constructs are the source of considerable conceptual headaches, individual indicator variables (items, symptoms, response times, etc.) often are tractable and associated with precisely the kinds of progressive scientific research that applied psychometrics typically lacks.

In our view, the referential connection between construct labels and systems may therefore be a comparatively simple affair. A short outline is as follows. Scientists use terms like

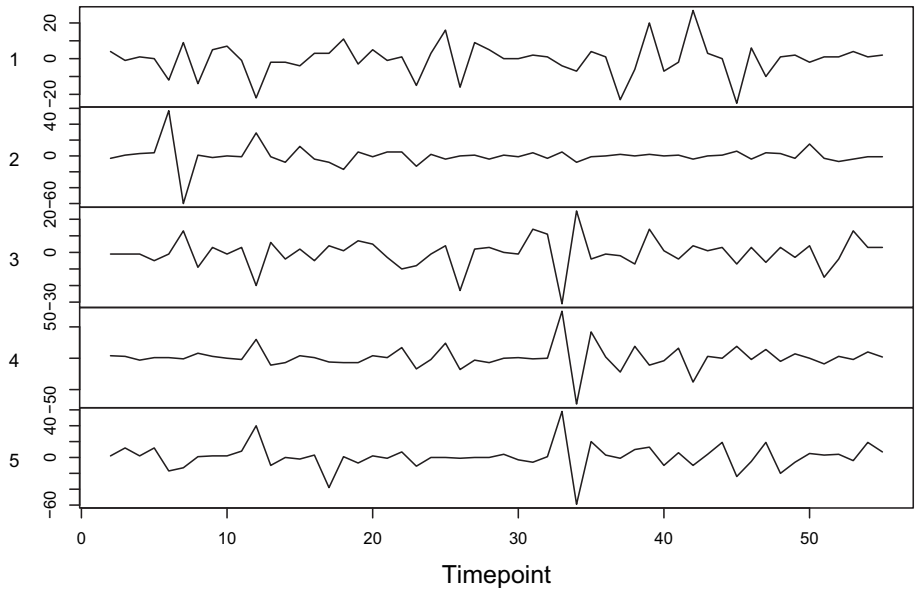

Fig. 5. The developmental trajectory of five constituents of depression in one participant. The *x*-axis represents discrete time points while the *y*-axis is divided in five parts, from top to bottom: 1) pleasure in current activity, 2) sad mood, 3) self-content, 4) concentration difficulties and 5) tiredness. For each part, the *y*-axis represents the continuous scale on which the participants rated the constituents.

"depression" to indicate a system that can be identified through its constituents (e.g., symptoms) and systematic connections between them: an *architecture*. The term "depressed", in contrast, indicates a certain *state* of the person that is analyzable in terms of the network's dynamics, characterized for instance by a attractor state in which a significant number of symptoms is present. This state may either be a gradual property of the system on which we can place a measure, or a discretely identifiable state that we can name. Which of these situations obtains depends on the network architecture and the resulting dynamic properties, with the interesting corollary that, say, levels of depression may be measurable continua in some people but discrete states in others. Networks are likely to differ over people ("traits") and over time ("states") and a considerable psychometric adventure may be entered by figuring out exactly how to determine these from observed data. Conceptually, however, little more is needed to furnish the connection of a construct label to a network.

Importantly, even though borders between networks are likely to be fuzzy, this does not make the systems themselves arbitrary, in the sense that collections of formative measures are arbitrary; systems to which constructs like depression refer can have definite characterizations and are eligible for scientific inquiry. However such characterizations are inherently complex, and scientific theorizing ideally respects that complexity. In this sense, our position is closely related to that of McGrath (2005) who suggests that complexity is an intrinsic property of many psychological constructs, and to that of Kendler, Zachar, and Craver (2010) who argue the very similar position that psychopathological constructs are best construed as mechanistic property clusters.

4.1. Validity

If the question of validity is constructed as whether a set of items "really measures" a given attribute, the answer to that question requires an account of item response processes in which that attribute plays a causal role (Borsboom, Cramer, Kievit, Zand-Scholten, & Franic, 2009; Borsboom et al., 2004). Such an account has not been forthcoming in most areas where the test validity is at issue. This is unsurprising from a network perspective. The essence of a network construct is not a common cause; rather, it resides in the relations between its constituents. These relations lead to a clustering of symptoms picked up both by formal methods to detect clustering (e.g. factor analysis) and by people (e.g. psychiatrists constructing the DSM). However, if a construct like depression is a network, searching for the common cause of its symptoms is like searching for actors inside one's television set. For this reason, the question whether symptoms "really measure" depression, understood causally, is probably moot, and causal processes that

Fig. 6. The best fitting confirmative time series model of the following five constituents of depression: tiredness; concentration difficulties (*concentration*); self-content; sad mood; pleasure in current activity (*activity*).

connect depression to item responses on a questionnaire will not be found because they do not exist.

4.2. Relations with other constructs

Importantly, it should be noted that unless a network is completely isolated – an unlikely situation in psychology – construct labels like "depression" denote in an inherently fuzzy sense. In particular, the distinction between different traits or disorders or abilities is itself a matter of degree, depending on the extent to which the networks are separated. Networks that are not well separated are likely to show entangled behavior that may often cause researchers to wonder whether they are dealing with one or two constructs.

A highly interesting example in clinical psychology where such a situation may arise is comorbidity, i.e., the simultaneous satisfaction of symptoms that belong to multiple disorders. In earlier work we have shown that a network perspective on comorbidity is feasible (Cramer, Waldorp, Van der Maas, & Borsboom, 2010) using the example of the extremely high comorbidity between depression and generalized anxiety. Depression and anxiety each have unique symptoms, but also share symptoms (e.g., fatigue and loss of concentration). Such symptoms may function as bridge symptoms that transfer symptom activation from one network to the other, like a virus may spread from one community to another via people who are in contact with both.

Where the dividing line between intertwined networks lies is a question that has no sharp factual answer, even though the optimal allocation of symptoms to disorders may of considerable practical significance. In accordance, any theory on such a system will have to admit a multiplicity of exceptions arising from the fact that science requires the isolated study of a system, and thus neglects the fact that it is ordinarily situated in a larger network of connections.

4.3. Causes and effects in a network structure

In a network perspective, causes do not work on a latent variable, and effects do not spring from it. Since the individual observables are viewed as causally autonomous, they are responsible for incoming and outgoing causal action. This motivates the study of such observables themselves as gateways of causal action, a perspective that has rarely been taken in psychometric thinking.

Depression again illustrates these issues nicely. Even though its symptoms appear to behave as a unidimensional scale in psychometric research, causal antecedents of depression seem to impact (clusters of) symptoms differentially. For example, adverse life events seem to have stronger ties to psychological symptoms (e.g., depressed mood) than to vegetative symptoms of depression (e.g., concentration problems; see Lux & Kendler, 2010; Tennant, 2001). This suggests that etiological pathways into depression may themselves depend on external events. In addition, it is likely that there are many such pathways, since individual symptoms of depression (e.g., concentration problems or being unable to sleep) can be activated by anything from lower back pain to babies. Any combination from these factors, like babies that cause lower back pain, may be involved as well. Moreover these causes may themselves form new, ever more complicated networks, such as when the sleep loss caused by the lower back pains leads to new babies. As result, even for network constructs whose dynamics are understood there is little hope that science in time will come up with a manageable laundry list of their causes.

Similar concerns involve the study of outgoing effects (e.g., consequences like losing one's job in the context of depression). In some cases, these consequences may be seen as a result of the overall state of the network; in other cases it is more plausible that only a few of the symptoms are responsible. The network perspective offers a natural way to accommodate this, as in dynamical systems even simple interactions between variables may cause emergent phenomena to arise as a result of nonlinear interactions between components of the system (most psychological systems must feature nonlinear relations because at least some of their variables are naturally bounded; e.g., one cannot sleep less than 0 h a night).

Take for instance a suicide attempt by someone who is depressed. Such an act may not be a result of the latent variable, "depression", but rather the result of interactions between symptoms like depressed mood, self-reproach, and suicidal ideation, i.e., three symptoms of depression. Such interactions may lead to a downward spiral from which a person cannot escape, and that spiral could be viewed as an emergent phenomenon with novel causal powers that none of its generating elements possessed. The network approach accommodates these issues naturally, and in a way that no reflective or formative model can do, because it allows us to reason about dynamics within the psychometric context of the indicator variables themselves.

5. Conclusion

In this paper, we have presented a network approach in which the constituents of psychological constructs are directly related in a nontrivial and non-spurious manner. The network approach is intuitively attractive and naturally accommodates what we know about the elusive nature of psychological constructs. It also offers an explanation of why our traditional psychometric approaches have met with so little success, that is, of why after all these years we still do not know whether typical psychometric instruments really measure something and, if so, what that something could be.

While the network approach is not necessarily adequate for all psychological constructs, it may turn out to be so for more constructs than intuition suggests. For instance, one may think that while depression is a nice test case, network models are unlikely to be useful in other domains, like intelligence testing. However, Van der Maas et al. (2006) present a convincing case for a network model of intelligence, the explanatory resources of which rival those of any other contemporary theory. Similarly, personality research seems a feasible area for network applications, because personality items typically list items that are plausible causes and effects (consider "I plan ahead when doing a job" and "I always finish jobs on time" as indicators of

conscienciousness). Longitudinal studies and novel methodologies are needed to investigate the dynamics of psychological constructs. An important question is whether network structure and dynamics differ between persons. In addition, studies investigating the relation between network properties (e.g., the distribution of connection weights, or the number and type of equilibrium points) on one hand, and the possible range of configurations of cross-sectional data obtained at a single time point (e.g., data conforming to a one-factor model or a five factor model with correlated factors) on the other may provide useful starting points for dynamic accounts of psychological constructs. For instance, if a cross-sectional data set conforms to a five factor model, what are the necessary and possible properties of a network that generated these data, and vice versa?

Past decades have resulted in a significant set of tools that can be used to study and evaluate the structure and dynamics of networks. Such approaches have gone largely unnoticed in psychometrics, validity theory, and psychological testing; however, they may offer significant potential for advancing our understanding of psychological constructs.

Acknowledgements

This work was supported by NWO innovational research grant no. 451-03-068. The authors wish to thank Conor Dolan and Jelte Wicherts for providing the NEO-PI-R data set.

References

Achermann, P. (2004). The two-process model of sleep regulation revisited. *Aviation, Space, and Environmental Medicine, 75*, A37–A43.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (DSM-IV)*. Washington DC: APA.

Bagozzi, R. P. (2007). On the meaning of formative measurement and how it differs from reflective measurement: comment on Howell, Breivik, and Wilcox (2007). *Psychological Methods, 12*, 229–237.

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

Baxter, R. (2009). Reflective and formative metrics of relationship value: a commentary essay. *Journal of Business Research, 62*, 1370–1377.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord, & M. R. Novick (Eds.), *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Blalock, H. M. (1964). *Causal inferences in nonexperimental research*. Chapel Hill: University of North Carolina Press.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D.-U. (2006). Complex networks: structure and dynamics. *Physics Reports, 424*, 175–308.

Bollen, K. A. (1984). Multiple indicators: internal consistency or no necessary relationship? *Quality and Quantity, 17*, 377–385.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A. (2007). Interpretational confounding is due to misspecification, not to type of indicator: comment on Howell, Breivik, and Wilcox (2007). *Psychological Methods, 12*, 219–228.

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: a structural equation perspective. *Psychological Bulletin, 110*, 305–314.

Borsboom, D., Cramer, A. O. J., Kievit, R. A., Zand-Scholten, A., & Franić, S. (2009). The end of construct validity. In R. W. Lissitz (Ed.), *The concept of validity*. Information Age Publishers.

Borsboom, D., Cramer, A. O. J., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (submitted for publication). The small world of psychopathology.

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*, 203–219.

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*, 1061–1071.

Burt, R. S. (1976). Interpretational confounding of unobserved variables in structural equation models. *Sociological Methods and Research, 5*, 3–52.

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: two applications of formative measurement. *Journal of Business Research, 61*, 1250–1262.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281–302.

Cramer, A. O. J., Waldorp, L. J., Van der Maas, H. L. J., & Borsboom, D. (2010). Comorbidity: a network perspective. *Behavioral and Brain Sciences, 33*, 137–193.

Csárdi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *Inter Journal Complex Systems*, 1695, Available from http://www.interjournal.org/.

Diamantopoulos, A. (2006). The error term in formative measurement models: interpretation and modeling implications. *Journal of Modelling in Management, 1*, 7–17.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research, 61*, 1203–1218.

Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: a comparison and empirical illustration. *British Journal of Management, 17*, 263–282.

Dolan, C. V., Oort, F. J., Stoel, R. D., & Wicherts, J. M. (2009). Testing measurement invariance in the target rotates multigroup exploratory factor model. *Structural Equation Modeling, 16*, 295–314.

Edwards, J. R. (2011). The fallacy of formative measurement. *Organizational Research Methods, 14*, 370–388, doi: 10.1177/1094428110378369.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods, 5*, 155–174.

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2011). qgraph: Network representations of relationships in data. R package version 0.4.10. Available from http://CRAN.R-project.org/package=qgraph.

Finelli, L. A., Baumann, H., Borbély, A. A., & Achermann, P. (2000). Dual electroencephalogram markers of human sleep homeostasis: correlation between theta activity in waking and slow-wave activity in sleep. *Neuroscience, 101*, 523–529.

Franke, G. R., Preacher, K. J., & Rigdon, E. E. (2008). Proportional structural effects of formative indicators. *Journal of Business Research, 61*, 1229–1237.

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software – Practice & Experience, 21*, 1129–1164.

Hoekstra, H. A., Ormel, J., & De Fruyt, F. (2003). *NEO-PI-R/NEO-FFI: Big five persoonlijkheidsvragenlijst. Handleiding [Manual of the Dutch version of the NEO-PI-R/NEO-FFI]*. Lisse, The Netherlands: Swets and Zeitlinger.

Howell, R. D., Breivik, E., & Wilcox, J. B. (2007a). Reconsidering formative measurement. *Psychological Methods, 12*, 205–218.

Howell, R. D., Breivik, E., & Wilcox, J. B. (2007b). Is formative measurement really measurement? Reply to Bollen (2007) and Bagozzi (2007). *Psychological Methods, 12*, 238–245.

Jarvis, C. B., Mackenzie, S. B., & Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecication in marketing and consumer research. *Journal of Consumer Research, 30*, 199–217.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*, 109–133.

Kendler, K. S., Zachar, P., & Craver, C. (2010). What kinds of things are psychiatric disorders? *Psychological Medicine*,. doi: 10.1017/S0033291710001844.

Lawley, D. N., & Maxwell, A. E. (1963). *Factor analysis as a statistical method*. London: Butterworth.

Lazarsfeld, P. F. (1959). Latent structure analysis. In S. Koch (Ed.), *Psychology: A study of science, Vol. 3* (pp. 476–543). New York: McGraw Hill.

Lux, V., & Kendler, K. S. (2010). Deconstructing major depression: a validation study of the DSM-IV symptomatic criteria. *Psychological Medicine*, doi: 10.1017/S0033291709992157.

Maraun, M. D., & Peters, J. (2005). What does it mean that an issue is conceptual in nature? *Journal of Personality Assessment, 85*, 128–133.

McGrath, R. E. (2005). Conceptual complexity and construct validity. *Journal of Personality Assessment, 85*, 112–124.

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York: John Wiley & Sons.

Mellenbergh, G. J. (1994). Generalized linear item response theory. *Psychological Bulletin, 115*, 300–307.

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*, 201–218.

Neapolitan, R. (2004). *Learning Bayesian networks*. Upper Saddle River, NJ: Pearson Education.
Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge: Cambridge University Press.
Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Paedagogiske Institut.
Samejima, F. (1969)*Estimation of latent ability using a response pattern of graded scores, Vol. 17*. Psychometrika Monograph.
Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search*. MIT Press.
Tennant, C. (2001). Life events, stress and depression: a review of recent findings. *Australian and New Zealand Journal of Psychiatry, 36*, 173–182.
Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences, 21*, 615–665.
Van der Maas, H. L. J., Dolan, C., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychological Review, 113*, 842–861.
Van der Maas, H. L. J., & Molenaar, P. C. M. (1992). Stagewise cognitive development: an application of catastrophe theory. *Psychological Review, 113*, 842–861.
Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*, 440–442.
Wilcox, J. B., Howell, R. D., & Breivik, E. (2008). *Questions about formative*.