Psychometric Perspectives on Diagnostic Systems

▼

Denny Borsboom
University of Amsterdam

The author identifies four conceptualizations of the relation between symptoms and disorders as utilized in diagnostic systems such as the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (*DSM-IV*; American Psychiatric Association, 1994): A constructivist perspective, which holds that disorders are conveniently grouped sets of symptoms; a diagnostic perspective, which holds that disorders are latent classes underlying the symptoms; a dimensional perspective, which holds that symptoms measure latent continua; and a causal systems perspective, which holds that disorders are causal networks consisting of symptoms and direct causal relations between them. Advantages and disadvantages of these conceptualizations are discussed. The author concludes that the psychometric analysis of diagnostic systems is not settled, and that these systems require deeper psychometric analysis than they currently receive. © 2008 Wiley Periodicals, Inc. J Clin Psychol 64: 1089–1108, 2008.

Keywords: diagnostic systems; psychometrics; theoretical psychology; latent variable models; causal networks

Over the past decades, considerable developments have taken place in the diagnosis of mental disorders. The fields of clinical psychology and psychiatry have moved from the ill-described and idiosyncratic methods of assessment, used during the greater part of the 20th century, to standardized methods of diagnosis involving a widely used diagnostic system as captured in, for instance, the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (*DSM-IV*; American Psychiatric Association, 1994). The *DSM* results partly from consensus and partly from compromise, and inherits the weaknesses inherent in both; however, there

Correspondence concerning this article should be addressed to: Denny Borsboom, Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands; e-mail: d.borsboom@uva.nl

seems to be a widespread agreement that, in general, its development has been beneficial to the understanding of mental disorders. At the very least, the use of standardized assessment methods facilitates communication between researchers, and, to some extent, increases the comparability of studies carried out at different locations by different researchers. In any event, diagnostic systems such as the *DSM* currently form the basis for much of the scientific research in clinical psychology and psychiatry.

It should be noted, however, that this movement towards standardization has not been paralleled by theoretical advances in understanding the conceptual and psychometric underpinnings of diagnostic systems in general, and the *DSM* in particular. One can see this clearly by asking a simple question: "What is it that a researcher, who uses the *DSM* for classification, really does?" There is no single theoretically and psychometrically convincing answer to this question; rather, as I will show in this article, there are several distinct answers, which are all plausible to some extent. The primary aim of the present article is to clarify these interpretations of diagnostic systems on the basis of insights taken from psychometric theory.

In what follows, I describe three accounts that may be used to conceptualize the relation between observables (in the diagnostic world commonly designated as symptoms) and theoretical constructs (mental disorders) and associated views of the diagnostic process. First, the *constructivist view*, which holds that disorders are constructed by researchers and clinicians on the basis of convenient groupings of symptoms; second, the *diagnostic view*, which says that these symptoms measure categorical latent classes; and third, the *dimensional view*, which maintains that symptoms measure latent continua. Subsequently, I will discuss several problems with traditional psychometric models, which have been identified in the psychometric literature over the past few years, because these problems appear to be particularly salient in models for diagnostic systems. Finally, I will sketch an alternative conceptualization based on the idea that disorders may be causal systems, rather than constructions of the researcher or latent variables measured through groups of symptoms.

The Constructivist View

One possible response to our focal question is that the researcher who uses the *DSM* for classification constructs classes of people based on a convenient grouping of symptoms into syndromes. In this view, the classification system is seen as relatively arbitrary, which renders the resulting classes of people socially constructed kinds rather than naturally existing ones (e.g., see Hacking, 1999). From this point of view, a classification label such as "depressed" is more similar to, say, "yuppie," than to "suffering from type 1 diabetes." The concept of a yuppie is a socially constructed kind in the sense that it is implicitly defined by a convenient grouping of key attributes (being young, urban, financially well-off, etc.). Although such attributes may hang together statistically, the concept that describes them does not identify a homogeneous group of people, or at least it does not do so in a scientifically interesting sense. In contrast, the label, "suffers from type 1 diabetes," does identify such a group. This term identifies a subpopulation of people who are homogeneous at a level deeper than that of their manifest symptoms: for they share a deficit in a causal mechanism that produces these symptoms (i.e., insulin production in the pancreas). If one considers "depressed" to be a label that functions in a similar way as the label "yuppie"—as a label that is merely useful to delineate a group of people

who share some key attributes (e.g., having little pleasure or interest in life; depressed mood), but does not "cut nature at its joints"—then one views depression as a social or logical construction. In what follows, I will designate this view as *constructivist*.

Constructivist conceptualizations have an arbitrary component, but it should be clearly recognized that this does not imply that the whole process of diagnosis, and the results of scientific research on mental disorders, are also arbitrary. Obviously, for instance, the symptoms of depression hang together reliably, in the sense that they are moderately positively correlated, and for this reason, the syndromes constructed out of them will have a sense of reliability as well. Perhaps the best way to interpret this sort of reliability is in the classical psychometric sense of internal consistency. The higher the intercorrelations between a set of measures, the higher internal consistency will be. Further, under the assumptions of classical test theory (Lord & Novick, 1968), internal consistency estimates based on a set of items are a lower bound to the reliability of the composite formed from these items (i.e., the sum score).[1] In a similar way, a constructivist may hold that syndromes have a sense of consistency and stability, and thus acknowledge that in this sense, they are not arbitrary. In addition, because people involved in diagnostic activity are supposed to behave in a standardized fashion, different persons who diagnose the same person are expected to produce comparable scores (i.e., the agreement among observers will be considerable). Moreover, the constructivist may freely admit that people so diagnosed can respond to treatments (e.g., serotonin reuptake inhibitors; cognitive–behavioral therapy) with some homogeneity (e.g., many or most may show a decrease in the number or severity of their symptoms). To see that this does not contradict the constructivist position, it is illustrative to note that people may respond to treatment (e.g., aspirin) with a reliable change of symptoms (e.g., decreasing fever) while they suffer from very different conditions (e.g., influenza, measles, pneumonia, malaria, AIDS). Naturally, fever itself may be a causally homogeneous symptom, in the sense that it results from the same bodily processes whenever it occurs, but that is not the point here. The point is that to deny that symptoms like fever map to a homogeneous underlying syndrome is consistent with acknowledging that the symptom can be uniformly responsive to a given treatment like taking aspirin. The constructivist can acknowledge all of these nonarbitrary aspects involving the diagnosis of mental disorders without contradicting himself or herself.

What a constructivist does deny, however, is that a group of symptoms is anything more than just that: a group of symptoms. Unlike the term, "suffers from type 1 diabetes," which identifies homogeneity at a deeper level than the symptoms of diabetes themselves, the term depressed in this view *only* produces a reliable classification. In a psychometric sense, one could say that a constructivist accepts that a set of symptoms may have high internal consistency, but denies that they all measure the same latent variable (unidimensionality). Now, it is important, at this point, not to get trapped in a widely accepted but false idea of the relation between internal consistency and unidimensionality, which is the idea that high internal consistency is evidence for unidimensionality. Empirically speaking, internal consistency is nothing more than a summary statistic of the intercorrelations

---

[1]The meaning of the word "reliability" in this sentence is not clear to all who think about psychometric matters (Borsboom, 2005), but researchers seem to get along quite well with it in the sense that they have an intuitive understanding of reliability. I will not go into the interesting, but difficult question of what it is precisely, that they have an understanding *of* because it is tangential to my present purposes.

between a set of variables, and these correlations may come from everywhere and nowhere. Any set of positively correlated variables (IQ, educational attainment, salary, health, etc.) will show high internal consistency if run through the relevant statistical analyses, even though they do not measure the same latent variable. Conversely, a set of indicators that do measure the same latent variable may have low internal consistency if the observables are contaminated by sizeable amounts of measurement error. Therefore, there is no simple inference ticket from internal consistency to unidimensionality.

Another way to see this is to realize that, as Haig (2005a) has convincingly argued (see also Borsboom, Mellenbergh, & Van Heerden, 2003), the inference that a set of indicators are affected by the same latent variable requires an abductive step: this means that the acceptance of the latent variable hypothesis is not mandated by the data alone, but requires an appeal to the explanatory merits of this hypothesis. Such explanatory merits may, for instance, involve strong theory about the processes that connect the latent variable to its indicators (Borsboom, Mellenbergh, & Van Heerden, 2004) and coherence of the hypothesis with a body of accepted knowledge (Haig, 2005a, 2005b). High internal consistency by itself, however, is not one of these explanatory merits because it is merely a function of the correlations between observed variables, and such correlations by themselves have no explanatory force.

Various statistical modeling schemes are at the disposal of one who endorses constructivism. These modeling schemes, generically known as *formative modeling*, have in common that they consider a theoretical term, such as depression, to be a function of the observable symptoms rather than a common cause of them. Examples of such models are described in Bollen and Lennox (1991) and Edwards and Bagozzi (2000); a well-known instantiation of formative modeling is Principal Components Analysis (PCA; see Bollen & Lennox, 1991). For a general discussion on the status of formative models, see Bagozzi (2007), Bollen (2007), and Howell, Breivik, and Wilcox (2007a, 2007b).

In exactly the same way that a psychometrician can accept that a set of indicators may have high internal consistency, interobserver reliability, show a homogeneous response to experimental manipulations, and allow the construction of pragmatically useful composites, but at the same time deny that these indicators measure the same latent variable, so the constructivist can accept that symptom groups have all of those merits, but at the same time deny that they measure an underlying syndrome. The constructivist position is thus crucially dependent on a negative appraisal of the hypothesis that the symptoms hang together because of an underlying condition. In other words, a constructivist refuses to take this abductive step, as described by Haig (2005a, 2005b) in the context of latent variable models—namely, from correlations between symptoms to an underlying condition. Two positions that explicitly do make this abductive step are described next.

The Diagnostic View

A second possible description of the diagnostic process, as followed by a researcher who uses the *DSM* for classification, is that such a researcher is involved in the determination of latent class membership on the basis of manifest responses to diagnostic questions. From this perspective, which I will designate as the *diagnostic view*, because of its similarity to the classical idea of diagnosis as it originated in medicine, symptoms in the *DSM* are more than conveniently grouped variables; they are indicators of some underlying condition that, although we may not have direct

observational access to it, does exist as a phenomenon independent of any diagnostic activities. That is, there is such a thing as depression, in the sense that we could be objectively right or wrong in diagnosing people as depressed. This means that there is more to our being right or wrong than, say, being merely consistent or inconsistent with a set of conventions (e.g., "diagnose as depressed when the *DSM*-criteria for depression are met," or "match diagnosis as closely as possible to a conventional grouping of symptoms"). When the notion of error (i.e., misdiagnosis) is taken to depend not on the adherence to, or violation of, such a set of conventions, but on the actual state of affairs in the world, then one is prepared to take a realist position with respect to mental disorders. That is, in such a view, the notion of error (i.e., wrong diagnosis) is crucially dependent on the existence of a true value on the measured variable (i.e., the underlying condition); in this sense, the notion of falsity is parasitic on the notion of truth.

To flesh out the difference with the constructivist perspective, it is instructive to note that one cannot be misdiagnosed as a yuppie; if one is young, rich, urban, etc., one cannot fail to be a yuppie because that is how the concept of yuppie is defined. There is no deeper reality to the term. In the diagnostic view, however, there *is* such a deeper level of reality; to return to our previous example, a person may suffer from the symptoms of type 1 diabetes (e.g., thirst, weight loss, nausea) for other reasons than the condition itself; that is, the symptoms may be the result of an altogether different disease. Thus, to admit the possibility of an erroneous diagnosis implies, in a relevant sense, the acceptance of the hypothesis that the condition itself exists. That is, one has to minimally accept that "John has disease x" has a definite truth value, which is independent of the outcome of attempts diagnose him; otherwise a diagnosis like "John does not have disease x" cannot be erroneous. Thus, in a relevant sense, the condition of suffering from some disorder is something that is (ontologically) distinct from the symptoms; it is not purely a function of them.

Because conditions like being depressed are not directly observable, the distinction between people who do and people who do not suffer from a mental disorder necessarily has hypothetical elements. However, the hypothesis that there is such a distinction is not inconsequential. For instance, if one supposes that symptoms are fallible indicators of class membership, where class membership is discrete (e.g., "does or does not suffer from depression" or "suffers from depression type a, b, c, . . ."), the implication that follows from this is that the symptoms should be statistically independent conditional on class membership. This is a standard testable implication of the latent class model (Heinen, 1996; Lazarsfeld & Henry, 1968). The constructivist position does not have such implications.

Another consequence that follows from adopting the diagnostic view is of an altogether different nature; namely, it suggests future courses of research. Because the latent structure of mental disorders, according to the diagnostic view, is categorical, there must be something deeper than the mere symptoms; and this something homogenizes the people suffering from a given mental disorder like depression, just like the failure of insulin production in the pancreas homogenizes the population of people suffering from type 1 diabetes. By "homogenizes" I mean that the diagnostic view implicitly promises that, with respect to some deeper level than the symptoms, those who suffer from a mental disorder form an equivalence class: They are "exchangeable" at a deeper level than that of the symptoms. If such a level does not exist, then the hypothesized latent classes are purely fictional, and the diagnostic view misses out on an ontological foundation that is essential to it; in this case the position collapses to constructivism with fancy statistics.

Many deeper levels of reality could be envisioned to do the job of homogenization, but in our time the level that many people think is the right level to look at is the level of biology. Therefore, researchers who adhere to the diagnostic view for instance, may hope to someday get a handle on some neural mechanism that determines a mental disorder like depression. Such researchers may expect that, at some currently unknown level, patients who suffer from a given psychopathology will turn out to be homogeneous in a nontrivial sense; they might share, for instance, a genetic deficit or be characterized by a disturbance in the equilibrium of neurotransmitters. Proponents of the diagnostic view would thus be prepared to invest money and time into the search for, say, the neural mechanisms that underlie depression. A constructivist obviously would not expect serious scientific payoff from such an exercise, for much the same reason that one would not expect much of attempts to uncover the neural mechanism that produces yuppies.

It is beyond the scope of the present article to assess the successes and failures of biological approaches in achieving the task of homogenization, but it is nevertheless interesting to evaluate tentatively some cases in which this task may turn out to fail. To give one example, especially in popular media it is sometimes suggested that depression is a simple function of disturbances of the serotonin balance in the brain. However, at least some authors argue that this hypothesis entertains very limited support, suggesting that even if there is such a relation, it must be a weak one (Lacasse & Leo, 2005). Similarly, despite the high heritability coefficients for most psychologically interesting variables, including depression (Boomsma, Busjahn, & Peltonen, 2002), the search for specific genetic deficits underlying this condition has so far achieved limited success. Some studies have suggested that specific genetic loci are relevant to disorders like depression (e.g., Lesch et al., 1996; Zubenko, Highers, Stiffler, Zubenko, & Kaplan, 2002), but these claims have not proven robust in replication studies (Beem et al., 2006; Middeldorp et al., 2007).

It is interesting to speculate what would happen if the task of homogenization fails in the case of mental disorders like depression, that is, when it turns out that such a disorder is not uniformly realized in different people (e.g., one's liability to develop depression is strongly polygenic and the actual realization of a depressive disorder is not biologically or otherwise homogeneous). This may imply that there is no deeper level of reality than the symptoms themselves to justify the realist assumption of the diagnostic view; that is, there is no level at which we find equivalence classes of people that correspond to the distinction of "depressed" and "not depressed" (or more complicated categorical schemes). What should we conclude if this turned out to be the case? One possible response, of course, is to give up the hypothesis that there are underlying conditions that give rise to symptoms of mental disorders, and take a constructivist view; another alternative, however, is to suggest that the categorical latent structure that characterizes the diagnostic view is too simple. The latter approach may follow a line of reasoning in which depression is not seen as a categorical structure, but as a continuum that smoothly extends into the normal population (Solomon, Haaga, & Arnow, 2001). The corresponding psychometric view is that symptoms depend not on causally homogeneous latent classes, but on latent continua that determine psychopathology. This perspective leads to a dimensional view of psychopathology.

The Dimensional View

A third description of the diagnostic processes followed in clinical research is that such processes involve the determination of persons' positions on a latent continuum

on the basis of their manifest responses to diagnostic questions. This view, which I will designate as the dimensional view, differs from the diagnostic view mainly because its proponents conceptualize disorders as continua rather than as discrete classes.

From this point of view, which is heavily inspired by traditional psychometric theory, the continua are real, but the cut points that define disorders may be arbitrary. Proponents of the dimensional view commonly see patient populations as extremes on a continuum that may extend into the normal population, so that people who suffer from a mental disorder need not be homogeneous in the sense that they share something that normal people do not have (Solomon et al., 2001). In the dimensional view, madness is a matter of degree; for instance, people who suffer from depression are really just people who occupy a higher position on a continuous latent variable rather than being qualitatively different from the normal population. This is a marked difference from the categorical system defining the diagnostic view. For obvious reasons, those who adhere to the dimensional view typically have difficulty squaring the categorical classification procedures implied by the *DSM* with the presumed continuous character of disorders, and would favor, for instance, the use of (weighted) sum scores on diagnostic checklists, instead of categorical assignments to disorders, as focal variables in research.

To get a clear view of the difference between the diagnostic and dimensional perspectives, it is instructive to consider again the analogy with medicine. A factor analogous to the dimensional perspective would not be a causally homogeneous disorder like diabetes, but a continuous factor that underlies certain problems. For instance, extremely tall people may have an increased probability of various problems that are causally related to their height being out of the ordinary (mostly problems that relate to the environment—the design of offices, cars, etc.—being tailored to people of average height; e.g., repetitive strain injury, suffering trauma in accidents, back pain). However, they do not have something that normal people do not have; they merely have a high position on the continuum of bodily height that is causally relevant to these problems. The dimensional view is based on the idea that a similar situation occurs for psychopathology; depressed people are positioned high along some mood-related continuum, and it is the position on this continuum that determines their liability to develop various symptoms. Symptoms are considered to also have a location on the continuum, called a *threshold*, which determines how easily they develop: The lower the threshold, the easier one will develop the symptom. The resulting model, which is sometimes called a *liability-threshold model*, hypothesizes a trade-off between symptom properties (thresholds) and person properties (liabilities) that is essentially the same as the trade-off modeled in educational assessments (between the item property of difficulty and the person property of ability). Formal modeling of this trade-off is done through the application of the same models in both fields: item response theory (IRT) models (e.g., see Aggen, Neale, & Kendler, 2005).

Because the dimensional view conceptualizes mental disorders as continuous attributes, it naturally allows for greater heterogeneity at the level of such attributes. This is a consequence of invoking a continuous latent variable; on a continuum, there is an infinite number of positions for any patient to take. Researchers and clinicians find this an attractive property because it naturally allows for discrimination between different levels of severity of a disorder—a possibility that the diagnostic view does not automatically accommodate (although it must be noted that the extension of a latent class model into a system with categories like not

depressed, mildly depressed, severely depressed is psychometrically trivial; see, for instance, Sullivan, Kessler, & Kendler, 1998). It is perhaps worth noting, however, that this property of the continuous model does not come free. The assumption that a latent continuum underlies the symptoms, or more importantly, the assumption that only one latent continuum underlies them, is not empirically empty—it has consequences for the structure of the probability distribution over the item responses. Thus, one cannot just say that depression is a latent continuum, and be done with it. For we may just as well be talking about a set of latent continua, intertwined in various ways, perhaps even about different sets of continua for different groups of people (e.g., men and women, or ethnic groups). Given that latent variables are latent, we have no direct way of knowing what their structure is.

Fortunately, the latent continuum hypothesis—like the latent class hypothesis—has testable consequences that allow us to investigate whether it is tenable (it should be noted that this is by no means easy, but at least the possibility exists). For instance, if symptoms are dichotomous (present/absent), monotonically increasing in the latent variable (such that for every symptom the probability of having it is greater for higher positions on that latent variable), and unidimensional (so that the probability of symptoms depends solely on the latent continuum), the hypothesized system is an IRT model (Sijtsma, 1998). At the level of observed symptoms, such a model has implications for the frequency distribution of people over symptoms. If these implications are not borne out, then something is wrong with the hypothesized model; if they are, then the model is to some extent corroborated.

To get a clear view of the sort of implications that follow from assuming a continuous latent variable, it is helpful to consider a mundane example of a trait commonly viewed as continuous, say, working memory capacity. The idea is that this capacity can be represented as a single dimension (a line) on which everybody has a position (a point on the line), and that the probability to answer an item, that requires a certain amount of working memory capacity, correctly increases with higher levels of ability: The higher your working memory capacity, the better you do on the item. This is a simple way of stating two of the core assumptions of the IRT model, to wit, unidimensionality and monotonicity. We will get to the third core assumption, local independence, shortly, but for now we will skip over it.

What does such a dimensional structure imply? It implies some quite important things. Suppose we measure working memory capacity with a digit span task, in which the respondent has to repeat sequences of digits (typically the numbers 0–9). Now, if you can repeat the sequence, 3–5–8–4–6–0–7–2–1–9, then we are quite confident that you can also repeat the sequence, 4–8–5–3–7, and virtually certain that you can repeat 6–2. In general, the model structure implies that if a person masters a highly difficult item, she or he will most likely master the great majority of the items that are less difficult. The key assumption that gives rise to this implication is that the items are unidimensional, which means that all of them measure (depend on) the same latent variable. This is another way of saying that given an item, the only thing that matters for whether a person will solve it is the distance of her or his point on the ability scale to the difficulty of the item. Now, because people differ on their positions on the continuum, and items do so too, we may expect to see heterogeneity in the item responses, in the sense that not everybody solves exactly the same items. However, we do expect a very strict ordering in the items and persons; persons who master very difficult items like 3–5–8–4–6–0–7–2–1–9, but fail on 6–2 should be extremely rare, for instance.

It is questionable whether this is the sort of heterogeneity that we see (or that we would expect) in data on, say, depression. For instance, the least prevalent symptom in *DSM-IV*-based criteria for depression is whether the patient has attempted to commit suicide. A much more prevalent symptom is whether the person is fatigued. If the depression symptoms measured the same continuous latent variable, and fatigue and suicide attempts were merely items that differed in their difficulty, then the analogy to difficult and easy digit span items would not be merely metaphorical, but exact. The model structure would then imply that there are much fewer people who have attempted to commit suicide, but were not fatigued, than there are people who have attempted suicide and were fatigued. Moreover, this should be the case for exactly the same reason that there are fewer people, who succeed on a difficult digit span item but fail on an easy one, than there are people with the reversed pattern. The items are exchangeable save for their difficulty; the people are exchangeable save for their ability: What determines the response frequencies is a trade-off between these factors and nothing more.

Now, in this respect I think that most will agree that there is a difference between the relation between fatigue and suicide attempts, on the one hand, and the relation between repeating 3–5–8–4–6–0–7–2–1–9, and repeating 6–2 on the other. That is, it is hard to see why having attempted suicide would have strong implications for fatigue, whereas it is easy to see why being able to repeat 3–5–8–4–6–0–7–2–1–9 has strong implications for being able to repeat 6–2. Where does this difference come from? Plausibly, repeating 3–5–8–4–6–0–7–2–1–9 requires essentially the same resources as repeating 6–2, so to successfully respond to a more difficult item just requires more of the same—here, greater working memory capacity. A similar connection between the variables, "fatigue" and "having attempted to commit suicide" is, however, quite a lot less obvious. That is, it is not at all clear that the difference in prevalence between attempting to commit suicide and being fatigued is merely a matter of suicide attempts requiring "more of the same" than fatigue. There is, of course, an empirical connection between these variables—they are not in the coding scheme for depression by accident, and they do prove to be positively correlated—but it is truly doubtful whether this empirical connection occurs as a result of the fact that these variables measure the same attribute. Yet, this is what standard models like the IRT model hypothesize.

Thus, although dimensional models allow for some heterogeneity, this may not be the kind of heterogeneity that we would expect for psychopathology symptoms like those of depression. That is, such symptoms are likely to be heterogeneous in a way that unidimensional latent variable models do not allow for; that is, they may be heterogeneous in the sense that the symptoms measure completely different things. This kind of heterogeneity, which I will denote as strong heterogeneity, is not solved by a dimensional view of psychopathological syndromes. Should disorders like depression be heterogeneous in this strong sense, then the dimensional view would be a cover-up for—and not a solution to—the real problem, which is that such disorders are themselves heterogeneous. In that case, their symptoms are not measures of a single latent continuum in the way that different digit span items may measure working memory capacity; yet that *is* the way that unidimensional models picture the situation.

Of course, the treatment above is based on a unidimensional modeling scheme, and the division of a diagnostic system into subscales that measure different facets of depression (e.g., vegetative, mood-related, and cognitive dimensions) may be more appropriate in such situations. It certainly represents a possibility to accommodate

strong heterogeneity. However, it is important to see that this does not speak to the problem we are currently considering. One reason is that the solution simply acknowledges that the symptoms do not measure the same continuum, and hence their correlations cannot be explained based on the hypothesis that they do (remember that this was what we were trying to achieve). Another reason is that the solution simply reintroduces the same problem that the introduction of a latent continuum was supposed to solve, only at a different level: instead of the question, "Why are the symptoms correlated?" it yields the question, "Why are the dimensions correlated?" In this case, the central question has been pushed back to a higher level of abstraction, but not solved.

Thus, the dimensional view does create some room for heterogeneity, but it is not clear that it does so in a way that is truly appropriate for the diagnostic systems used in psychopathology. Despite the fact that psychometric analyses of psychopathological symptoms may be consistent with the empirical implications of IRT models (see Aggen et al., 2005, in the context of depression; however see also Keller & Kempf, 1997, who reach a different conclusion), the conditions for fruitfully applying such models are not trivial and, perhaps more importantly, it is not at all easy to imagine how they could be met in the first place. By far the most important hypothesis in these models is that the different symptoms measure the same attribute, and it is precisely this hypothesis that is difficult to back up, either with theory or common sense. Therefore, it is important to consider the meaning of the word *measurement*, as we are using it presently, in detail, and to connect it to causal structures that measurement models encode.

Measurement and Causality

We have discussed three ways of looking at the relation between symptoms and constructs like depression, two of which—the diagnostic and dimensional views—construct the relation between symptoms and construct as one of measurement. Whereas the constructivist view accepts empirical relations between symptoms as a fact, but makes no assumptions on the origin of these relations, the diagnostic and dimensional views share the idea that the symptoms hang together empirically because they measure the same latent structure. This structure is categorical in the diagnostic view and continuous in the dimensional view, but in both cases it plays the same role; namely it enters in the model as a representative for "that which the symptoms measure"—whatever it may turn out to be.

The intuitions that render the diagnostic and dimensional views attractive are that (a) symptoms do not correlate by accident (i.e., there must be some reason why they correlate), and (b) symptoms within a disorder correlate more strongly than symptoms between disorders (i.e., even though comorbidity in *DSM* diagnoses is extremely high, symptoms do cluster in systematic ways; e.g., see also Hartman et al., 2001). Latent structure models explain these facts, that is, symptoms hang together because they measure the same thing, and correlations between symptoms within a disorder are higher than correlations between symptoms belonging to different disorders because they measure different things. Note that these are just the implications of the logic of convergent–divergent validity, as first explained by Campbell and Fiske (1959). As Hartman et al. (2001) show, the diagnostic system of the *DSM* does not fare altogether badly with respect to these implications; although there is certainly room for improvement, factor models consistent with them fit the data relatively well.

Measurement, however, requires more than model fit; there must be a plausible account of how the attributes to be measured are causally connected to a set of indicators (here, the symptoms of the *DSM-IV*). That is, for a set of indicators to measure an attribute it is required that differences in position on the attribute structure (John is depressed, while Jane is not) cause differences in the symptoms (John sleeps badly, while Jane does not). When such causal relevance of the attribute for the indicators is absent, it is hard to defend the supposition that the indicators are valid measures of the attribute, because in this case they are not measures of the attribute at all (Bollen, 1989; Borsboom, Mellenbergh, & Van Heerden, 2004).

There exist several problems with the requirement of a causal connection as applied to the relation between symptoms and disorders. First, to my knowledge there exists no substantively motivated account of what this causal relation should be. In fact, insofar as a causal interpretation of this relation could be given at all, it is likely to involve the emulation of a psychometric model in quasi-substantive terms (e.g., with liabilities and thresholds) rather than a genuinely substantive model of the measurement process that could steer and motivate psychometric models. As is very often the case in psychology, we have many candidate measurement models—the dimensional and diagnostic view are merely based on two typical psychometric models, but these by no means exhaust the psychometric possibilities—but much less is available in the way of theory to guide the choice between them (Borsboom, 2006).

Apart from the absence of scientific theory that could justify the interpretation of the relation between symptoms and disorders as one of measurement, there are considerable problems in the causal ontogenesis of symptom patterns. A latent variable model, which must be the basis for the diagnostic and dimensional views discussed above, views correlations between indicator variables as spurious, in the sense that they do not reflect direct causal relations between the indicators, but arise as a result of the fact that the indicators measure the same attribute. One can think of the indicators as a number of (noisy) thermometers; the fact that the thermometers rise and drop in step with each other (i.e., the fact that they are correlated) originates from their common dependence on the temperature of their environment. Thus, in this situation, temperature functions as the common cause of the readings of the different thermometers, which is fully consistent with the structure presumed in measurement models.

One important consequence of a common cause relation is that conditioning on the levels of the common cause screens off correlations between its effects. Therefore, although our different thermometers will be strongly correlated when we examine their readings over a range of temperatures, if we look at their correlation while temperature is held constant, we will see that this correlation vanishes. The reason is that the variation left in the individual thermometer's readings, after the effect of temperature is controlled for, is random error. So conditional on a value of their common cause, indicator variables are uncorrelated. The same implication exists in latent variable models, where this property is called *local independence* ("local" in the sense that one position on the attribute is considered at a time, and "independence" because the indicators are statistically independent in the subpopulation of people who occupy this position). So latent variable models bear a strong resemblance to common cause models (in fact, they are formally indistinguishable). This should not be considered surprising as the very idea of a latent variable model is that we can learn about differences in position on an attribute (conceptualized as a latent variable) from observed differences in a set of

indicators. It is hard to see how this may happen if the indicators do *not* share a common dependence on this latent structure.

There are several problems with this view as it is applied in psychology, two of which stand out clearly. The first problem involves the lack of correspondence between models for interindividual differences and intraindividual processes, a problem that has been well documented over the last few years (Borsboom et al., 2003; Cervone, 2005; Hamaker, Dolan, & Molenaar, 2005; Hamaker, Nesselroade, & Molenaar, 2007; Molenaar, 2005; Molenaar, Huizenga, & Nesselroade, 2003). The problem is that the structure of a model, as derived from interindividual differences research (e.g., correlations between symptoms as computed over many people at a single time point) has no discernable implication for the structure of the processes that go on within an individual (e.g., which would apply to the correlation between symptoms as computed within a single person over many time points, as these would be present in the etiology of symptoms). Thus, the processes that generate data at the individual level may have a completely different structure from that present in a model for the differences between people.

Hamaker et al. (2007) present the problem clearly. Briefly, they show that a latent variable model with a single latent variable may fit the correlations between individual differences, even when the data are generated from arbitrarily complex generating processes (e.g., the dynamic model that governs the development of indicators over time may be a model with five factors rather than one). In addition, they show that if everybody has the same dimensionality and structure at the intraindividual level (a single factor drives correlation between indicators over time), we may find an arbitrarily complex structure at the level of individual differences (a 5-factor model). An exception to the second point occurs when subjects are not only identical in the structure of their dynamic processes, but in addition strict measurement invariance over subjects holds, so that any differences in observed indicator means are a pure function of differences in latent means (Meredith, 1993; see also, Muthén, 1989). In the latter case, the dimensionality of an invariant intraindividual structure should limit at least the dimensionality yielded by a factor analysis of the covariance matrix computed over individual differences, although it will not give estimates of the factor covariance structure that are interpretable at the intraindividual level (e.g., see Muthén, 1989, Eq. 6).[2] The conditions under which the structure of interindividual differences is strictly isomorphic to the structure of intraindividual processes, however, are extremely strong (Molenaar, 2005) and cannot be expected to be met in common applications of psychometric modeling.

Note that this does not mean that we cannot investigate what the relation between intraindividual and interindividual structures looks like; this is certainly possible (e.g., with models that can combine intra- and interindividual variation; Hamaker et al. 2007; Timmerman, 2006; Timmerman & Kiers, 2003). However, it does follow that we cannot routinely assume that the model that describes the dynamics of the individual is isomorphic to the model that describes individual differences. Thus, if we think of liabilities and thresholds as dynamic concepts (so that, say, John's probability of attempting suicide increases as he moves forward on the latent variable of depression) we are doing so entirely on our own account, that is, we cannot adduce the fit of a liability-threshold model to interindividual differences on depression data to substantiate this interpretation. Now, if the dynamic structure of intraindividual development is very different from the structure present in

---

[2]I thank Marieke Timmerman for pointing this out to me.

interindividual differences, it is not entirely clear in what sense we may still have a causal interpretation of the relation between the latent variable depression (i.e., the structure describing interindividual differences) and the observed interindividual differences in the observed symptoms. This point is more complicated than it may at first seem, and a full discussion of it is beyond the scope of this article (e.g., see Borsboom, 2005; Borsboom et al., 2003, 2004). For present purposes, the important thing for the reader to recognize is that the causal relation between latent variables (depression) and indicators (symptoms of depression) is not straightforward; it requires theoretical and empirical justification that goes beyond fitting a latent variable model to individual differences.

This brings us to a second difficulty of the hypothetical structure coded in a latent variable model, as it occurs in the context of psychopathology data. For if we cannot routinely assume that the latent variable model (e.g., the liability-threshold idea) is valid for describing the causal ontogenesis of symptom groups within individual people, then we are forced to devote considerable attention to a theoretical analysis of the processes that may generate the data on which we execute our psychometric analyses. At this level, the diagnostic and dimensional views run into serious problems that may, in some cases, largely invalidate them as plausible candidates for a correct view of the relation between symptoms and theoretical attributes like depression.

The problem is most easily explained through a simple example. Consider the symptom group that defines panic disorder in the *DSM-IV*. Four of the symptoms of panic disorder are (1) recurrent unexpected panic attacks; (2) at least one of the attacks has been followed by 1 month (or more) of one (or more) of (2a) persistent concern about having additional attacks, or (2b) worry about the implications of the attack or its consequences (e.g., losing control, having a heart attack, "going crazy"); and (3) there is a significant change in behavior related to the attacks. A significant change of behavior that is often observed is that people who have suffered from panic attacks tend to avoid public places (i.e., they develop agoraphobia). For clarity, I will restrict the discussion to this particular change of behavior.

Think about the relation between these symptoms and panic disorder as a measurement relation (i.e., the symptoms *measure* panic disorder) as considered in the dimensional and diagnostic views. The model that corresponds to this idea is graphically represented in the left panel of Figure 1. The model says that the symptoms are correlated because they measure the same latent variable. Thus, the observed variables are causally dependent (in some way) on the latent variable that we measure.

Now think about how these symptoms relate to each other. One does not have to do a deep literature search to encounter the not-altogether-implausible idea that, at the level of the individual person, the symptoms are not effects of a common cause at all; rather, they stand in direct causal relations to each other. For instance, a plausible causal ontogenesis of the symptoms is (1) people have a panic attack in a public place, which causes them (2b) to worry about the implications of that event, and (2a) to worry that they may have another one in a public place, as a result of which (3) they do not get out of the house anymore. This sequence of events describes how each symptom arises as a result of the previous one, and therefore describes a *causal network*. This network is represented in the right panel of Figure 1.

Now look at the models in the left and right panels of Figure 1. Which one makes more sense to you? My experience suggests (in fact, up until now the verdicts of scientific researchers, clinicians, and laypeople have been unanimous with respect to
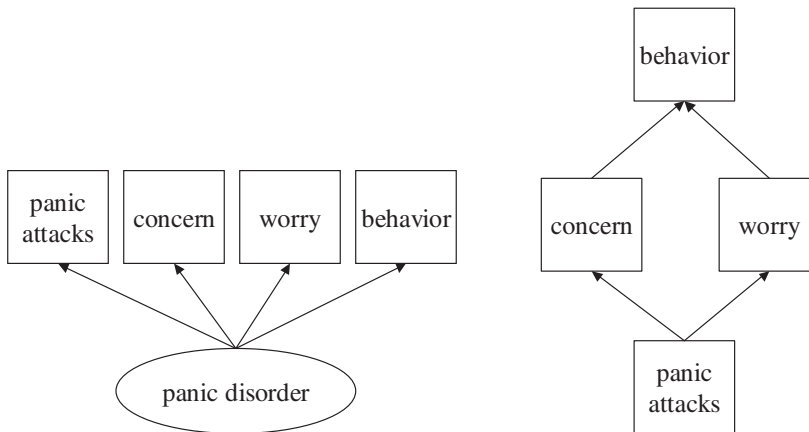
Figure 1. The left panel shows the relation between panic disorder and its symptoms from a latent variable modeling point of view. The right panel shows a representation of these symptoms as a causal system.

this issue) that you will favor the figure in the right panel. Rightly you should, for it makes a good deal of theoretical sense. The implications of this model (should that model be correct) for the idea that we are measuring the same theoretical attribute with these different symptoms are, however, devastating. The existence of direct causal relations between symptoms is in plain contradiction with the core idea that motivates a latent variable model. The reason is that correlations between Symptoms 1, 2a, 2b, and 3 do *not* result from some underlying variable, but reflect the direct effects of Symptom 1 on Symptoms 2a and 2b, and of Symptom 2a and 2b on Symptom 3. Hence, the correlations between the symptoms are not spurious in the sense that a latent variable model assumes them to be; they reflect direct causal effects.

Thus, with respect to this set of symptoms, theoretical considerations do not univocally support the diagnostic or dimensional views. For although the concept of a latent variable, interpreted as the liability to develop symptoms, may be initially attractive, when one recognizes the hypothesized causal structure (a *common cause* structure) and contrasts this with an alternative (a *causal chain*) the measurement model has a strong competitor. Naturally, the proof of the pudding is in the eating, and the analysis of empirical data is necessary to validate the relative empirical successes of the different approaches (although this may not be easy; it is notable, in this respect, that many latent variable models are statistically equivalent to dynamic state-space models, see Molenaar, 2003). In addition, the example given is a relatively clear and plausible account of the ontogenesis of one particular mental disorder. Much more theoretical and empirical research is necessary to establish whether similar accounts can be given of other disorders. However, even a cursory glance through diagnostic systems like the *DSM* suggests a multitude of direct causal links between indicators of disorders. For instance, in depression one encounters symptoms like (a) lack of sleep, (b) fatigue, and (c) lack of concentration, which may plausibly make up a causal chain. Also, seldom encountered symptoms like suicide attempts may be plausibly seen as distal effects of a persevering disturbance in other variables commonly viewed as indicators (e.g., lack of pleasure, depressed mood, feeling of worthlessness), rather than as indicators that stand at the same causal level as all the other indicators (i.e., as effects of a common cause). For symptoms of

various addictive disorders, such causal chains also stand to reason. It is beyond the scope of this article to discuss these structures in detail. Rather, the objective of the present analysis was to point out that the causal ontogenesis of symptoms as a latent variable model pictures it is (a) not at all straightforward, and (b) there are alternative ways of explaining why symptoms hang together.

If panic disorder turns out to be properly conceptualized as a causal system, as represented in the right panel of Figure 1, then the question arises what diagnostic activities really come down to. In this respect, the concept of mental disorders as being *identical* to causal systems of symptoms (analogous to the way van der Maas et al., 2006, conceptualize intelligence as a causal system of mutually supportive modules and processes) may be a fruitful alternative to the hitherto discussed views. In fact, it may actually offer an integrated account of mental disorders that incorporates several of the intuitively attractive properties of the constructivist, diagnostic, and dimensional views. To this topic we now turn.

Disorders as Causal Systems

Suppose, as a matter of interesting speculation, that the relations between symptoms as captured in diagnostic systems are, for a large part, direct causal relations.[3] This means that the conceptualization of disorders as latent variables (either continuous or categorical) is basically wrong. Of course, one may fit a latent variable model to a dataset, and that model may give a reasonable description of the relations between variables. However, if the relations between symptoms are direct, it follows that the latent variables invoked are purely fictional; they cannot be interpreted as real entities. It would moreover be a case of language abuse to say that one measures such latent variables; they would be convenient fictions, invented by the researcher for some pragmatic reason. To say one measures a convenient fiction is surely to stretch the limits of language beyond what is reasonable. At the same time, it would appear that a purely constructivist view fails to acknowledge a very important property of the diagnostic categories, namely that the relations between symptoms that define them are far from arbitrary. That is, disorders do not consist of merely conveniently grouped variables; they consist of sets of symptoms that are connected through a system of causal relations.

Several important questions arise within such a scheme of thinking. The first is the question how we should conceive of the relation between indicators (e.g., lack of sleep) and constructs (e.g., depression). Clearly, we could no longer say that the indicators measure the construct, as in the diagnostic and dimensional views, because the causal structure of measurement (a causal effect of the measured attribute on its indicators) is not satisfied. Neither would we be inclined to say that the symptoms are merely conveniently grouped into the disorder, as in the constructivist view would have it; the relations between the symptoms are causal, i.e., the grouping represents much more than convenience alone. It would seem that the relation between indicators and constructs, in a causal systems perspective, is a *mereological* (i.e., part–whole) relation rather than a causal one. That is, the symptoms are part of a larger system of symptoms and causal connections that we refer to when we use the word depression. They do not measure the construct, but constitute it—and,

---

[3]Various analyses of psychological constructs available in the literature are based on this idea; examples are the analysis of developmental disorders as given in Morton and Frith (2004) and the analysis of the positive manifold of intelligence test data by Van der Maas et al. (2006). In a broader sense, the conceptualization is consistent with the basic tenets of dynamical systems theory (Smith & Thelen, 1994).

importantly, they constitute it in a nonarbitrary way. So viewed, a causal systems' perspective evades the problems inherent in making sense of measurement models for diagnostic systems (namely, figuring out how constructs may have causal relevance of their indicators). At the same time, it does not collapse into the relativistic consequences of constructivist thinking (i.e., all groupings of variables are empirically equally defensible, only some are more convenient for our purposes than others).

A second question that arises within a causal systems perspective is how we should conceptualize diagnostic activity, i.e., what does the application of a diagnostic system really come down to? It is important to note that although the cutoff scores for diagnoses (i.e., five out of nine depression symptoms equals major depressive disorder) are no less arbitrary from a causal systems perspective than from a measurement point of view, it is likely that they convey very nonarbitrary information about (a) whether the causal system of symptoms is at all activated in a person, and (b) where in the causal sequence a person is located at the time of the diagnostic interview. Diagnostic activity could thus be conceptualized as a twofold process, which involves the qualitative step of deciding whether a system is activated (akin to latent class assignment in the diagnostic view) and the quantitative step of deciding what a person's position in the causal structure currently is (akin to the determination of a position on a continuum in the dimensional view).

For instance, satisfaction of the first criterion of panic disorder (recurrent unexpected panic attacks) indicates that the causal system has been entered by the person being diagnosed. The other criteria (worry about attacks and behavioral change, like avoidance of certain situations) indicate "how far down the line" a person is. Panic disorder, as a diagnosis, requires that one is at least sufficiently far in the causal sequence that the symptom "behavior change" has been activated; panic disorder with agoraphobia entails that one can no longer leave one's home without fear. These are somewhat arbitrary lines, which may well be drawn based on pragmatic rather than empirical considerations, but that does not render the structure in the causal system, or the diagnostic activity itself, purely pragmatic. Moreover, one could imagine that, for example, the distinction between panic disorder with and without agoraphobia is empirically nonarbitrary in the responsiveness of conditions to interventions (i.e., one might need different therapies for the disorders).

Similarly, exclusive satisfaction of the criteria lack of sleep and fatigue suggests that the causal system called depression has been entered, but that the person is not far enough down the line to justify a diagnosis of major depressive disorder. To that effect, one must not only proceed farther in the causal sequence, but in addition activate symptoms that are central in the causal system (like depressed mood and lack of pleasure). The notion of centrality, which plays an important role in diagnostic systems in the form of criteria that are necessary for a diagnosis, may receive various interpretations that are beyond the scope of this article. However, to get a feel for the concept one might suppose that central symptoms are symptoms with many incoming and outgoing relations from and to other symptoms that make up the disorder. The converse of centrality occurs when variables are peripheral, that is, are on the border of a causal system; suicide attempts, which can be imagined to be located at the edge of the depression system, may be a peripheral variable in the sense that they are an outcome that many of the other symptoms lead up to. The ability to define central and peripheral variables may be a significant advantage over traditional psychometric models, which cannot conceptualize such distinctions (for

exactly the same reason that there is no rationale to decide which of a set of thermometers is central to the definition of temperature—in a latent variable model, all indicators are on equal footing; Bollen, 1989). However, given the ubiquity of such distinctions in diagnostic systems (as inclusive and noninclusive criteria) it would seem that a psychometric model that cannot capture or formalize them is missing out on something important.

In conclusion, it seems that the idea that constructs in psychopathology may refer to causal systems rather than to latent variables has prima facie plausibility and requires further conceptual analysis. However, it is beyond the scope of the current article to present such an analysis. At this point, I think that the most important message here is that the notion of diagnosis, in psychopathology, is not exhausted by conceptual accounts that parallel traditional psychometric models (e.g., the diagnostic and dimensional views); plausible alternatives that do not imply a collapse into relativism (as implicit in the constructivist view) are available. Thus, one does not have to cling to realism about latent variables (Borsboom et al., 2003) to defend the notion that diagnostic systems are not purely arbitrary.

Discussion

In this article, I have attempted to give a systematic clarification of the process of diagnosis by connecting conceptual views of diagnostic systems to psychometric theories on the relation between symptoms and constructs. Three viewpoints were discussed: the *constructivist view*, which holds that disorders are conveniently grouped sets of symptoms; the *diagnostic view*, which maintains that symptoms measure categorical latent classes of people; and the *dimensional view*, which says that symptoms measure continuous latent dimensions. Finally, I presented an alternative view of mental disorders as *causal systems*, in which the relation between symptoms and disorders is not one of convenient grouping (as in the constructivist view) and neither one of measurement (as in the diagnostic and dimensional views), but one of mereology. The symptoms are parts of a larger causal network; we designate this network when we use words like depression or panic disorder. The process of diagnosis, in the latter perspective, is a two-stage activity in which it is first assessed whether a person has entered a causal network, and second, where in the network a person's current location is.

Some caveats regarding the implications of the present discussion are in order. First, in scientific wildlife, the positions sketched are often encountered as all-inclusive philosophies (e.g., scholars who are dismissive of a realist interpretation of, say, depression, are likely to be dismissive of realism on all other disorders; similarly, those who favor a dimensional perspective for one disorder generally favor a dimensional perspective for all other disorders). However, it should be clearly recognized that there is absolutely no inconsistency in adhering to a constructivist perspective when it comes to, say, personality disorders, a diagnostic perspective when it comes to schizophrenia, and a dimensional perspective when it comes to depression. Thus, there is naturally room for a multitude of models and associated frameworks within the study of a diagnostic system, as long as they do not apply to the same disorders.

Despite this, it is clear that existing psychometric theories of the relation between symptoms and disorders face problems that are of a very general nature. The constructivist view would naturally translate into a formative modeling approach in psychometrics (Borsboom et al., 2003). The formative model is ontologically loose,

in the sense that it does not require the existence of a latent structure and a measurement process that connects this structure to the observations, but it fails to capture the fact that many groupings of symptoms do not appear to be arbitrary in the sense of "merely convenient." Something more is clearly going on in the data. The diagnostic and dimensional views provide an explanation for the nonarbitrary nature of many symptom groups through the hypothesis that the grouped symptoms measure the same latent structure. However, it is not clear how the causal connection between the presumed latent structure and the observable symptoms, which is required to sensibly speak of measurement (Borsboom, 2005; Borsboom et al., 2004), can be fleshed out in the case of mental disorders. Models for interindividual variation (i.e., all the models that have hitherto been fitted to diagnostic systems data) have no clear implications for intraindividual processes (Borsboom et al., 2003; Hamaker et al., 2007), and it is far from clear whether in the case of psychopathology they have any relevance whatsoever at the level of the individual. Thus, the idea that models for intraindividual variation are isomorphic to models for interindividual variation currently has little support, which means that such models cannot be routinely assumed to describe the data generating process at the intraindividual level.

Now, I do not know of alternative conceptualizations of the data generating process in the context of psychopathology, which would adequately tackle this problem in the sense of providing a convincing alternative data generating mechanism to connect variation in a latent structure to variation at the symptom level. This means that there is a serious possibility that the entire latent variable framework is inappropriate for the analysis of diagnostic systems; hence, both the diagnostic and dimensional views are purely metaphorical. Moreover, at least for disorders such as panic disorder, a dynamic causal structure involving direct causal relations between symptoms would seem to be appropriate; this may well be the case for many other disorders. However, such a dynamic structure is in direct contradiction to the tenets of the latent variable model because it violates the common cause structure that such models instantiate (Borsboom et al., 2003). In other words, the diagnostic and dimensional views are in an awkwardly general sort of trouble.

Turning a vice into a virtue, I sketched a conceptual viewpoint based on a causal systems view, akin to van der Maas et al.'s (2006) theory of intelligence, that is based on the idea that symptoms may bear direct causal relations to each other. Unfortunately, however, there is currently no worked-out psychometric theory to go with that perspective. Neither is there any empirical evidence that such a viewpoint may be correct. Both the theoretical and empirical analysis of causal systems as psychometric constructs thus need to be elaborated.

In conclusion, it seems that currently available psychometric conceptualizations of diagnostic systems and of diagnostic activity as such, suffer from serious shortcomings. Hence, whatever else one may think about the considerations presented in this article, I think that one implication is relatively uncontestable: Diagnostic systems deserve more elaborate psychometric thinking and analysis than they currently receive.

References

Aggen, S.H., Neale, M.C., & Kendler, K.S. (2005). DSM criteria for major depression: Evaluating symptom patterns using latent-trait item response models. Psychological Medicine, 35, 475–487.

American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Washington, DC: Author.

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.

Beem, A.L., De Geus, E.J.C., Hottenga, J., Sullivan, P.F., Willemsen, G., Slagboom, P.E., et al. (2006). Combined linkage and association analyses of the 124-bp allele of marker D2S2944 with anxiety, depression, neuroticism, and major depression. Behavior Genetics, 26, 127–136.

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

Bollen, K.A. (2007). Interpretational confounding is due to misspecification, not 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.

Boomsma, D.I., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. Nature Reviews Genetics, 3, 872–882.

Borsboom, D. (2005). Measuring the mind: Conceptual issues in contemporary psychometrics. Cambridge: Cambridge University Press.

Borsboom, D. (2006). The attack of the psychometricians. Psychometrika, 71, 425–440.

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.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81–105.

Cervone, D. (2005). Personality architecture: Within-person structures and processes. Annual Review of Psychology, 56, 423–452.

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

Hacking, I. (1999). The social construction of what? Cambridge, MA: Harvard University Press.

Haig, B.D. (2005a). Exploratory factor analysis, theory generation, and scientific method. Multivariate Behavioral Research, 40, 303–329.

Haig, B.D. (2005b). An abductive theory of scientific method. Psychological Methods, 10, 371–388.

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

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

Hartman, C.A., Hox, J., Mellenbergh, G.J., Boyle, M.H., Offord, D.R., Racine, Y., et al. (2001). DSM-IV internal construct validity: When a taxonomy meets data. Journal of Child Psychology and Psychiatry, 42, 817–836.

Heinen, T. (1996). Latent class and discrete latent trait models: Similarities and differences. Thousand Oaks: Sage.

Howell, R.D., Breivik, E., & Wilcox, J.B. (2007a). Reconsidering formative measurement. Psychological Methods, 12, 201–218.

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

Keller, F., & Kempf, W. (1997). Some latent trait and latent class analyses of the Beck-Depression-Inventory (BDI). In J. Rost & R. Langeheine (Eds.), Applications of latent trait and latent class models in the social sciences (pp. 314–323). Münster: Waxmann Verlag.

Lacasse, J.R., & Leo, J. (2005). Serotonin and depression: A disconnect between the advertisements and the scientific literature. PloS Medicine, 2, 1211–1216.

Lazarsfeld, P.F., & Henry, N.W. (1968). Latent structure analysis. Boston: Houghton Mifflin.

Lesch, K.P., Bengel, D., Heils, A., Sabol, S.Z., Greenberg, B.D., Petri, S., et al. (1996). Association of anxiety-related traits with a polymorphism the serotonin transporter gene regulatory region. Science, 274, 1527–1531.

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

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

Molenaar, P.C.M. (2003). State space techniques in structural equation modeling: transformation of latent variables in and out of latent variable models. Retrieved July 3, 2008, from http://www.hhdev.psu.edu/hdfs/faculty/pubs/StateSpaceTechniques.pdf.

Molenaar, P.C.M. (2005). A manifesto on psychology as ideographic science: bringing the person back into scientific psychology, this time forever. Measurement, 2, 201–218.

Molenaar, P.C.M., Huizenga, H.M., & Nesselroade, J.R. (2003). The relationship between the structure of interindividual and intraindividual variability: A theoretical and empirical vindication of developmental systems theory. In U.M. Staudinger & U. Lindenberger (Eds.), Understanding human development: Dialogues with lifespan psychology (pp. 339–360). Dordrecht: Kluwer Academic Publishers.

Middeldorp, C.M., de Geus, E.J.C., Beem, A.L., Lakenberg, N., Hottenga, J., et al. (2007). Family-based association analyses between the serotonin transporter gene plomorphism (5-HTTLPR) and neuroticism, anxiety and depression. Behavior Genetics, 37, 294–301.

Morton, J., & Frith, U. (2004). Understanding developmental disorders: A causal modeling approach. Oxford: Blackwell.

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

Sijtsma, K. (1998). Methodology review: Nonparametric IRT approaches to the analysis of dichotomous item scores. Applied Psychological Measurement, 22, 3–31.

Smith, L.B., & Thelen, E. (1994). A dynamic systems approach to development. Cambridge, MA: MIT Press.

Solomon, A., Haaga, D.A.F., & Arnow, B.A. (2001). Is clinical depression distinct from subthreshold depressive symptoms? A review of the continuity issue in depression research. The Journal of Nervous and Mental Disease, 189, 498–506.

Sullivan, P.F., Kessler, R.C., & Kendler, K.S. (1998). Latent class analysis of lifetime depressive symptoms in the national comorbidity survey. American Journal of Psychiatry, 155, 1398–406.

Timmerman, M.E. (2006). Multilevel component analysis. British Journal of Mathematical and Statistical Psychology, 59, 301–320.

Timmerman, M.E., & Kiers, H.A.L. (2003). Four simultaneous component models of multivariate time series from more than one subject to model intraindividual and interindividual differences. Psychometrika, 86, 105–122.

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

Zubenko, G.S., Highers, H.B. III, Stiffler, J., Zubenko, W.N., & Kaplan, B.B. (2002). D2S944 identifies a likely susceptibility locus for recurrent, early onset, major depression in women. Molecular Psychiatry, 7, 460–467.