RETHINKING THE

DSM

A Psychological Perspective

EDITED BY LARRY E. BEUTLER
AND MARY L. MALIK

AMERICAN PSYCHOLOGICAL ASSOCIATION
WASHINGTON, DC

9

SIMPLIFYING DIAGNOSIS USING A PROTOTYPE-MATCHING APPROACH: IMPLICATIONS FOR THE NEXT EDITION OF THE DSM

DREW WESTEN, AMY KEGLEY HEIM, KATE MORRISON, MARCUS PATTERSON, AND LAURA CAMPBELL

Accurate diagnosis is crucial for both clinical work and research. In the last three decades, major strides forward have been made in the development of diagnostic categories and criteria that can improve research into etiology, prognosis, and treatment response. Yet problems persist, as evidenced in debates about the validity of various diagnoses (e.g., atypical depression), the criteria for diagnoses (e.g., the number of weeks of continuous depression required for a diagnosis of major depression, or the number of binges per week required for a diagnosis of bulimia nervosa), and the separation of clinical syndromes from personality syndromes known to predispose individuals to certain disorders (e.g., major depression and borderline personality disorder). Equally important are concerns by clinicians that the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM–IV*; American Psychiatric Association, 1994) is cumbersome to apply and not tied closely enough to clinical decision-making.

In this chapter we address some of these problems and suggest a strategy for generating diagnostic groupings empirically and diagnosing patients clinically. Throughout, our focus is twofold: how to use empirical strategies to develop a valid classification system and how to keep that classification system close to clinical reality and readily usable by clinicians. Classification systems always exist for a purpose. The aim of classification in psychopathology is not only to "carve nature at its joints" but also to guide clinical observation and treatment. Thus, an optimal classification system is one that is not only nature friendly but also user friendly.

HOW WELL DO THE CURRENT DIAGNOSTIC CATEGORIES WORK?

In this first section we address the question of how well the current diagnostic categories in *DSM–IV* fulfill the functions of carving nature at its joints and facilitating practice and research. We begin by briefly describing a series of problems that recur across diagnoses, and then we illustrate these problems by considering the evolution of three major classes of Axis I syndromes: mood, anxiety, and psychotic disorders.

Common Problems Across Diagnoses

Every decision made in creating a diagnostic classification system has its costs and benefits. The *DSM–IV*, like any evolved "organism," represents not only adaptations of the recent past but also older adaptations that constrain future ones. Categorical diagnosis, for example, was an "adaptation" of a disease model of illness made by the pioneering psychiatric taxonomists of a century ago that remains a part of the nucleus of the current diagnostic system.

DSM–IV represents the best of psychiatric diagnostic thinking as of the early 1990s. Since that time, several problems have come to the fore that suggest the need for rethinking some old and newer adaptations. Here we focus on six central issues.

First is the question of whether psychopathology should be classified categorically, as discrete syndromes; dimensionally, as continua; or both. For example, major depression could be on a continuum with less severe depression, but it could also represent a biologically distinct syndrome, characterized, for example, by a disturbance of the hypothalamic–pituitary axis, which only emerges when the severity of depression crosses a threshold. Alternatively, depression could be represented as a continuous dimension, with a relatively arbitrary cutoff defined as major depression, much as physicians diagnose and treat high blood pressure when a dimensional measure of blood pressure exceeds an arbitrary threshold.

The most persuasive arguments for dimensional diagnosis have come from the literature on personality disorders (e.g., Frances & Widiger, 1986; Livesley, Schroeder, Jackson, & Jang, 1994; Widiger, 1993), largely because of the rich tradition of dimensional trait measurement in personality psychology. However, as we will see, research on the range of Axis I disorders has begun to raise many of the same questions, even in disorders such as schizophrenia, which are clearly discontinuous from normal functioning and hence seemingly most amenable to categorical diagnosis.

A second and related problem is the existence of "subclinical" phenomena. A growing body of research suggests that subclinical cases are at least as prevalent as clinical cases of many if not most disorders. For example, roughly 60% of patients treated for enduring, maladaptive personality patterns cannot be diagnosed on Axis II (Westen & Arkowitz-Westen, 1998). Nevertheless, these patients suffer from clinically significant problems recognized and treated by clinicians of all theoretical orientations, ranging from difficulties regulating self-esteem and problems with assertiveness or aggression to repetitive interpersonal patterns that interfere with relational functioning and satisfaction.

Third is the problem of comorbidity. The virtual explosion of research on comorbidity has been, at least in part, a byproduct of the more careful and systematic delineation of diagnostic categories and criteria in the last three editions of the DSM. Whether this represents incremental knowledge about psychopathology or simply the amassing of data documenting the problems of distinguishing disorders best conceptualized as fuzzy sets is unclear. A classic case is the literature on comorbidity of personality disorders and virtually every Axis I syndrome, such as depression, eating disorders, or panic. As suggested by theoretical traditions as disparate as psychoanalysis (e.g., Kernberg, 1984), trait psychology (e.g., Eysenck, 1994), and Millon's (1990) evolutionary social learning approach, the roughly 50% comorbidity of virtually every Axis I disorder with Axis II pathology of some sort likely reflects the fact that most forms of psychopathology (e.g., anxiety disorders) emerge from personality vulnerabilities (e.g., high neuroticism, trait anxiety, or borderline personality disorder) that can often be observed long before the development of the first Axis I episode. (On the longitudinal prediction of Axis I symptoms in early adulthood from Axis II symptoms in adolescence, see Johnson et al., 1999). Of particular importance in this regard are numerous studies showing that the presence of multiple Axis I syndromes in a patient is essentially a proxy measure for the presence of Axis II pathology, with an exponential rise in likelihood of Axis II diagnosis with each additional Axis I diagnosis (e.g., Lewinsohn, Rohde, Seeley, & Klein, 1997; Newman, Moffitt, Caspi, & Silva, 1998).

Fourth is the proliferation of mixed, atypical, and not otherwise specified categories with each successive revision of the DSM. Careful delimitation of virtually every Axis I category has brought with it the recognition

of border cases that require a new categorical diagnosis, because strict adherence to a set of specific, and often quasi-arbitrary, diagnostic criteria necessarily leads to nondiagnosis of subclinical or atypical syndromes.

A fifth problem is the difficulty researchers have had in reproducing many current diagnostic categories and criterion sets using statistical aggregation procedures such as factor analysis, cluster analysis, latent class analysis, and structural equation modeling (e.g., Brown, Chorpita, & Barlow, 1998). This could, of course, reflect problems in the items, samples, or algorithms used to develop classifications (algorithms that sometimes have difficulty reproducing known structures in Monte Carlo simulations; see, e.g., Waller & Meehl, 1998). Alternatively (or, more likely, additionally), the problem may lie in the way diagnostic groupings have evolved, namely, through gradual clinical and empirical refinements of distinctions first made more with less precise tools by pioneering taxonomists such as Kraepelin, Bleuler, and Schneider.

A final problem is the difficulty in implementing the fine distinctions made in *DSM–IV* in clinical practice. As we discuss later, the current method of combining criteria to diagnose a patient—namely, counting criteria and subcriteria—often reflects arbitrary cutoffs and algorithms and is too cumbersome to be used by clinicians in everyday practice. With several hundred criteria for several dozen disorders, actually following the decision rules outlined in *DSM–IV* would be incredibly time-consuming. Not surprisingly, clinicians usually do not make diagnoses this way (e.g., Jampala, Sierles, & Taylor, 1988).

More important, perhaps, is the question of whether the procedure specified in the *DSM–IV* for making diagnoses has any advantage in terms of predictive validity (e.g., accurately predicting prognosis or treatment response) over the intuitive prototype-matching process clinicians are more likely to use if left to their own devices. The *DSM–IV* procedure often seems to clinicians (the primary consumers of the diagnostic manual) artificial or irrelevant to treatment decisions. In clinical practice, we suspect most clinicians get the "gist" of the patient's pathology (e.g., is the patient depressed, having trouble sleeping, losing weight, and thinking about suicide) and diagnose accordingly (the patient has major depression), whether or not the patient has three, four, or five of the criteria required for the diagnosis. The real question, then, pertains to the incremental validity of counting symptoms over prototype matching—particularly if clinicians were to learn criterion sets as prototypes rather than as sets of isolated symptoms to be counted.

The symptom-counting algorithm for arriving at a diagnosis was developed in an effort to increase reliability of diagnosis in the move from the 2nd to the 3rd editions of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM–II*; American Psychiatric Association, 1968; *DSM–III*; American Psychiatric Association, 1980). An unintended consequence,

however, was an increasing disconnect between clinical and research diagnoses and a growing antagonism between researchers and clinicians. Researchers tend to view clinicians as sloppy diagnosticians who do not use structured interviews (which, in fact, now provide the only way to make reliable diagnoses if accurate diagnosis requires knowing exactly how many criteria out of four, five, six, or more per disorder the patient meets). Clinicians, in contrast, often view researchers as symptomatic bean counters who require precise answers to questions (such as exactly how long and how often a patient has had symptoms of major depression) that patients often cannot answer accurately and that may not even be relevant to treatment planning. Indeed, another unintended consequence of greater precision in diagnosis has been the loss of information about the large percentage of patients who fall just short of one diagnosis or another, because they are routinely excluded from studies of psychopathology and treatment. We know very little, for example, about the treatment of garden-variety depression, because most outcome studies focus on major depression (see Morrison & Westen, 2000).

Having counted the symptoms of *DSM–IV*, we now examine the way they manifest in three kinds of disorder: mood, anxiety, and psychotic disorders. We conclude our "case formulation" with recommendations for treatment of the *DSM*.

Classifying and Diagnosing Depression

Mood disorders are the most commonly diagnosed of the Axis I disorders, yet their specific nature continues to be an object of considerable debate (Chen, Eaton, Gallo, Nestadt, & Crum, 2000). The 1st edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM–I*; American Psychiatric Association, 1952) proposed three broad categories of mental illness: psychoses, personality disorders, and psychoneuroses. Each of these categories included depressive phenomena (e.g., the psychoses included manic–depressive reactions and psychotic depressive reactions). A central feature of both *DSM–I* and *DSM–II* (1968) was a reliance on etiological theories as a basis for taxonomic organization, including organization of the mood disorders. For example, depression resulting from early childhood experiences constituted a subcategory of mood disorder. Debate later ensued over the appropriateness of basing the definition of psychological disorders on etiological theories that often had little basis in research (see Skinner, 1986).

DSM–III marked a shift toward a Schneiderian taxonomic approach based exclusively on directly observable phenomena. This shift occurred in response to increasing demands for a classification system capable of yielding more reliable and empirically valid diagnoses. On the road to *DSM–III*, researchers developed specific criteria and interviews to assess

those criteria with the aim of providing an empirical basis for distinctions among subtypes of depression (Feighner, Robins, Guze, Woodruff, & Winokur, 1972; Spitzer, Endicott, & Robins, 1978). The resulting criteria included symptom, duration, and exclusion specifications that added precision to diagnoses and dramatically increased their reliability across treatment and research sites. The *DSM–III* also deleted all personality disorder diagnoses characterized primarily by depression, thereby confining depressive symptomatology to Axis I. One result of this has been continuing debate about the existence of a depressive personality style that meets all the general criteria for a personality disorder outlined in the introductory material to Axis II in *DSM–IV* (e.g., Klein, 1999; Phillips & Gunderson, 1999; Westen & Shedler, 1999b).

Attempts to clarify the classification of mood disorders empirically also led to application of a variety of statistical aggregation procedures, including factor analysis, cluster analysis, and, more recently, structural equation modeling. Early studies utilizing factor analysis (e.g., Mendels & Cochrane, 1968) corroborated the inclusion of some version of the endogenous–reactive distinction that appeared in the first two editions of the *DSM*. Many researchers then turned to cluster analysis as a way of producing categorical diagnoses (see, e.g., Everitt, 1979; Fleiss, Lawlor, Platman, & Fieve, 1971; Grove & Andreasen, 1986). Whereas factor analysis groups symptoms together on the basis of their co-occurrence in the population of interest, cluster analysis groups patients together on the basis of similarity of their profiles on the criteria of interest. Thus, cluster analysis appeared potentially more appropriate for uncovering symptoms that may covary in different ways in different groups (if such true taxa, or discrete groups, exist). In a number of studies, researchers used cluster analysis to try to group patients with depression to see if they could uncover syndromes (e.g., Andreasen & Grove, 1982; Everitt, 1979).

Despite the initial promise of this method, it eventually fell into disuse. One reason had to do with the varied success of efforts to validate empirically derived clusters on the basis of treatment response (Paykel, 1971; Raskin & Crook, 1976). Perhaps the main reason for the decline of cluster analysis, however, was the lack of replicability of cluster solutions. In their review of 11 cluster-analytic studies of depression, Blashfield and Morey (1979) found that all studies isolated an endogenous subtype but that little agreement emerged on any other subtypes (although several yielded an anxious subtype as well).

Related to the problem of replicability were a number of other issues that led to dampened enthusiasm for cluster analysis. Cluster analysis always yields cluster solutions, even if no orderly or coherent classification scheme is inherent in the data, and different cluster algorithms frequently yield disparate groupings for the data set (Blashfield & Morey, 1979; Everitt, 1979). Furthermore, most researchers using cluster analysis tended
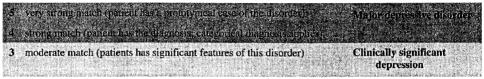
to assume that the data were best understood categorically rather than dimensionally, leading to the problem of subclinical and atypical diagnoses encountered currently in *DSM–IV* (Grove & Andreasen, 1986).

DSM–IV introduced several changes to the classification of mood disorders, such as the inclusion of an atypical subtype, characterized by increase in appetite, weight gain, hypersomnia, and psychomotor agitation (and preferential response to MAO inhibitors, although this is not a diagnostic criterion). As was the case with previous cluster-analytic studies, investigations of this new category using statistical aggregation techniques such as latent class analysis have yielded some support, although different analyses have suggested different criteria (Kendler et al., 1996; Sullivan, Kessler, & Kendler, 1998).

Two major problems with the diagnostic system for mood disorders are subthreshold cases and comorbidity. The extent to which the current categories provide a way to diagnose most patients who present clinically with depression is a matter of some debate, with some data suggesting that most cases can be encompassed using current criteria and other data suggesting that subthreshold cases may be common (Barrett, Barrett, Oxman, & Gerber, 1988; Keller et al., 1995). Depression is also highly comorbid with a variety of disorders (see, e.g., Kessler et al., 1994), raising the question of whether it is a discrete syndrome. In an attempt to deal with the high comorbidity of mild forms of anxiety and depression, a category of mixed anxiety-depression (MAD) was proposed during the development of *DSM–IV*. The common co-occurrence of major depressive disorder and dysthymic disorder also raises the issue of whether two discrete mood disorders are present, whether "double depression" is actually a category unto itself, whether depression really falls on a continuum, or whether the vulnerability for major depression seen in 'double depression is more readily understood in terms of personality and temperament (e.g., Keller & Lavori, 1984). Figure 9.1 is a prototype approach to diagnosing depression.

Classifying and Diagnosing Anxiety

The systematic classification of anxiety disorders has a relatively short history when compared with depression and schizophrenia. Freud's descrip-

5	very strong match (patient has a prototypical case of the disorder)	**Major depressive disorder**
4	strong match (patient has the diagnosis, categorical diagnosis applies)	
3	moderate match (patients has significant features of this disorder)	**Clinically significant depression**
2	slight match (patient has minor features of this disorder)	
1	no match	

Figure 9.1. A Prototype Approach to Diagnosing Depression

tion of "anxiety-neurosis" paved the way for the syndromal classification of anxiety disorders used today. As noted above, DSM–I and DSM–II used the term *neurosis* as a superordinate diagnostic term. In addition to deleting the ties to psychodynamic etiological theories and turning to a more descriptive diagnostic system, DSM–III assigned greater importance and specificity to the anxiety disorders by creating a separate category. The *anxiety neuroses* became *panic disorder* and *generalized anxiety disorder* (GAD), whereas the *phobic neuroses* became *agoraphobia, social phobia,* and *simple phobia.*

A key change in the 3rd revised edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM–III–R; American Psychiatric Association, 1987) involved a shift in emphasis from agoraphobic behavior to panic, with the designation of panic disorder with agoraphobia replacing the previous diagnosis of agoraphobia with panic attacks. DSM–III–R also eliminated much of the hierarchical ordering of diagnoses, which had, for example, relegated GAD to a residual status not to be diagnosed in the presence of other diagnosable disorders. In DSM–III–R, GAD became a nonresidual diagnostic category with its own defining feature of excessive worry that was more diffuse than worry associated with other Axis I anxiety disorders. In DSM–IV, GAD was retained as a diagnosis after some debate. Changes to the GAD diagnosis in the most recent edition of the DSM included an emphasis on the uncontrollability of the worry process and removal of associated symptoms that reflected autonomic hyperactivity rather than motor tension or vigilance.

The current classification system for anxiety disorders suffers from many of the same problems as the classification of depression, notably subclinical cases and comorbidity. Numerous studies have documented the prevalence of cases of subthreshold anxiety disorders (e.g., Zinbarg, Barlow, Liebowitz, & Street, 1994). Olfson et al. (1996) surveyed 1,001 primary care patients in a large health maintenance organization and found that 30% of them met their criteria for subthreshold symptoms of a variety of Axis I disorders; the most prevalent diagnoses were panic (11%), depression (9%), and anxiety (7%).

Comorbidity of anxiety and other disorders is also extremely high, particularly anxiety and depression (see Clark, 1989; Kessler et al., 1996). Clark and Watson (1991) found that many items on measures of anxiety and depression symptoms do not discriminate between individuals with anxious and depressive symptoms. Indeed, instruments such as the Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), Spielberger State–Trait Anxiety Inventory (Spielberger, Gorsuch, & Lushene, 1970), and Neuroticism factor of the NEO–PI–R (McCrae & Costa, 1990) tend to intercorrelate upwards of .60 in both clinical and nonclinical populations.

As noted below, one way of resolving this problem is to look for a combination of general and specific factors that both unite and distinguish

anxiety and depression, much as researchers in the field of intelligence have distinguished g factors (general intelligence) and s factors (specific intellectual factors). Another proposal is to create a mixed anxiety and depression category (MAD), which is included in DSM–IV as a diagnosis needing further study. Indeed, many patients with subclinical levels of co-occurring anxious and depressive symptoms do not meet criteria for any DSM–IV anxiety or mood disorder (Katon & Roy-Byrne, 1991). Far from being an anomalous group, patients fitting the MAD profile frequently present for treatment and overuse primary care services (Barrett et al., 1988). In a field trial designed to investigate the MAD diagnosis, Zinbarg et al. (1994) found that patients presenting with subthreshold anxious or depressed symptoms were "at least as common as patients with several of the already established anxiety and mood disorders in each of the seven sites."

Research using factor analysis and structural equation modeling has provided both support for and challenges to the current classification of anxiety disorders. Clark and Watson (1991) found that although anxiety and depression share a common negative affect factor, depression is uniquely associated with low positive affect, and anxiety is specifically associated with autonomic arousal. Zinbarg and Barlow (1996) found that a single, higher order factor of negative affect distinguished those diagnosed with an anxiety or mood disorder from those who received no diagnosis, but they also found a number of lower order factors that differentiated between patients meeting criteria for the distinct categories in the anxiety disorders section of the DSM–IV. Using structural equation modeling, Brown et al. (1998) similarly found a shared negative affect factor as well as a latent factor of autonomic arousal that was differentially related to panic disorder (sympathetic hyperarousal) and GAD (sympathetic inhibition).

Recent data suggest, however, that the current classification system may create distinctions that are so fine-grained that they miss broader, underlying characteristics that may be as or more important to understanding and treating anxiety disorders (Brown et al., 1998; Zinbarg & Barlow, 1996). Studies by Barlow, Clark, and others suggest that personality dispositions, notably negative affect, may put individuals at risk for a variety of conditions. Indeed, the current taxonomy of anxiety disorders focuses only minimally on the relationship between anxiety and personality. Although personality constellations were implicit in the early conceptualization of anxiety as anxiety neurosis, DSM–IV focuses on anxiety states rather than traits that may predispose to these states. It is interesting that although DSM–IV includes depressive personality disorder as a diagnosis worthy of further study, no analogous category has been proposed for anxious personality (although the avoidant diagnosis has evolved in that direction over multiple editions of the DSM). A much more complex but likely possibility is that a mapping of trait anxiety onto state anxiety as a

diathesis may be too simple, given data suggesting that patients with co-morbid anxiety and depressive symptoms are likely to have a variety of personality disorders (e.g., Newman et al., 1998). Genetic factors as well as environmental factors such as sexual abuse can create diatheses for multiple negative affect states, including major depression, minor depression, panic, generalized anxiety, and posttraumatic phenomena (see., e.g., Barlow, 2002).

Classifying and Diagnosing Schizophrenia

Although a syndrome akin to schizophrenia was identified early in the 19th century (see Morel, 1852; Pinel, 1801/1962), the first systematic conceptualizations of the disorder emerged around a century ago, in the work of Emil Kraepelin (1898) and Eugene Bleuler (1911/1950). Kraepelin distinguished *dementia praecox* (later renamed *schizophrenia* by Bleuler) and manic-depressive illness and linked phenotypically diverse forms of schizophrenia (i.e., hebephrenic, catatonic, and paranoid) on the basis of what he regarded as their shared underlying features: early onset, a deteriorating course, and poor prognosis.

Bleuler viewed the deficit shared by these diverse phenotypic expressions as a disconnection among emotions, thoughts, and behaviors and introduced the term *schizophrenia* (literally "split brain") to represent this view. Unlike Kraepelin, Bleuler emphasized signs and symptoms over outcome and course (Andreasen & Carpenter, 1993). Bleuler viewed the hallucinations and delusions essential to the Kraepelinian concept of the disorder as secondary to four key symptoms: affective disturbance, autism, ambivalence, and associational looseness. Bleuler's definition cast a wider net than Kraepelin's and resulted in a substantial increase in the number of patients diagnosed with schizophrenia (especially in the United States, where Bleuler's views held sway, as opposed to the Kraepelinian view, which was favored in Europe).

Bleuler's view was reflected in both *DSM-I* and *DSM-II*, which contained brief, relatively vague definitions of schizophrenia with no operational criteria (Tsuang, Stone, & Faraone, 2000). A "neo-Kraepelinian revolution" began in the 1960s and early 1970s (Lenzenweger, 1999), as Kraepelin's narrower, tighter approach to defining schizophrenia lent itself well to improving diagnostic reliability and validity. The narrower definition of schizophrenia resulted in a significant reduction in diagnoses of schizophrenia, so much so that within 5 years of the publication of *DSM-III*, diagnoses of schizophrenia in a large university hospital decreased by 50% (Loranger, 1990). The move to a more Kraepelinian definition also resulted in a greater emphasis on blatantly psychotic symptoms (e.g., hallucinations) and a commensurate de-emphasis of negative symptoms (e.g.,

affective flattening, alogia; Andreasen & Carpenter, 1993; Lenzenweger, 1999).

DSM–III also saw the inclusion of Kraepelin's subtypes (catatonic; hebephrenic, renamed *disorganized*; and paranoid), although the heterogeneity of symptoms seen in schizophrenia has led to other classification schemes, notably the distinction among positive, negative, and disorganized symptoms (Strauss, Carpenter, & Bartko, 1974). Researchers have also attempted to subtype schizophrenia on the basis of prognosis, adaptive functioning, and biological measures, such as electrodermal response (Lencz, Raine, & Sheard, 1996) and EEG (John et al., 1994), and will no doubt do so in the future using functional neuroimaging.

The move from *DSM–III* to *DSM–III–R* resulted in relatively minor changes to the schizophrenia category, such as elimination of the age limit for diagnosis (age 45) and changes in the criteria for paranoid schizophrenia (see Kendler, Spitzer, & Williams, 1989). *DSM–IV* saw more extensive changes, perhaps the most important being the placement of schizophrenia within the broader category of psychotic disorders, with the aim of facilitating differential diagnosis. Other significant changes to the schizophrenia diagnosis made it both broader (increased emphasis on negative and disorganized symptoms) and narrower (increase in threshold for duration of active symptoms from 1 week to 1 month; American Psychiatric Association, 1998).

Despite these efforts to revise the *DSM* nomenclature, critics continue to voice concern over several aspects of the schizophrenia diagnosis. Two prominent concerns include problems with the categorical approach and the absence of biological factors among the diagnostic criteria. A major problem with categorical diagnosis has been the difficulty in distinguishing a clear schizophrenic syndrome from other psychotic syndromes (such as schizoaffective disorder) and the correlative problem of reproducing the subtypes in the diagnostic manual using statistical aggregation procedures such as cluster and latent class analyses (e.g., Dollfus et al., 1996; John et al., 1994; Kendler, Karkowski-Shuman, et al., 1997; Kendler, Karkowski, & Walsh, 1998; Lencz et al., 1996; Sham, Castle, Wessely, Farmer, & Murray, 1996; Van der Does, Dingemans, Linszen, Nugter, & Scholte, 1995). As with anxiety and mood disorders, reviewed above, many researchers have argued that a categorical classification system creates artificial boundaries and does not describe significant numbers of patients with schizophrenia (e.g., Van der Does, Dingemans, Linszen, Nugter, & Scholte, 1993). In response, several researchers have proposed dimensional diagnostic systems, one of which is included in Appendix B of *DSM–IV* for further study (American Psychiatric Association, 1994). This system, which is particularly promising, has clinicians rate the extent to which the patient has positive symptoms, disorganized symptoms, and negative symp-

toms, using a 4-point severity scale (from *absent* to *severe*). Lenzenweger (1999) suggested adding a fourth dimension, premorbid social functioning.

A second concern with the diagnosis of schizophrenia as defined in *DSM–IV* involves the absence of biological and neuropsychological abnormalities among the diagnostic criteria (e.g., Lenzenweger, 1999; Tsuang et al., 2000). Tsuang et al. argued that the exclusive focus on phenomenological criteria may have been appropriate in a prior era in which data on pathophysiology were lacking, however, amassing data support a diagnosis first suggested by Meehl (1962), *schizotaxia*, that can be assessed using markers such as eye tracking and structural brain abnormalities. They suggested a diagnosis of schizotaxia (characterized in large measure by negative symptoms) with and without psychosis, much as we currently distinguish psychotic and nonpsychotic depression. According to Tsuang and colleagues (2000), psychosis may be the fever of severe mental disorders—a relatively nonspecific symptom seen in a range of neuropsychiatric conditions such as schizophrenia, bipolar disorder, dementia, and Huntington's disease.

Andreasen (1999) has offered an alternative, neo-Bleulerian approach, a unified theory of schizophrenia that links the multiple phenotypic expressions of the disorder to an underlying "misconnection" of neural circuits. This misconnection produces generalized cognitive dysfunction or "dysmetria" (Andreasen, Paradiso, & O'Leary, 1998). On the basis of functional neuroimaging studies, Andreasen argued that schizophrenia is a neurodevelopmental disorder involving faulty wiring in a circuit running from the frontal lobes through the thalamus and cerebellum.

CLASSIFYING PSYCHOPATHOLOGY AND IDENTIFYING CASES: HOW SHOULD WE CREATE AND IMPLEMENT A CLASSIFICATION SYSTEM?

In the preceding sections, we reviewed evolving efforts to classify the forms of psychopathology for which we have the best data. The review suggests both the enormous progress made in the 20th century and the enormous task that remains ahead in the current century in resolving basic taxonomic issues. In this section we take a step back from specific diagnoses to ask two questions: How should we create a classification system, and how should we diagnose patients after we have settled on an appropriate set of categories or dimensions?

It is important to note that the questions of *how to create a classification system* and *how to identify cases once such a system is in place* are in fact distinct. An important way these processes differ is in the number of criteria used (Sokol, 1974). Developing a classification system means using all available data to select the variables (criteria) that best distinguish

patients with different forms of psychopathology. Whether this is done impressionistically, as when Kraepelin (1898) listed all the patients he had observed on note cards and sorted them into piles based on similarity of their symptoms, or statistically, using procedures such as factor and cluster analyses, the process of creating a diagnostic system requires inclusion of many more variables than the process of diagnosing a patient after the most discriminating variables have been identified. The reason for this difference is apparent if one considers the process of constructing a psychological test. In test construction, a researcher always does well initially to maximize content validity, by including items that comprehensively cover the domain in question, using all available clinical, theoretical, and empirical knowledge to develop items that sample the domain as exhaustively as possible. After applying this item set to multiple samples, the researcher is then in a position to examine the factor structure of the item set and the intercorrelations of the items and eliminate items, many of which will be redundant or minimally predictive of criterion measures theoretically related to the construct.

After a researcher has developed an instrument, or a committee has developed a diagnostic system, the next question is how to use that "instrument" to diagnose individual cases. (On the analogy between diagnostic systems and psychological tests, see Livesley & Jackson, 1992.) The method specified for making diagnoses in the shift from *DSM–II* to *DSM–III* and *DSM–III–R* was for the clinician or interviewer to make a dichotomous forced-choice decision about each criterion as present or absent and then to count the number present, sometimes following algorithms specifying a certain number of subcriteria from Criteria A, B, and so forth). As we show, this is only one of a range of options for combining criteria to make a diagnosis, and one that may have been a useful first step toward a simpler and psychometrically more useful algorithm.

Central Questions in Developing a Classification System

In creating a diagnostic system, its framers have to make a number of key decisions, each of which should be explicitly considered. Here we briefly outline some of the most important (see Westen, 1999).

The first issue concerns the content of the variable list or item set to be used to distinguish disorders or dimensions. The variables used in defining syndromes historically have been derived from two leading contenders, phenomenology (symptomatology) and etiology. At different times, different authors have called for greater attention to one or the other. *DSM–II* was filled with etiological statements, few of which had a firm empirical basis. In large part as a response to the growing theoretical pluralism in the field by the 1970s (with psychoanalysis no longer dominant), *DSM–III* eliminated all such statements in the criterion sets and moved to a

focus on symptomatology. Researchers such as Tsuang and colleagues (2000) have suggested that, for disorders in which etiological data are accumulating, we begin to reconsider etiological criteria.

This situation in many ways parallels developments in both medicine and biological taxonomy. In medicine, disorders often begin with descriptive, phenomenological classification until etiological agents are better understood (e.g., in the evolution of the diagnosis of AIDS). In biology, Linne developed a classification of species based on overt features, which was later challenged by, and ultimately integrated with, evolutionary theories (comparable to etiology in psychiatry; Skinner, 1981).

Although factor- and cluster-analytic studies in psychiatric nosology have typically used exclusively phenomenological item sets, nothing about these statistical techniques requires one form of data or another. Indeed, some recent studies of schizophrenia have applied cluster-analytic techniques to psychophysiological data (e.g., Tsai et al., 1998). Note that etiological agents need not be genetic; environmental factors ranging from exposure to viruses and malnutrition, to sexual abuse and emotional criticism, have been linked to various disorders. To date, it may well be worth assembling a list of genetic and environmental variables linked to one or more psychiatric conditions and include them in studies aimed at classifying psychological disorders. In our own cluster- and factor-analytic work on personality disorders (Westen & Chang, 2000; Westen & Shedler, 1999a, 1999b), we are experimenting with mixed models, in which the units of analysis include both personality variables and etiological variables.

A second key question concerns the level of inference required in rating the variables used for aggregation. An assumption made by most researchers is that the lower the level of inference, the more valid, reliable, and useful the data are likely to be. This has inspired the architects of successive editions of the DSM to become progressively more specific in their diagnostic criteria. For example, the diagnostic criteria for paranoid personality disorder are now all behavioral examples of ways of being distrustful, which could be collapsed into a single, higher order statement such as "is distrustful of people and preoccupied with fears of betrayal, maltreatment, and so on."

As we discuss below, the preference for minimal inference may not be well founded. In our own research, we are finding substantially higher correlations between interview and clinician diagnoses of personality disorders (median correlations around .80) than previously reported in the literature, using the least structured, most highly inferential rating procedure currently available for assessing personality disorders. (We use a Q-sort based on either the clinician's knowledge of the patient over the course of multiple sessions, or an interviewer's inferences regarding the 200 items that constitute the Q-sort after completing an interview focused primarily on narratives rather than questions about diagnostic criteria; Westen &

Muderrisoglu, 2001). Some of the most reliable and valid personality questionnaires, such as the NEO–PI–R (McCrae & Costa, 1990), require individuals to make a number of generalizations about themselves that go well beyond a descriptive behavioral level, such as whether they tend to be anxious, which is a highly subjective and inferential question, given that the metric can mean different things to different people (how much anxiety is "moderate"?).

A third question related to the development of a classification system concerns the observer or rater: Whose observations are the most reliable and valid for rating the variables to be entered into a factor analysis, cluster analysis, or other structural model? The implicit assumption of the early taxonomists such as Kraepelin and Bleuler was that skilled clinicians with many years of experience were in the best position to try to discriminate types of patients, although they lacked the kinds of statistical procedures for doing so that we have today. The implicit assumption of most personality psychologists and psychopathologists is that self-observation, perhaps filtered through the eyes of a research assistant administering a structured interview, is sufficient or even optimal. This assumption should be carefully considered, given the problems with self-reports, including lack of expert knowledge, the problems of using explicit (self-report) measures to report an implicit processes, and self-deception (see, e.g., Shedler, Mayman, & Manis, 1993; Westen, 1995, 1997). An alternative procedure that we have been pursuing is to use the observations of experienced clinicians, not to offer their best guesses about how patients should be categorized, but simply to describe patients using a psychometrically sound measure, which then provides data for statistical aggregation.

A fourth question related to a classification system is whether diagnoses should be categorical, dimensional, some combination of the two, or functional. As described earlier, the issue of categorical versus dimensional diagnosis has received considerable attention over the past two decades, particularly in relation to personality disorders. The basic issues are similar for many Axis I syndromes, for which subclinical presentations are common. Categorical classification is familiar in everyday life (e.g., an object in the office is classified as a chair, not .85 chairlike and .32 sofalike, even though it may bear some resemblance to a sofa) and feels equally "natural" in clinical practice. Categorical diagnosis is efficient and parsimonious, and it renders communication among professionals relatively easy (e.g., "the patient suffers from major depression," vs. "the patient has a Hamilton depression score of 22"). Furthermore, some psychiatric disorders, like some medical disorders, appear to be taxonic—that is, to represent discrete categories—or at the very least to be something other than located along a continuum with healthy functioning, as suggested by data on the genetic epidemiology of syndromes such as schizophrenia and bipolar disorder, for which a genetic diathesis is virtually a *sine qua non*.

By contrast, dimensional systems tend to be complex and cumbersome. Their advantages, however, are twofold. First, they tend to fit the data better for most disorders and prevent many of the problems of comorbidity and creation of atypical, mixed, and not otherwise specified categories that dog the current diagnostic system. This advantage is amplified when the dimensions are selected empirically through procedures such as factor analysis. Second, and related, dimensional systems do not arbitrarily cut continuous variables into dichotomous variables (present–absent) and hence tend to be much more reliable and valid.

It is worth noting an alternative (or perhaps more accurately, a complement) to both categorical and dimensional diagnosis, one that is particularly useful clinically: functional diagnosis (Westen, 1998; Westen & Arkowitz-Westen, 1998). A functional diagnosis is an assessment of how the patient is functioning in each of several central psychological domains. A functional diagnosis is fundamental to case formulation. Rather than asking whether the patient has major depressive disorder (categorical diagnosis) or whether the patient is high on neuroticism (dimensional diagnosis), a functional assessment asks questions such as these: What is the patient's affective functioning like (e.g., is he depressed, anxious)? How effectively can he regulate his emotions (e.g., when he is angry, can he contain it, or does he express it through physical assault or passive aggression)? How is the patient's interpersonal functioning (e.g., do current or long-standing problems interfere with his ability to maintain a job or relationships)?

In all likelihood, our classification systems ultimately have to integrate functional with diagnostic (categorical and dimensional) assessment, particularly for personality. The closest to a functional component of the DSM–IV is the Global Assessment of Functioning (GAF) scale. However, GAF scores provide little insight into which aspects of the patient's functioning are problematic and what psychological processes underlie aspects of both pathological and healthy functioning. As argued elsewhere (Westen, 1998; Westen & Shedler, 2000), a functional assessment of personality is not only quantifiable (e.g., assessing the extent to which the patient has various characteristics that interfere with the capacity to maintain relationships) but also translatable into a personality diagnosis.

A fifth question is whether we should use exploratory or confirmatory procedures in developing a classification system. At the heart of this question are two corollary issues, regarding the role of theory and the extent to which we should begin with familiar classification systems. According to one point of view, a diagnostic system is essentially a theory (Livesley & Jackson, 1992; Skinner, 1981, 1986), which includes propositions about how the criteria that constitute a diagnosis should relate to one another (internal consistency) and how the diagnosis should relate to external criteria (criterion or construct validity). In this view, the crucial question is

how the current classification system compares to another theoretically-derived system. A somewhat different, yet related, approach is to use theory and observation not to construct diagnoses a priori but to select the *item set* to which statistical aggregation procedures can be applied. Theory and clinical judgment again enter into the equation in deciding which factor or cluster solutions to retain.

A final question is how to validate a diagnostic system. In general, there are three broad classes of validating criteria: *internal criteria, external criteria,* and *clinical criteria* (see, e.g., Skinner, 1981, 1986). *Internal criteria* refer to characteristics such as the coherence, nonredundancy (discriminant validity), replicability, and comprehensiveness of a diagnostic system, irrespective of its ability to predict external variables. These internal characteristics are necessary but not sufficient to validate a classification system. The current diagnostic system fails on a number of internal criteria: Although most of its categories are coherent (i.e., they "hang together" as syndromes), many diagnoses are redundant (highly comorbid or lacking in discriminant validity), have proven only variably reproducible through empirical procedures such as cluster and factor analysis, and leave too many patients undiagnosed (borderline and subclinical cases).

With respect to external criteria, the central question is whether the diagnostic categories or dimensions predict theoretically relevant criterion variables (such as prognosis, treatment response, level of adaptation, and etiology; see Livesley & Jackson, 1992; Robins & Guze, 1970). With respect to clinical criteria, the validity of a diagnostic system depends on the extent to which its diagnoses appear faithful to clinical reality, provide clinically useful information, and are practical and user friendly. Although taxonomists may disagree on the relative weight to place on these three types of validating criteria, all three are clearly important in evaluating alternative diagnostic systems.

A Prototype-Matching Approach to Diagnosis

Creating versus applying a diagnostic system are two different enterprises and require very different methods. A central thesis of this chapter is that a painstaking, symptom-counting approach, even more systematic than the current procedure required by *DSM–IV*, is essential for creating a more valid diagnostic system, but that a much more inferential, intuitive, less obsessional approach is optimal for applying that system.

Diagnosing a patient using the symptom-counting approach of *DSM–IV* requires a lengthy, systematic, structured interview that inquires about each of several hundred criteria outlined in Axis I and Axis II. We would argue that this is not optimal for either refinement of the current diagnostic system or for clinical diagnosis. To refine the current system, factor- or cluster-analyzing the current criteria would be of only limited utility be-

cause the criteria have been selected over the last three editions of the DSM precisely to maximize internal consistency and minimize comorbidity. We can learn much about where these efforts have succeeded or failed but very little about how to improve them unless we include alternative criteria. If we are to apply statistical aggregation procedures to any item set, it should include current diagnostic criteria for all DSM–IV disorders as well as at least as many other potential criteria for the constructs included in the DSM–IV and their subclinical variants (e.g., subclinical eating disorder symptoms, such as preoccupation with food preparation or irrational restriction of food intake that produces weight loss not severe enough to warrant a DSM–IV diagnosis of anorexia nervosa). Variables included in efforts to develop diagnoses empirically should also be coded dimensionally unless we have good reason to believe that they are dichotomously distributed in nature, given the loss of validity, reliability, and statistical power that generally occurs when continuous variables are dichotomized.

With respect to diagnosing cases using a refined diagnostic system, the current rules for diagnostic decision making, which require assessing and counting hundreds of criteria, are neither clinically practical nor, one may argue, optimal for making valid and reliable clinical or research diagnoses. We propose, instead, a simple, intuitive approach based on prototype theory.

Prototype Theory and the Current Polythetic Diagnostic Criteria

The most recent editions of the DSM have shifted away from a defining features approach to categorization toward a prototype approach, based in large part on changing views of classification in cognitive psychology (see Cantor & Genero, 1986; Cantor, Smith, French, & Mezzich, 1980; Horowitz et al., 1981; Smith, 1995). Defining features views of categorization suggest that people categorize an object by comparing its features with a list of qualities that are essential (that is, necessarily present). Most concepts used in daily life, however, are fuzzy concepts, whose members bear a family resemblance to one another but do not share a set of necessary and sufficient features (Malt, 1993; Rosch, 1978). Thus, people usually classify objects by matching the similarity of the object to a prototype in memory—that is, to a mental representation of members of the category that has been abstracted across multiple instances, or to a prominent or prototypical exemplar of the category, such as the best example of a patient with borderline personality disorder one has seen previously (or, we suspect, although no one has tested the roles of affect or primacy in prototype matching, a borderline patient seen while one was in training who was the source of particular anxiety, anger, or rescue fantasies).

The shift from monothetic criterion sets (classification based on the presence of a set of singly necessary and sufficient attributes) to polythetic

criterion sets (classification based on the presence of multiple attributes, no one of which is sufficient, and most of which are not necessary for diagnosis) was in fact an attempt to operationalize a prototype view of diagnosis (Frances, 1982; Widiger & Frances, 1985). The architects of the DSM since DSM–II were concerned with several problematic aspects of DSM–II, including two of particular importance here: its lack of reliability and its assumption of categories defined by necessary and sufficient features. The solution to both problems was to develop specific criteria that could be evaluated one at a time as present or absent and to develop a simple procedure for diagnosing a given patient: counting the number of symptoms present from each criterion set and determining whether the patient meets a cutoff for diagnosis. The explicit assumption was that clinical judgment is inherently unreliable and that more specific, operational criteria would increase reliability of diagnosis. As Widiger and Frances (1985) put it, "If interrater reliability is to be achieved, the amount of inference required by the diagnostic criteria must be decreased . . ." (p. 617).

An Alternative Operationalization of a Prototype-Matching Approach

An alternative hypothesis (which we are currently testing with respect to Axis II diagnosis) is that clinical inference can in fact be reliable if clinicians are not forced to make dichotomous (present–absent) decisions about either diagnoses treated as a whole (DSM–II) or individual diagnostic criteria treated individually (DSM–III through DSM–IV). Thus, for clinical purposes, diagnoses could be made rapidly and efficiently by having clinicians make simple Likert-type ratings of the extent to which the patient's symptoms taken as a whole match each of several diagnostic prototypes, with the prototypes developed empirically through application of statistical aggregation techniques to large data sets. The result would be a series of prototypicality ratings that would provide a symptom profile, much like that provided by the Minnesota Multiphase Personality Inventory–2nd edition profile.

Westen and Shedler (1999b, 2000) have attempted to develop such a procedure for the classification and diagnosis of personality disorders. Over the course of several studies, they presented large, randomly selected samples of clinicians with a 200-item personality pathology Q-sort procedure that included not only versions of the 80-plus criteria currently included on Axis II but also a broad range of items assessing aspects of personality functioning and pathology not currently included on Axis II. The task of the clinician was simply to describe a randomly selected patient's personality in two studies of a patient currently diagnosed on Axis II: in one study, a patient with subclinical personality pathology, and in another, an adolescent patient being treated for enduring, maladaptive patterns of thought, feeling, motivation, or behavior (whether or not severe

enough to warrant an adult Axis II diagnosis). The investigators aggregated these descriptions using Q-analysis (inverted factor analysis), a clustering procedure that groups cases together on the basis of their similarity across the 200-variable item set.

The results, in both adult and adolescent samples (see Westen & Chang, 2000; Westen, Shedler, Glass, Zimmerman, & Martens, 2001), were a series of clinically and empirically coherent prototypes, some of which (e.g., narcissistic, paranoid) resemble current Axis II categories, whereas others (e.g., dysphoric personality disorder) do not. Patients' scores on each diagnosis are calculated by correlating their 200-item profile with the 200-item empirically derived prototype (Q-factor) for each diagnosis, and converting these to T scores (with a mean of 50 and standard deviation of 10). Using this approach, patients can receive both dimensional diagnoses (which index the extent to which their personality profile matches —that is, correlates with—each prototype) and categorical diagnoses (e.g., T-score elevation of 1.5 standard deviations). In several samples, these T scores have yielded theoretically predicted correlations with relevant external criterion variables, such as measures of adaptive functioning and etiology (see Westen & Shedler, 2000).

We are currently testing whether diagnosis by Q-sort with 200 variables is as or more reliable and valid than current diagnostic procedures for research purposes. For clinical purposes, we are proposing for the next edition of the *DSM* a simplified prototype-matching procedure that may prove a useful alternative to the current symptom-counting algorithm.[1] In this model, Axis II would consist of a set of prototype personality descriptions, each including 15–20 statements about the patient's characteristic patterns of thought, feeling, motivation, interpersonal functioning, and so forth. Rather than consider symptoms one at a time, dichotomize them as present or absent, and count the number of symptoms present, the clinician's task is much simpler: Examine each criterion set as a gestalt and decide to what extent the patient's personality matches the prototype.

Figure 9.2 presents the narcissistic prototype, empirically derived through Q-analysis of a sample of 496 patients, and converted to paragraph form (by grouping together items with related content). After reading through the items that constitute this prototype, the clinician makes a simple prototype rating, using the scale reproduced in the figure. The patient's dimensional score represents the extent to which he or she resembles the prototype. From a categorical point of view, patients who receive a score of 4 or 5 would be considered to have the disorder.

A similar procedure could be used to diagnose disorders such as those currently coded on Axis I. For example, the Axis I description of depression might include 8–10 symptoms that constitute the syndrome of major de-

---

[1]The current version of this prototype matching approach was developed in consultation with Robert Spitzer and Michael First.

pression, and the clinician's task is to rate the extent to which the patient's condition matches the prototype taken as a whole (see Figure 9.2). A score of 4 or 5 would mean that the patient's symptoms approximate the complete prototype and would warrant a categorical diagnosis of major depression. A score of 3 would mean that the patient's symptom picture resembles the prototype in many respects but not enough to warrant a categorized diagnosis of major depression. The patient would thus receive a subclinical diagnosis. There is little need, using this method, for the not otherwise specified categories. The same patient may or may not receive a high prototype rating on depression, panic, generalized anxiety, social phobia, and so on.

What is important in this example, as in the example of narcissistic personality disorder, is that the clinician is not asked to dichotomize symptoms that are probably continuously distributed in nature or to count the number of symptoms rated as present. This simplifies the clinicians' task immensely—and is probably much closer to the prototype-matching process clinicians intuitively use in everyday practice. Using the current diagnostic algorithms, clinicians who engage in intuitive prototype matching of this sort, which we suspect most do, invariably make invalid and unreliable diagnoses, precisely because of the diagnostic decision rules built into DSM–IV. Using a more straightforward prototype-matching process such as the one described here, a clinician familiar with a patient could rate the entire range of syndromes in the diagnostic manual in 3 to 5 minutes.

Four points deserve brief mention here. First, the method of diagnosis we are proposing represents in some respects a hybrid between DSM–II

5	very strong match (patient's personality *exemplifies* this disorder; prototypical case)	
4	strong match (patient *has* this disorder; diagnosis applies)	**Diagnosis**
3	moderate match (patient has *significant features* of this disorder)	**Features**
2	slight match (patient has minor features of this disorder)	
1	no match (description does not apply to the patient)	

Figure 9.2. A prototype-matching approach to diagnosis of narcissistic personality disorder. Individuals who match this prototype have fantasies of unlimited success, power, beauty, talent, brilliance, and so on. They appear to feel privileged and entitled, and they expect preferential treatment. They have an exaggerated sense of self-importance and believe they can only be appreciated by, or should only associate with, people who are high-status, superior, or otherwise "special." Individuals who match this prototype seek to be the center of attention and seem to treat others primarily as an audience to witness their own importance, brilliance, beauty, etc. They tend to be arrogant, haughty, or dismissive; to be competitive with others (whether consciously or unconsciously); to feel envious; and to think others are envious of them. They expect themselves to be "perfect" (e.g., in appearance, achievements, performance) and are likely to fantasize about finding ideal, perfect love. They tend to lack close friendships and relationships; to feel life has no meaning; and to feel like they are not their true selves with others, so that they may feel false or fraudulent. Adapted from Westen and Shedler, 2000.

and more recent editions of the diagnostic manual and would lead to similar hybrid forms of clinical and research interviewing. DSM–II had two virtues: brevity and prototypic descriptions of disorders whose "gist" clinicians could readily capture. However, it had several problems, including lack of empirically derived or empirically testable criteria (because diagnostic criteria were not separated), lack of empirically derived diagnostic groupings, built-in but untested etiological assumptions, assumptions about the categorical nature of disorders that rendered reliability of diagnosis impossible (a patient either had or did not have a given diagnosis, depending on whether the patient fit the description), and an implicit if not explicit assumption of a defining-features approach to categorization.

We are essentially suggesting a return to a manual consisting of prototype descriptions of disorders, with the brevity of DSM–II but the systematic empirical selection of criteria—and hence the minimization of comorbidity that may be an artifact of redundant and mixed diagnoses—and the prototype model of categorization characteristic of later editions of the DSM. The implication for assessment would be that interviewers can use clinical skill in determining how much to ask about each disorder, as in the DSM–II era, but follow a more explicit structure that guarantees comprehensiveness of diagnosis, which makes use of some of the semistructured interview techniques that emerged in the late 1970s. In our research on Axis II, we are using an interview that resembles a 3-session exploratory psychiatric-psychotherapy intake assessment (see Westen & Muderrisoglu, 2001; Westen, Muderrisoglu, Fowler, Shedler, & Koren, 1997). The interviewer begins by asking patients to describe themselves, what brought them in for treatment, the history of their symptoms, and their family and developmental history. The interviewer then elicits a series of narrative descriptions of patients' significant family, friend, work, and love relationships over the course of their lives.

To assess current states and Axis I syndromes, the clinician is guided by some general rules (e.g., always ask about mood, substance use, eating patterns, antisocial behavior, and clinical and subclinical thinking disturbance, whether or not the patient notes problems along these lines) but only asks about most criteria of most disorders if the clinical material indicates a reason to do so. Thus, an interviewer does not typically ask a patient who is in a stable relationship and functioning well at work whether he hears voices, and the interviewer does not ask a patient who shows no signs of depression in a 150-minute interview whether his or her eating patterns have changed in the last 2 weeks. Although this procedure may produce a small number of false negatives, it is more parsimonious, less cumbersome, and provides, we believe, a better balance between data collection and alliance maintenance. It differs little from decision-tree approaches to structured diagnosis in which the patient's answers to certain questions lead the interviewer either to inquire or not inquire further about

symptoms of particular syndromes. The difference is that our method is less dependent on a scripted list of questions and is more interactive and sensitive to the alliance, which we believe is likely to offset any losses associated with less standardization of wording.

Second, and implicit in the first point, if a prototype-matching algorithm for diagnosis were built into the next edition of DSM, the prototypes should, to the extent possible, be developed empirically, just as researchers have distinguished positive, negative, and disorganized symptoms of schizophrenia; distinct aspects of panic, generalized anxiety and depression, and so forth. We have been able to eliminate comorbidity from personality diagnosis in our studies of adult and adolescent patients with personality pathology, and these data are cross-replicating in new samples. Numerical taxonomic efforts are no panacea, and expert committees with large pools of clinical consultants will always be required to choose among alternative solutions. Nevertheless, a committee with access to the results of four or five studies applying statistical aggregation techniques to large samples using large item sets, each including data against which to validate the various potential solutions, would likely make better decisions than a committee attempting to refashion the categories and criteria that evolved from Bleuler and Kraepelin's extraordinary but intuitive taxonomic methods from a century ago, as modified by data and committee compromises over the last half century.

Third, using a simple prototype-matching system such as the one outlined here, the diagnostic manual could avoid arbitrary temporal cutoffs (e.g., the episode must have lasted at least 2 weeks), just as it could avoid arbitrary cutoffs for number of criteria present to constitute the diagnosis. For all diagnoses on which the patient receives a score of 3 or more (indicating at least moderate match to the prototype), the clinician would simply rate duration of the current episode other potentially useful variables such as severity of current symptoms and age of onset of first episode. This approach would render a diagnosis much closer to a case formulation: The clinician would describe the symptom and then the duration and one or two other variables relevant to the treatment decision, rather than only describe the symptom if it meets arbitrary duration criteria.

Finally, in the next edition of DSM, we will need to rethink the multiaxial system, so that the system has three characteristics that it currently approximates only imperfectly. First, the axes should be defined more clearly. Axis I, for example, includes both states and enduring personality conditions, and it arbitrarily includes some enduring conditions (such as dysthymic disorder) while excluding others (such as personality disorders). We suspect that the only consistent way to create a first axis that codes symptoms is to limit it to states, assessed prototypically, so that patients would receive a series of 1–5 ratings for panic, generalized anxiety, depression, mania, positive symptoms of psychosis, negative symptoms, and

so forth. A clinician could make this kind of Axis I diagnosis very quickly, by only listing diagnoses on which the patient receives a score greater than 1, with 1 (i.e., no match to the prototype) being the default score. Second, the axes should yield a case formulation, each providing nonredundant information about the major things a clinician needs to know and communicate about a patient, including current condition (state), duration of the condition, history of psychiatric conditions, personality context, etiological context, and recent stressors. Third, the axes should be statistically nonredundant: We should consider using a multiple regression model in deciding whether to add any potential axes, asking whether it provides data that predict incremental variance in clinically relevant variables.

CONCLUSION

Any proposals for refinement of the current diagnostic system should acknowledge the considerable divergence between the aims and observations of clinicians and researchers and the importance of incorporating both perspectives. A diagnostic manual should not be moored in unsystematic clinical observation and clinical hypotheses, as were *DSM–I* and *DSM–II*; these early editions of *DSM* were first approximations of a classification system developed without the tools we now have at our disposal. On the other hand, if a diagnostic system diverges so far from clinical experience that clinicians begin to disregard major aspects of it, such as the algorithms it provides for making diagnoses, we should be careful before we pathologize clinicians for not using the manual appropriately (or set up training programs for their remediation) and consider the equally plausible explanation that the manual requires reconfiguration.

We believe this is what has happened with aspects of *DSM–IV*. If assessing hundreds of diagnostic criteria by asking a series of questions about each one; making forced-choice, present–absent decisions about each criterion; and then counting them yielded information that helped clinicians treat their patients, we suspect that, 20 years after the introduction of *DSM–III*, clinicians would have caught on to the benefit of doing so. Clinical practice is an imperfect mechanism for assessing clinical utility, but it is certainly a useful bellwether.

In our own research, we have enlisted the help of hundreds of experienced clinicians, not by pooling their intuitive theories or biases about the nature of psychopathology, but by enlisting their expertise by asking them to describe a patient using a psychometrically sound instrument, and then pooling their knowledge through statistical aggregation procedures. There are many ways to approach the classification of psychiatric disorders, and this is just one of them. The prototype-matching approach we have proposed here requires much more research on implementation to see

whether clinicians can use it reliably and whether doing so leads to increases or decreases in reliability and predictive validity. We suspect, however, that the approaches that will ultimately prove most successful are ones that will engage clinicians and researchers in an ongoing, collaborative effort that acknowledges the strengths and limitations of each of their vantage points.

REFERENCES

American Psychiatric Association. (1952). *Diagnostic and statistical manual of mental disorders*. Washington, DC: Author.

American Psychiatric Association. (1968). *Diagnostic and statistical manual of mental disorders* (2nd ed.). Washington, DC: Author.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

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

American Psychiatric Association. (1998). *DSM–IV sourcebook (Vol. 4)*. Washington, DC: Author.

Andreasen, N. (1999). A unitary model of schizophrenia: Bleuler's "fragmented phrene" as schizencephaly. *Archives of General Psychiatry, 56*(9), 781–787.

Andreasen, N. C., & Carpenter, W. T., Jr. (1993). Diagnosis and classification of schizophrenia. *Schizophrenia Bulletin, 19*(2), 199–214.

Andreasen, N. C., & Grove, W. M. (1982). The classification of depression: Traditional versus mathematical approaches. *American Journal of Psychiatry, 139*, 45–52.

Andreasen, N. C., Paradiso, S., & O'Leary, D. S. (1998). "Cognitive dysmetria" as an integrative theory of schizophrenia: A dysfunction in cortical–subcortical–cerebellar circuitry? *Schizophrenia Bulletin, 24*(2), 203–218.

Barlow, D. (2002). *Anxiety and its disorders* (2nd ed.). New York: Guilford Press.

Barrett, J. E., Barrett, J. A., Oxman, T. E., & Gerber, P. D. (1988). The prevalence of psychiatric disorders in a primary care practice. *Archives of General Psychiatry, 45*, 1100–1106.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561–571.

Blashfield, R. K., & Morey, L. C. (1979). The classification of depression through cluster analysis. *Comprehensive Psychiatry, 20*, 516–527.

Bleuler, E. (1950). *Dementia praecox or the group of schizophrenias* (J. Zinkin, Trans.). New York: International Universities Press. (Original work published 1911)

Brown, T. A., Chorpita, B. F., & Barlow, D. H. (1998). Structural relationships among dimensions of the *DSM–IV* anxiety and mood disorders and dimen-

sions of negative affect, positive affect, and autonomic arousal. *Journal of Abnormal Psychology, 107,* 179–192.

Cantor, N., & Genero, N. (1986). Psychiatric diagnosis and natural categorization: A close analogy. In T. Millon & G. Klerman (Eds.), *Contemporary directions in psychopathology: Toward the DSM–IV* (pp. 233–256). New York: Guilford Press.

Cantor, N., Smith, E., French, R., & Mezzich, J. (1980). Psychiatric diagnosis as prototype categorization. *Journal of Abnormal Psychology, 89,* 181–193.

Chen, L., Eaton, W. W., Gallo, J. J., Nestadt, G., & Crum, R. M. (2000). Empirical examination of current depression categories in a population-based study: Symptoms, course, and risk factors. *American Journal of Psychiatry, 157,* 573–580.

Clark, L. A. (1989). The anxiety and depressive disorders: Descriptive psychopathology and differential diagnosis. In P. C. Kendall & D. Watson (Eds.), *Anxiety and depression: Distinctive and overlapping features* (pp. 83–129). San Diego, CA: Academic Press.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100,* 316–336.

Dollfus, S., Everitt, B., Ribeyre, J. M., Assouly-Besse, F., Sharp, C., & Petit, M. (1996). Identifying subtypes of schizophrenia by cluster analyses. *Schizophrenia Bulletin, 22,* 545–555.

Everitt, B. S. (1979). Unresolved problems in cluster analysis. *Biometrics, 35,* 169–181.

Eysenck, H. J. (1994). Personality and temperament. In D. Tantam & M. J. Birchwood et al. (Eds.), *Seminars in psychology and the social sciences* (College Seminars Series, pp. 187–202). London: Gaskell Royal College of Psychiatrists.

Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R. A., & Winokur, G. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry, 26,* 56–73.

Fleiss, J. L., Lawlor, W., Platman, S. R., & Fieve, R. (1971). On the use of inverted factor analysis for generating typologies. *Journal of Abnormal Psychology, 77,* 126–133.

Frances, A. (1982). Categorical and dimensional systems of personality diagnosis: A comparison. *Comprehensive Psychiatry, 23,* 516–527.

Frances, A., & Widiger, T. A. (1986). Methodological issues in personality disorder diagnosis. In T. Millon, G. Klerman (Eds.), *Contemporary directions in psychopathology: Towards the DSM–IV* (pp. 381–400). New York: Guilford Press.

Grove, W. M., & Andreasen, N. C. (1986). Multivariate statistical analysis in psychopathology. In T. Millon & G. L. Klerman (Eds.), *Contemporary directions in psychopathology: Towards the DSM–IV* (pp. 347–362). New York: Guilford Press.

Horowitz, L. M., Post, D. L., de Sales French, R., Wallis, K. D., & Siegelman, E. Y. (1981). The prototype as a construct in abnormal psychology: 2. Clarifying disagreement in psychiatric judgments. *Journal of Abnormal Psychology, 90,* 575–585.

Jampala, V. L., Sierles, F. S., & Taylor, M. A. (1988). The use of *DSM–III* in the

United States: A case of not going by the book. *Comprehensive Psychiatry, 29,* 39–47.

John, E. R., Prichep, L. S., Alper, K. R., Mas, F. G., Cancro, R., Easton, P., & Sverdlov, L. (1994). Quantitative electrophysiological characteristics and subtyping of schizophrenia. *Biological Psychiatry, 36,* 801–826.

Johnson, J. G., Cohen, P., Skodol, A., Oldham, J., Kasen, S., & Brook, J. S. (1999). Personality disorders in adolescence and risk of major mental disorders and suicidality during adulthood. *Archives of General Psychiatry, 56,* 805–811.

Katon, W., & Roy-Byrne, P. (1991). Mixed anxiety and depression. *Journal of Abnormal Psychology, 100,* 337–345.

Keller, M. B., Klein, D. N., Hirschfeld, R. M. A., Kocsis, J. H., McCullough, J. P., Miller, I., First, M. B., Holtzer, C. P., Keitner, G. I., Marin, D. B., & Shea, T. (1995). Results of the *DSM–IV* mood disorders field trial. *American Journal of Psychiatry, 152,* 843–849.

Keller, M. B., & Lavori, P. W. (1984). Double depression, major depression, and dysthymia: Distinct entities or different phases of a single disorder? *Psychopharmacology Bulletin, 20,* 399–402.

Kendler, K. S., Karkowski, L. M., & Walsh, D. (1998). The structure of psychosis. *Archives of General Psychiatry, 55,* 492–499.

Kendler, K. S., Karkowski-Shuman, L., O'Neill, A., Straub, R. E., MacLean, C. J., & Walsh, D. (1997). Resemblance of psychotic symptoms and syndromes in affected sibling pairs from the Irish study of high-density schizophrenia families: Evidence for possible etiologic heterogeneity. *American Journal of Psychiatry, 154,* 191–198.

Kendler, K. S., Neale, M. C., Prescott, C. A., Kessler, R. C., Heath, A. C., Corey, L. A., & Eaves, L. J. (1996). Childhood parental loss and alcoholism in women: A causal analysis using a twin family design. *Psychological Medicine, 26,* 79–95.

Kendler, K. S., Spitzer, R. L., & Williams, J. B. W. (1989). Psychotic disorders in *DSM–III–R. The American Journal of Psychiatry, 146,* 953–962.

Kernberg, O. (1984). *Severe personality disorders.* New Haven, CT: Yale University Press.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., Wittchen, H.-U., & Kendler, K. S. (1994). Lifetime and 12-month prevalence of *DSM–III–R* psychiatric disorders in the United States. *Archives of General Psychiatry, 51,* 8–19.

Kessler, R., Nelson, C., McGonagle, K., Liu, M., Swartz, M., & Blazer, D. (1996). Comorbidity of *DSM–III–R* major depressive disorder in the general population: Results from the US National Comorbidity Survey. *British Journal of Psychiatry, 168,* 17–30.

Klein, D. N. (1999). Depressive personality in the relatives of outpatients with dysthymic disorder and episodic major depressive disorder and normal controls. *Journal of Affective Disorders, 55*(1), 19–27.

Kraepelin, E. (1898). *The diagnosis and prognosis of dementia praecox.* Paper presented at the 29th Congress of Southwestern German Psychiatry, Heidelberg.

Lencz, T., Raine, A., & Sheard, C. (1996). Neuroanatomical bases of electrodermal hyporesponding: A cluster analytic study. *International Journal of Psychophysiology, 22*, 141–153.

Lenzenweger, M. F. (1999). Schizophrenia: Refining the phenotype, resolving the endophenotypes. *Behaviour Research and Therapy, 37*, 281–295.

Lewinsohn, P., Rohde, P., Seeley, J. R., & Klein, D. N. (1997). Axis II psychopathology as a function of Axis I disorders in childhood and adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*, 1752–1759.

Livesley, W., & Jackson, D. (1992). Guidelines for developing, evaluating, and revising the classification of personality disorder. *Archives of General Psychiatry, 55*, 941–948.

Livesley, W. J., Schroeder, M., Jackson, D. N., & Jang, K. (1994). Categorical distinctions in the study of personality disorder: Implications for classification. *Journal of Abnormal Psychology, 103*, 6–17.

Loranger, A. W. (1990). The impact of DSM–III on diagnostic practice in a university hospital: A comparison of DSM–II and DSM–III in 10,914 patients. *Archives of General Psychiatry, 47*, 329–344.

Malt, B. (1993). Concept structure and category boundaries. In G. V. Nakamura, D. L. Medin, & R. Taraban (Eds.), *Categorization by humans and machines. The psychology of learning and motivation: Advances in research and theory* (Vol. 29, pp. 363–390). San Diego, CA: Academic Press.

McCrae, R. R., & Costa, P. T. (1990). Personality disorders and the five-factor model. *Journal of Personality Disorders, 4*(4), 362–371.

McCrae, R. R., & Costa, P. T. (1997). Personality trait structure as a human universal. *American Psychologist, 52*(5), 509–516.

Meehl, P. (1962). Schizotaxia, schizotypy, schizophrenia. *American Psychologist, 17*, 827–838.

Mendels, J., & Cochrane, C. (1968). The nosology of depression: The endogenous-reactive concept. *American Journal of Psychiatry, 11*, Supplement.

Millon, T. (1990). *Toward a new personology: An evolutionary model.* New York: Wiley.

Morel, B. A. (1852). *Traité de maladies mentales* [Treatise on mental disorders]. Paris: Masson.

Morrison, K., & Westen, D. (2000). *The external validity of psychotherapy trials: An empirical investigation.* Unpublished manuscript, Boston University.

Newman, D. L., Moffitt, T., Caspi, A., & Silva, P. A. (1998). Comorbid mental disorders: Implications for treatment and sample selection. *Journal of Abnormal Psychology, 107*, 305–311.

Olfson, M., Broadhead, W. E., Weissman, M. M., Leon, A. C., Farber, L., Hoven, C., & Kathol, R. (1996). Subthreshold psychiatric symptoms in a primary care group practice. *Archives of General Psychiatry, 53*, 880–886.

Paykel, E. S. (1971). Classification of depressed patients: A cluster analysis derived grouping. *British Journal of Psychiatry, 137*, 275–288.

Phillips, K. A., & Gunderson, J. G. (1999). Depressive personality: Fact or fiction? *Journal of Personality Disorders, 13*(2), 128–134.

Pinel, P. (1962). *A treatise on insanity*. New York: Hafner. (Reprinted from article published in 1801 in *Psychology, 100*, 337–345).

Raskin, A., & Crook, T. H. (1976). The endogenous–neurotic distinction as a predictor of response to antidepressant drugs. *Psychological Medicine, 6*, 59–70.

Robins, E., & Guze, S. B. (1970). Establishment of diagnostic validity in psychiatric illness: Its application to schizophrenia. *American Journal of Psychiatry, 126*, 983–986.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization*. New York: Wiley.

Sham, P. C., Castle, D. J., Wessely, S., Farmer, A. E., & Murray, R. M. (1996). Further exploration of a latent class typology of schizophrenia. *Schizophrenia Research, 20*(1–2), 105–115.

Shedler, J., Mayman, M., & Manis, M. (1993). The illusion of mental health. *American Psychologist, 48*, 1117–1131.

Skinner, H. A. (1981). Toward the integration of classification theory and methods. *Journal of Abnormal Psychology, 90*, 68–87.

Skinner, H. A. (1986). Construct validation approach to psychiatric classification. In T. Millon & G. L. Klerman (Eds.), *Contemporary directions in psychopathology: Towards the DSM–IV* (pp. 307–329). New York: Guilford Press.

Smith, E. E. (1995). Concepts and categorization. In E. Smith & D. Osherson (Eds.), *Thinking: An invitation to cognitive science* (Vol. 3, 2nd ed., pp. 3–33). Cambridge, MA: MIT Press.

Sokol, R. R. (1974). Classification: Purposes, principles, progress, prospects. *Science, 185*, 1115–1123.

Spielberger, C. D., Gorsuch, R. I., & Lushene, R. E. (1970). *Manual for the State–Trait Anxiety Inventory (Self-Evaluation Questionnaire)*. Palo Alto, CA: Consulting Psychologists Press.

Spitzer, R. L., Endicott, J., & Robins, E. (1978). Research diagnostic criteria: Rationale and reliability. *Archives of General Psychiatry, 35*, 773–782.

Strauss, J. S., Carpenter, W. T., Jr., & Bartko, J. J. (1974). The diagnosis and understanding of schizophrenia: Part III. Speculations on the processes that underlie schizophrenia signs and symptoms. *Schizophrenia Bulletin, 1*(11), 61–69.

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*, 1398–1406.

Tsai, G., van Kammen, D. P., Chen, S., Kelley, M. E., Grier, A., & Coyle, J. T. (1998). Glutamatergic neurotransmission involves structural and clinical deficits of schizophrenia. *Biological Psychiatry, 44*, 667–674.

Tsuang, M. T., Stone, W. S., & Faraone, S. V. (in press). Towards reformulating the diagnosis of schizophrenia. *American Journal of Psychiatry*.

Van der Does, A. J. W., Dingemans, P. M., Linszen, D. H., Nugter, M. A., & Scholte, W. F. (1993). Symptom dimensions and cognitive and social functioning in recent-onset schizophrenia. *Psychological Medicine, 23*, 745–753.

Van der Does, A. J. W., Dingemans, P. M. A. J., LinSzen, D. H., Nugter, M. A., &

Scholte, W. F. (1995). Dimensions and subtypes of recent-onset schizophrenia: A longitudinal analysis. *Journal of Nervous and Mental Disease, 183,* 681–687.

Waller, N., & Meehl, P. (1998). *Multivariate taxometric analysis.* New York: Sage.

Westen, D. (1995). A clinical–empirical model of personality: Life after the Mischelian ice age and the NEO-lithic era. *Journal of Personality, 63,* 495–524.

Westen, D. (1997). Divergences between clinical and research methods for assessing personality disorders: Implications for research and the evolution of Axis II. *American Journal of Psychiatry, 154,* 895–903.

Westen, D. (1998). Case formulation and personality diagnosis: Two processes or one? In James Barron (Ed.), *Making diagnosis meaningful* (pp. 111–138). Washington, DC: American Psychological Association.

Westen, D. (1999). *Personality disorders in DSM–IV: A critical appraisal.* Unpublished manuscript, Boston University.

Westen, D., & Arkowitz-Westen, L. (1998). Limitations of Axis II in diagnosing personality pathology in clinical practice. *American Journal of Psychiatry, 155,* 1767–1771.

Westen, D., & Chang, C. M. (2000). Adolescent personality pathology: A review. *Adolescent Psychiatry, 25,* 61–100.

Westen, D., & Muderrisoglu, S. (2001). *Reliability and validity of personality disorder assessment using a systematic clinical interview.* Unpublished manuscript, Boston University.

Westen, D., Muderrisoglu, S., Fowler, C., Shedler, J., & Koren, D. (1997). Affect regulation and affective experience: Individual differences, group differences, and measurement using a Q-sort procedure. *Journal of Consulting and Clinical Psychology, 65,* 429–439.

Westen, D., & Shedler, J. (1999a). Revising and assessing Axis II: Part I. Developing a clinically and empirically valid assessment method. *American Journal of Psychiatry, 156,* 258–272.

Westen, D., & Shedler, J. (1999b). Revising and assessing Axis II: Part II. Toward an empirically based and clinically useful classification of personality disorders. *American Journal of Psychiatry, 156,* 273–285.

Westen, D., & Shedler, J. (2000). A prototype matching approach to personality disorders: Toward DSM–V. *Journal of Personality Disorders, 14,* 109–126.

Westen, D., Shedler, J., Glass, S., Zimmerman, C., & Martens, A. (2001). *Personality disorders in adolescence, II: An empirically derived classification.* Unpublished manuscript, Boston University.

Widiger, T. (1993). Validation strategies for the personality disorders. *Journal of Personality Disorders,* Supplement, 34–43.

Widiger, T., & Frances, A. (1985). The *DSM–III* personality disorders: Perspectives from psychology. *Archives of General Psychiatry, 42,* 615–623.

Zinbarg, R. E., & Barlow, D. H. (1996). Structure of anxiety and the anxiety disorders: A hierarchical model. *Journal of Abnormal Psychology, 105,* 181–193.

Zinbarg, R. E., Barlow, D. H., Liebowitz, M., & Street, L. (1994). The *DSM–IV* field trial for mixed anxiety–depression. *American Journal of Psychiatry, 151,* 1153–1162.