Journal of Experimental Psychology: General
2002, Vol. 131, No. 4, 451–476

0096-3445/02/$5.00 DOI: 10.1037//0096-3445.131.4.451

Clinical Psychologists' Theory-Based Representations of Mental Disorders Predict Their Diagnostic Reasoning and Memory

Nancy S. Kim
Yale University

Woo-kyoung Ahn
Vanderbilt University

The theory-based model of categorization posits that concepts are represented as theories, not feature lists. Thus, it is interesting that the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM–IV;* American Psychiatric Association, 1994) established atheoretical guidelines for mental disorder diagnosis. Five experiments investigated how clinicians handled an atheoretical nosology. Clinicians' causal theories of disorders and their responses on diagnostic and memory tasks were measured. Participants were more likely to diagnose a hypothetical patient with a disorder if that patient had causally central rather than causally peripheral symptoms according to their theory of the disorder. Their memory for causally central symptoms was also biased. Clinicians are cognitively driven to use theories despite decades of practice with the atheoretical *DSM*.

The theory-based view of categorization proposes that concepts are represented as theories or causal explanations. Murphy and Medin (1985) suggested that our naïve theories about the world hold the features of a concept together in a cohesive package. For instance, a layperson's concept of anorexia contains not only the features "fear of becoming fat" and "refuses to maintain minimal body weight" but also the notion that the fear of becoming fat helps cause the refusal to maintain minimal body weight (Kim & Ahn, 2002). Indeed, a growing body of evidence supports the notion that the human mind constantly seeks out rules and explanations that make sense of incoming data concerning its surroundings and forms concepts based on its theories about the world (Carey, 1985; Gelman, 2000; Keil, 1989; Murphy & Medin, 1985).

A considerable number of studies have demonstrated theory-based categorization to date (e.g., Ahn, Kim, Lassaline, & Dennis, 2000; Gelman & Kalish, 1993; Medin & Shoben, 1988; Ross, 1997; Wisniewski & Medin, 1994), mostly in artificial and common everyday categories. The aim of the current article is to examine what kind of reasoning occurs for a real-life domain in which official guidelines for categorization deliberately attempt to minimize a prescribed theoretical structure.

Specifically, the current study investigated how clinical psychologists operate in the domain of mental disorders. This population and domain are unique in that clinical psychologists have been guided since 1980 by atheoretical manuals for the diagnosis of mental disorders (Follette & Houts, 1996). Most mental disorders lack a single universally acknowledged pathogenesis, which in the past led to unreliability between clinicians in diagnosis. The *Diagnostic and Statistical Manual of Mental Disorders'* (4th ed.; *DSM–IV*; American Psychiatric Association, 1994) widely acclaimed solution was to ferret out syndromal clusters of symptoms that clinicians, regardless of theoretical orientation, could agree on. The manual itself represents disorders as checklists of symptoms and does not attempt to supply "an organizing theory that describes the fundamental principles underlying the taxonomy" (Follette & Houts, 1996, p. 1120).

The introduction of the *DSM–IV* states that "*DSM–III* introduced a number of important methodological innovations, including explicit diagnostic criteria . . . and a descriptive approach that attempted to be neutral with respect to theories of etiology" (American Psychiatric Association, 1994, pp. xvii–xviii). This approach was also adopted in the 4th version of the *DSM*. Indeed, the *DSM* casebook, used in training, encourages clinical psychologists to search for symptoms in their patients that match up with *DSM–IV* diagnostic criteria, without explicitly instructing them to incorporate any additional notions they may have of how these symptoms may affect each other (Spitzer, Gibbon, Skodol, Williams, & First, 1994). Furthermore, the *DSM–IV* states that if a subset of the diagnostic criteria list is present in the patient, that is sufficient for a diagnosis regardless of which combination of symptoms appears. The manual thereby assumes that all symptoms

Nancy S. Kim submitted this research as part of a doctoral dissertation to the Yale University Graduate School. This research was supported in part by a National Science Foundation Graduate Research Fellowship and a Yale University Dissertation Fellowship to Nancy S. Kim and by National Institute of Mental Health Grant RO1 MH57737 to Woo-kyoung Ahn. Many thanks to Paul Bloom, Marvin Chun, Marcia Johnson, Frank Keil, Donna Lutz, Laura Novick, Peter Salovey, and Andrew Tomarken for helpful comments and suggestions. We also thank Douglas Medin for inspiring the general framework for the methods used in this study and Judy Choi, Jessecae Marsh, and Talia Valdez for help in conducting a number of experimental sessions.

Correspondence concerning this article should be addressed to Nancy S. Kim, Department of Psychology, Yale University, 2 Hillhouse Avenue, P.O. Box 208205, New Haven, Connecticut 06520-8205, or to Woo-kyoung Ahn, Department of Psychology, Vanderbilt University, 534 Wilson Hall, 111 21st Avenue South, Nashville, Tennessee 37240. E-mail: nskim@aya.yale.edu or woo-kyoung.ahn@vanderbilt.edu

in the list are equally central to the disorder.[1] For example, any two of the following five symptoms should warrant a diagnosis of schizophrenia, according to the *DSM–IV*: hallucinations, delusions, disorganized speech, grossly disorganized or catatonic behavior, and negative symptoms.

Over the past 20 plus years, beginning with the *DSM–III* (APA, 1980), the *DSM* system has become widely accepted in the United States, forming the core of research, clinical assessment, diagnosis, and treatment in psychopathology. Research funding, journal titles, and health care reimbursements are all organized by, and dependent on, use of the categories defined by the *DSM–IV*. Clearly, use of the categories laid out by the *DSM–IV* is widespread. The question for the current article, then, is what kind of clinical reasoning emerges under use of the *DSM*?[2]

Do clinicians actually adhere to these guidelines outside formal diagnosis situations? If so, we would expect clinicians to give equal credence to all symptoms of a disorder. On the other hand, the theory-based approach argues that people treat features that are central to their domain theories as central in categorization as well. For instance, *being round* is treated as more central in categorizing basketballs than in categorizing cantaloupes because roundness is central in the naive physics underlying basketball concepts, but it is not central in the naive biology underlying cantaloupe concepts. Thus, clinicians reasoning in a theory-based manner might weigh symptoms differently depending on the theories they hold about the disorder.

In the remainder of this introduction, previous work on clinical decision making from the perspective of categorization research is first outlined. Next, it is suggested that clinicians' representations of mental disorders include notions of how the symptoms affect one another, and these representations can account for the relative importance clinicians assign to different symptoms in diagnosis and clinical reasoning. The possibility of expertise effects is also explored, followed by a brief overview of the current experiments.

Categorization in Clinical Decision-Making Research

The categorization literature has undergone several major shifts over the last few decades. The classical rule-based approach was followed by similarity-based models, which include prototype models (e.g., Rosch & Mervis, 1975; Posner & Keele, 1968) and exemplar models[3] (e.g., Medin & Schaffer, 1978). The theory-based view (e.g., Carey, 1985; Keil, 1989; Medin, 1989; Murphy & Medin, 1985) was also added to account for the role of explanation in categorization. This section discusses how these different approaches have been applied to mental disorder representations, as well as how these representations have been thought to be used in diagnostic reasoning. It is certainly not the intention here to suggest that only one of these views must be correct. Indeed, some, and perhaps all, of these approaches may account for some aspect of categorization and, furthermore, are likely to interact in some way with each other (Keil, Smith, Simons, & Levin, 1998; Wisniewski & Medin, 1994).

Rule-Based Approach

The rule-based view of categorization postulates that each category has individually necessary, and collectively sufficient, defining features (Medin, 1989; Smith & Medin, 1981). For example, the category *bachelor* may be defined by the conjunction of the features *adult*, *male*, and *unmarried*. The earliest known classification systems of mental disorders, espoused by Kraepelin (1913) and later by the *DSM–I* (1st ed.; American Psychiatric Association, 1958) and *DSM–II* (2nd ed.; American Psychiatric Association, 1968), also required necessary and sufficient features for diagnosis. For instance, Kraepelin (1913) postulated that the defining feature of schizophrenia was an early-life onset of dementia (Hill, 1983).

However, it proved difficult or impossible to come up with satisfactory defining features for most natural categories (Wittgenstein, 1953). For instance, *priest* fits the defining features for *bachelor*, but few people would actually refer to a priest as a bachelor. In addition, the classical approach cannot account for typicality effects in which some exemplars are rated as better members of a category than others. Such effects would not occur if categories were truly represented as defining features.

Similarity-to-Prototype Approach

One alternative to the rule-based approach is the prototype view (e.g., Rosch, 1978; Rosch & Mervis, 1975), which states that a category is represented as a prototype, an averaged, abstract representation of category members. Category membership is determined by an instance's similarity to that prototype. This approach can solve problems with the classical view. For instance, it can account for typicality effects in that the more similar the instance is to the prototype, the more likely it is to be categorized quickly, rated as more typical of the category, and so on.

Similarly, the *DSM–III* task force adopted a format more like the prototype approach. That format was retained in the *DSM–III–R* (3rd ed., rev.; American Psychiatric Association, 1988) and the current *DSM–IV*. This prototype-based nosology allows for more flexibility than did previous versions of the manual (Barlow & Durand, 1999). For example, the prototypical patient with schizophrenia has five symptoms, but a presenting patient need only have two of those five symptoms for a diagnosis of the disorder.

The validity of the prototype approach as the model of mental disorder diagnosis gained ground with additional studies. Cantor and her colleagues (Cantor, Smith, French, & Mezzich, 1980; Genero & Cantor, 1987), for instance, found that clinicians do make graded typicality ratings of patients, such that some are considered to be better examples of a disorder than others. They also found that clinicians were less accurate and confident when diagnosing atypical patients than when diagnosing patients highly or moderately typical of a disorder (see also Clarkin, Widiger, Frances, Hurt, & Gilmore, 1983; Horowitz, Post, French, Wallis,

[1] It should be noted that this is the case for many, though not all, *DSM–IV* disorders. For instance, for a diagnosis of major depressive episode, a subset of symptoms from each of two separate symptom lists must be present (two symptoms in one list and seven in the other).

[2] We emphasize that this article does not attempt to suggest how the *DSM* system could or should be changed. The research presented later on suggests a model of how clinicians do reason, but is not intended to show how clinicians ought to reason.

[3] The exemplar view has not been used as a model of mental disorder representation by the *DSM* systems and therefore is not discussed in detail here.

& Siegelman, 1981; Horowitz, Wright, Lowenstein, & Parad, 1981; Russell, 1991; Widiger, 1982).

Theory-Based Approach

The theory-based approach posits that the human mind forms categories and concepts based on its theories about the world (Carey, 1985). The current study investigated what role theory-based reasoning might also play in diagnosis by asking whether clinicians have internalized the atheoretical reasoning of the *DSM* system. One possibility is that experienced clinicians, after many years of following the prescribed *DSM* system (Spitzer et al., 1994), reason about mental disorder categories without being affected by their theories about how the symptoms affect each other. The alternative is that clinicians, despite being given atheoretical guidelines for diagnosis, are still influenced by their own idiosyncratic theories about disorders when reasoning about them. Indeed, Medin (1989) has suggested that the *DSM* system "provides only a skeletal outline that is brought to life by theories and causal scenarios underlying and intertwined with the symptoms that comprise the diagnostic criteria" (p. 1479).

The current research differentiates the two possibilities by examining two specific issues. First, we examined whether clinicians represent mental disorders as a list of independent symptoms, as represented in *DSM–IV* guidelines, or as a rich structure of symptoms that are highly interrelated, as assumed by the theory-based approach. Second, we examined whether clinicians give equal weights to symptoms as prescribed by the *DSM* system or, alternatively, whether symptoms central to clinicians' theories about mental disorders determine their weights, as claimed by the theory-based approach. The next section describes two known specific mechanisms by which domain theories determine feature weighting.

Ways in Which Theories Determine Feature Weighting

In this article, the focus is on the internal structure of concepts (i.e., Murphy & Medin, 1985), or how features are structured like theories, rather than on how concepts are connected to each other in domain theories. One specific mechanism by which the internal structure of concepts affects reasoning is the causal status effect (Ahn, 1998; Ahn et al., 2000). The causal status effect occurs when features causally central to an individual's theory of that category are treated as more important in categorization than less causally central features. For instance, if Symptom A causes Symptom B in a clinician's theory, then A is more causally central than B, and A is thereby predicted to have greater diagnostic importance than B. This effect has been shown in lay people with both *DSM–IV* disorders and artificial mental disorders (Kim & Ahn, 2002).

To derive the causal centralities of individual symptoms embedded in a complex theory, the following formula can be used:[4]

$$C_{i,t+1} = \sum_j d_{ij} c_{j,t}, \quad (1)$$

where d_{ij} is a positive number that represents how strongly symptom j depends on symptom i, and $c_{j,t}$ is the conceptual centrality of feature j, at time t (Sloman, Love, & Ahn, 1998). This model states that the centrality of feature i is determined at each time step by summing across the centrality of every other feature multiplied by that feature's degree of dependence on feature i. Thus, in the current studies, the theory-based view was operationalized as a systematic effect of relational structures on conceptual representation and use.[5]

Another way in which theories influence feature weighting is that features relationally connected to other features are treated as more important than isolated features in reasoning (Kim & Ahn, 2002). Gentner's (1983, 1989) structure-mapping theory, for example, argues that relational features (statements taking at least two arguments; for instance, x is smaller than y) are more important than attributes (statements taking only one argument; for instance, x is blue) in analogical inference. For instance, attributes such as *yellow*, *hot*, and *massive* are not particularly useful in making the analogy that an atom is like the solar system. In contrast, relational features such as *more massive than* and *revolves around* can be used to draw the analogy that electrons revolve around the nucleus in an atom as planets revolve around the sun in a solar system. (See also Lassaline, 1996.)

To summarize, the first hypothesis is that clinicians show a causal status effect when reasoning about symptoms, such that symptoms central to their theory of a disorder are treated as more important in diagnosis. The second hypothesis is that symptoms causally related to one another in a clinician's theory of a disorder are treated as more important than symptoms not thought to be part of the causal theory. Putting these two hypotheses together, we predicted that if Symptom A causes Symptom B but symptom C is isolated (it does not cause and is not caused by any other symptoms in a clinician's theory), C would be the least central symptom of the three. The presence of these effects is taken as evidence that domain theories influence clinical reasoning. Alternative explanations of such effects are considered in the General Discussion section.

It must be noted that the ecological validity achieved by using real-life mental disorders comes at the price of not being able to control for all other factors (see the General Discussion section). In our previous work with lay people, we were able to create artificial mental disorders and manipulate which symptoms were causally connected. In the present study, however, we were more interested in how clinicians deal with real disorders than in how they might learn to deal with hypothetical disorders for which they are fed the causal explanations (which, as we have pointed out, the *DSM* does not do). Thus, the studies reported here must be read with that caveat.

Expertise and the Use of Theories in Diagnosis

We also examined whether experts and trainees both adhere to the same strategy of mental disorder diagnosis. One alternative is

[4] Although other formulas are also consistent with the causal status effect, this formula showed the best fit in analyses of lay people's conceptual representations of common objects (e.g., apples and guitars; Sloman et al., 1998). Moreover, all of the reported analyses on causal centrality that follow are based on rank orders of causal centrality derived from this formula, and different formulas do not produce radically different rank orders.

[5] We do not intend to claim that theory-based categorization is limited to the effect of relational structures. Categorization may also be affected by what types of relationships are present between features, an issue that was not the focus of the current studies.

that experts are more likely to show theory-based reasoning than trainees because experts have developed more theories after decades of clinical work. A second alternative is that whereas trainees may behave like lay people, making theory-based diagnoses (Kim & Ahn, 2002), experts might diagnose atheoretically because of years of experience with the atheoretical *DSM* system. The third possibility is that both trainees and experts use their theories to weight features in diagnosis because theory-based reasoning is too cognitively compelling to be diminished even with a widely accepted set of atheoretical guidelines. Finally, it is possible that neither group is affected by theories in diagnosis, adhering strictly to *DSM–IV* guidelines. These last two alternatives, albeit for a different aspect of clinical reasoning, are consistent with previous findings that experts and novices differ little in the treatment outcome aspect of clinical work (i.e., Durlak, 1979; Faust & Zlotnick, 1995), an issue that we return to in the General Discussion.

Overview of Experiments

Three major experiments (Experiments 1, 2, and 4) are reported in this article. For purposes of comparison across studies, Figure 1 summarizes the methods and critical results for these experiments. In each of the experiments, the general methodology involved measuring each individual's theories (Task I in Figure 1) and seeing whether these theories could predict which symptoms were treated as more central in that individual's representation of mental disorders. Two sets of centrality measures were taken in each experiment: (a) importance of a symptom when diagnosing or thinking of a mental disorder (Tasks II and III in Figure 1) and (b) memory for patients' symptoms (Task IV in Figure 1). As described earlier, the main prediction was that a symptom on which another symptom depends in an individual's theory would be treated as more important in diagnosis and would be more likely to be remembered than dependent and isolated symptoms.

Experiment 1: Use of Causal Theories in Clinical Reasoning

Our question for this first experiment, then, was whether clinical psychologists conceptualized familiar mental disorders as the atheoretical, unweighted lists of the *DSM–IV* or whether they represented them as theories that affect their diagnostic reasoning. The focus was on explanatory, or causal, relations because these have been pegged as critical to the background knowledge used in categorization (Ahn, Marsh, Luhmann, & Lee, 2002; Carey, 1985; Wellman, 1990).

In the current experiment, we operationalized clinicians' and clinical trainees' use of theories in diagnosis with three different measures. First, participants were asked to rate the likelihood that hypothetical patients actually have the associated disorder. These patients had only either causally central, causally peripheral, or isolated symptoms, created according to each participant's theory of the disorder. Each type of hypothetical patient was given the same number of *DSM–IV* diagnostic criteria. Therefore, if participants did follow the guidelines of the *DSM–IV*, then no systematic differences in diagnosis likelihood ratings were expected. However, if the causal status model accurately reflects how participants used theory in diagnosis, patients with causally central symptoms were expected to have a higher likelihood of diagnosis than patients with causally peripheral symptoms. Furthermore, if the elevated importance for relational versus isolated features in analogical reasoning extends to diagnosis, then patients with causally central or peripheral symptoms were expected to have a higher likelihood of diagnosis than patients with isolated symptoms. Second, participants were asked to recall the symptoms of those hypothetical patients. If clinicians were biased to attend to symptoms causally central to their theories about the disorder, then they were expected to recall more causally central symptoms than causally peripheral or isolated symptoms. Alternatively, if they had internalized the *DSM*'s guidelines for diagnosis, then they

	Experiment 1		Experiment 2		Experiment 4	
Tasks	Method	Results	Method	Results	Method	Results
I. Theory drawing	To draw causal relations between symptoms	Figure 2 for a sample diagram	To draw any kind of relations between symptoms	Figures 7-11	Same as Experiment 2	Figures 13-16
II. Conceptual centrality	*How easily can you imagine a patient who has all the symptoms of [disorder X] except that he or she does not have [symptom Y]?*	Median *r* between causal centrality in (I) and conceptual centrality = 0.41 for experts and 0.27 for novices	*How important is the symptom of [Y] in diagnosing a person with [disorder X]?*	*r* between causal centrality in (I) and conceptual centrality = 0.85	*How important is the symptom of [Y] to your concept of [disorder X]?*	*r* between causal centrality in (I) and conceptual centrality = 0.46
III. Hypothetical patient diagnosis	*What is the likelihood that a patient [with either causally central, peripheral, or isolated symptoms in (I)] has [Disorder X]?*	Figure 3	*How well does a patient [with either causally central, peripheral, or isolated symptoms in (I)] fit in the diagnostic category of [X]?*	Figure 5	Same as Experiment 2	Figure 17
IV. Memory for symptoms presented in (III)	Recall	Figure 4	Recognition	Figure 6	Recall	Figure 18

Figure 1. Summary of results for each task in the three major experiments.

Table 1
Participant Characteristics in Experiments 1, 2, and 4

Characteristic	Experiment 1	Experiment 2	Experiment 3
Age in years; range (*Mdn*)			
Experts	38–71 (53)	45–73 (58)	39–67 (47)
Novices	22–30 (27)	25–32 (26)	23–29 (25)
Years seeing patients; range (*Mdn*)			
Experts	15–52 (28)	17–43 (26)	13–28 (18)
Novices	0–8 (1)	4–8 (5)	0.4–5 (1)
Hrs/week seeing patients, career; range (*Mdn*)			
Experts	5–45 (20)	5–30 (20)	5–35 (20)
Novices	0–20 (0.5)	3–20 (10)	0–18 (2)
Hrs/week seeing patients, current; range (*Mdn*)			
Experts	6–40 (20)	5–30 (15)	2–39 (18)
Novices	0–10 (0.5)	6–20 (14)	0–17.5 (2)
Mean current patients with Axis I disorders (%)			
Experts	95.0	90.0	69.5
Novices	77.5	92.5	97.1
Mean current patients with Axis II disorders (%)			
Experts	25.0	20.0	32.2
Novices	25.0	27.5	17.1
Psychoanalytic–humanistic clinicians (*n*)			
Experts	6	4	3
Novices	7	1	0
Cognitive–behavioral clinicians (*n*)			
Experts	3	6	3
Novices	0	2	8
Other orientation clinicians (*n*)			
Experts	2	4	4
Novices	3	3	1

were expected to recall equal proportions of all types of symptoms. Finally, we measured the conceptual centrality of each symptom and examined whether this measure correlated with causal centrality in each clinician's theory.

Method

Participants. Clinical experts were 10 clinical psychologists in independent practice in the New Haven, Connecticut area and 1 clinical psychologist from the Nashville, Tennessee area, each paid at the rate of $75 per hour. Clinical trainees were 10 clinical psychology graduate students at the Department of Psychology, Vanderbilt University, each paid $14 per hour.

On the basis of a 23-item questionnaire (adapted from Lehman, 1992) that participants completed at the end of the experiment, we compiled descriptive statistics for the two clinical samples (see Table 1).[6] Ten clinicians were licensed psychologists with doctorates of philosophy in clinical psychology, and 1 was a board-licensed psychiatrist and medical doctor. The students included 4 first-years, 3 second-years, 1 third-year, 1 fourth-year, and 1 eighth-year.

Materials and procedure. The disorders used as stimuli were selected such that all participants would be highly familiar with them. To do this, familiarity ratings were obtained from untrained undergraduates under the assumption that disorders familiar to them should also be familiar to expert and trainee clinicians. (See Appendix A for a detailed description of this preexperiment and our criteria for selecting disorders.) The *DSM–IV* disorders selected were anorexia nervosa, schizophrenia, major depressive episode, antisocial personality disorder, and specific phobia.

Participants each engaged in two sessions, spaced 10–14 days apart, to complete the eight tasks. The total time to complete both sessions ranged from 2.5 to 6 hr. In the first session, they completed a familiarity-rating task, a disorder-defining task, a theory-drawing task, and a conceptual centrality task. The familiarity-rating task was always completed first, followed by the disorder-defining task, and the order of the theory-drawing and conceptual centrality tasks was counterbalanced between participants. In the second session, they completed a hypothetical patient diagnosis task, an everyday categories theory-drawing task, an everyday categories conceptual centrality task, and a free-recall task. The tasks in Session 2 occurred in the order listed above, except that the order of the everyday categories theory-drawing and conceptual centrality tasks was counterbalanced. Within each task, the order of the five mental disorders was randomized. Participants were presented with one of two different randomized orders of symptoms and features within each disorder for the disorder-defining, conceptual centrality, and everyday categories conceptual centrality tasks. For all other tasks, the order of symptoms and features within each disorder was randomized for each participant.

In the familiarity-rating tasks, participants were asked to report their best estimates of how many patients they had seen with each of the five disorders within the past year. They were also asked to rate on a 9-point scale their familiarity with each disorder compared with the average clinical psychologist (for expert participants) or compared with the average clinical graduate student (for trainee participants).

Next, in the disorder-defining task, participants were given the names and characteristic symptoms of the five disorders on five separate sheets. The symptoms included both the *DSM–IV* diagnostic criteria and the associated symptoms listed in the manual for the disorder. Participants were told that "we would like to understand exactly what you personally consider the symptoms of each disorder to be." They were asked to read the

[6] See the General Discussion section for a discussion of the differences in theoretical orientation between expertise groups. Also, because the sample sizes of the different orientation groups are small, it was not possible to run analyses based on theoretical orientation. (However, please see Experiment 4 for an analysis collapsing the results from all three hypothetical patient experiments reported in this article.)

symptoms of each disorder and to add, delete, group, or split them as they saw fit. The on-site experimenter revised the materials for the other Session 1 tasks according to each participant's responses on this task.

In the theory-drawing task (Task I in Figure 1), they were given the names of the five disorders, each on a separate sheet of paper. They were also given five sets of small paper slips, each bearing the name of one symptom or characteristic feature of that disorder. The participants' task was to arrange the symptoms on each sheet and draw arrows (cause to effect) between any symptoms they felt were causally connected. Participants were also told that, to simplify the drawings, they should feel free to group symptoms and to draw arrows between those groups if they believed that each symptom in one group causes each symptom in the other group. They were also asked to assign a causal strength to each arrow on a 5-point scale. They were encouraged to again add, delete, group, or split symptoms as they saw fit. Immediately following this task, participants were asked to rate, on a 9-point scale, how confident they were about their drawings of causal relations for each disorder in the theory drawing task.

In the conceptual centrality task (Task II in Figure 1), participants were asked, "How easily can you imagine a patient who has all the symptoms of [disorder X] except that he or she does not have [symptom Y]?" for each disorder and symptom of that disorder (Sloman & Ahn, 1999; Sloman et al., 1998).[7]

In the second session, participants received two or three hypothetical case studies for each disorder.[8] These case studies concerned patients who had either three causally central symptoms present, three causally peripheral symptoms present, or three isolated symptoms present, according to that participant's own ratings in the theory drawing task in the first session. The hypothetical patients with isolated symptoms were constructed using symptoms that were not related to any other symptoms. The hypothetical patients with causally central and those with causally peripheral symptoms were developed by applying Equation (1) to the ratings obtained in the theory drawing task.

Specifically, a pairwise dependency matrix for each participant and disorder was first determined from participants' responses in the theory-drawing task. That is, the strengths participants assigned to the causal arrows constituted the cells of the matrix.[9] Model-predicted causal centrality rank orders were set to the initial arbitrary value of 0.5, following the procedure of Sloman et al. (1998).[10] For a more concrete example, consider the sample causal diagram, drawn by an expert clinician participant, shown in Figure 2 for anorexia nervosa. Figure 2 also shows the output given by Equation (1) after three iterations. "Refuses to maintain weight" and "lives in industrialized society" caused the most symptoms with the greatest causal strength and also had the highest model output value. Symptoms such as "absence of the period for 3+ months" and "concern about eating in public" are part of the causal diagram but did not cause any other symptoms and had the lowest model output values. In this way, rank orders of causal centrality were determined for the symptoms in each individual's theory for each disorder.

On the basis of these rank orders, we initially selected the three highest and three lowest ranked symptoms for a patient with causally central symptoms and a patient with causally peripheral symptoms, respectively. Then, if the number of *DSM–IV* diagnostic criteria was not equated across patients, we replaced a symptom in the patient with the next rank-ordered symptom. In this way, items were balanced within each disorder for the number of *DSM–IV* diagnostic criteria they contained. This manipulation was particularly important because, under these conditions, participants following *DSM–IV* guidelines should not see the hypothetical patients as different from each other from a diagnostic standpoint. Items were also balanced within each disorder for reference to gender (i.e., if gender had to be revealed in one patient, all patients for that disorder were said to be of that gender also).

For each hypothetical patient, participants were explicitly told that no other symptoms were present in the patient. Next, they were asked "what is the likelihood, in your opinion, that a patient with the following characteristics has [disorder X]?" Participants made their judgments on a scale of 0–100 ranging from *very unlikely* to *very likely* (Task III in Figure 1).

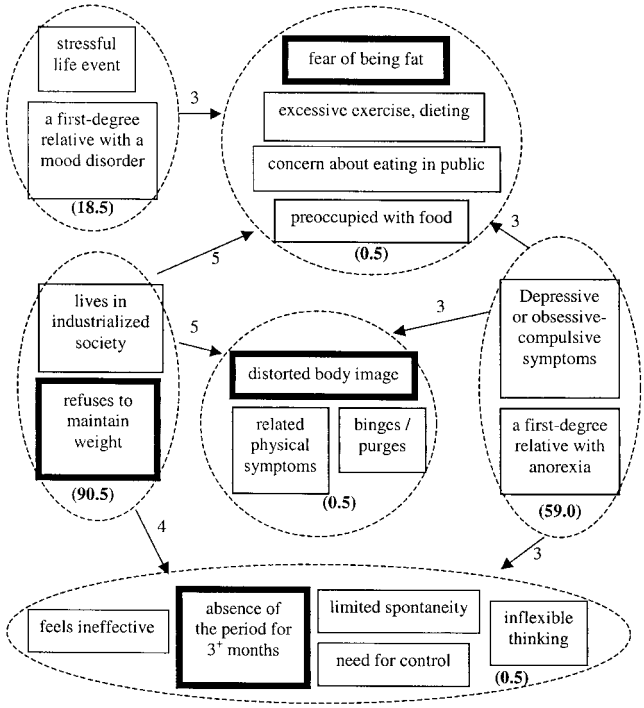

Figure 2. One participant's causal drawing of anorexia nervosa in Experiment 1. Dotted circles indicate groupings drawn in the diagram by the participant. Causal centrality predictions generated by Equation (1) after three iterations are shown in parentheses for each symptom group. Diagnostic criteria, shown in boldfaced boxes, are reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders,* Fourth Edition. Copyright 1994 American Psychiatric Association.

Participants were next asked to complete a theory-drawing task and a conceptual centrality task for four everyday categories: *robin*, *apple*, *acoustic guitar*, and *chair*, as used in Sloman et al. (1998). If expertise effects were found in the mental disorder domain, they could be said to occur because of a general response bias (e.g., older people are more

[7] Sloman et al. (1998) determined that people's theories predict their conceptual centrality ratings (including this specific ease-of-imagining rating) but not their base rate estimates (including ratings of cue and category validities).

[8] Only two case studies were used when a participant did not leave any symptoms isolated in a causal drawing. This occurred 62.8% of the time in Experiment 1, 57.0% of the time in Experiment 2, and 68.4% of the time in Experiment 4.

[9] It should be noted that a distinction in coding was made between joint and separate causes. For instance, if X and Y were each said to cause Z separately, then the relational strength listed was entered for each link. In contrast, if X and Y were said to cause Z jointly, then the relational strength listed was divided by the number of "causal" symptoms and the quotient was entered for each link.

[10] The matrix multiplication was performed repetitively until the Spearman's rank correlation of the model-predicted causal centrality ratings and the conceptual centrality ratings given directly by the participants converged to its terminal stable value as in Sloman et al. (1998).

theory-based regardless of domain). To eliminate this possibility, we included these tasks. No expertise effects were found in either domain (see Kim, 2002, for more details), however, so the results of this task are further discussed in this article. This task also served as a filler task to prevent a ceiling effect in participants' level of recall.

Next, participants were asked to recall as many symptoms as they could from the patients in the hypothetical case studies at the beginning of that day's session, cued only by the disorder name (Task IV in Figure 1).

Results and Discussion

The primary concerns were whether clinicians represent disorders as theories and whether these theories affect the way they reason about disorders in a systematic way. The bottom line for the large number of results presented here is that both experts and trainees showed the causal status effect in the hypothetical patient task, recall task, and conceptual centrality task. The sections that follow report the results roughly in the order of theoretical importance.

Hypothetical patient diagnosis. The main question was whether participants were more likely to diagnose hypothetical patients as having a disorder if those patients had causally central as opposed to causally peripheral or isolated symptoms, according to the participant's theory of the disorder. A 2 (expertise; clinicians vs. graduate students) × 3 (item type; causally central, causally peripheral, isolated) analysis of variance (ANOVA) revealed a main effect for item type, $F(2, 38) = 27.5$, $MSE = 157.90$; $p < .01$; $\eta^2 = .59$. Hypothetical patients with causally central symptoms were rated as more likely to have the disorder in question than were patients with causally peripheral symptoms (Ms = 61.0 and 42.0, respectively); $t(20) = 4.54$; $p < .01$; $\eta^2 = .51$, who, in turn, were rated as more likely to have the disorder than patients with isolated symptoms ($M = 34.5$); $t(20) = 3.32$; $p = .01$; $\eta^2 = .36$. The results went in the expected direction for all disorders and comparisons except for the causally peripheral versus isolated patients in phobia only. No main effect for expertise ($p > .1$; $\eta^2 = .09$) or interaction of Expertise × Item Type ($p > .2$; $\eta^2 = .07$) was found, indicating that both clinicians and graduate students showed the causal status effect in the diagnosis task. (See Figure 3.)

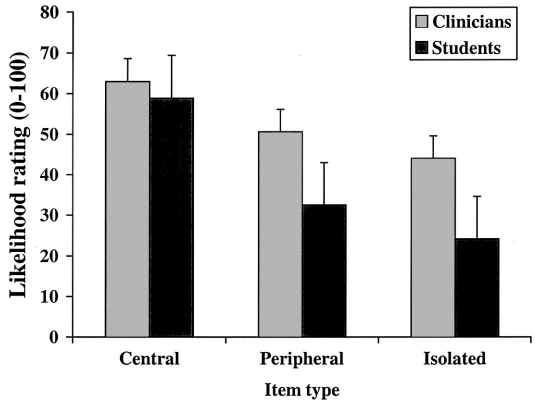

Figure 3. Clinical psychologists' and clinical psychology graduate students' likelihood ratings of mental disorder category membership for hypothetical patients in Experiment 1. Error bars indicate standard errors.

Free recall. The question of interest for this task was whether participants would better recall causally central as opposed to causally peripheral or isolated symptoms, according to the participant's theory of the disorder, that they saw earlier in the task. A 2 (expertise; clinicians vs. graduate students) × 3 (symptom type; causally central, causally peripheral, isolated) ANOVA revealed a main effect for symptom type, $F(2, 38) = 4.90$, $MSE = .04$; $p = .01$; $\eta^2 = .21$. Causally central symptoms ($M = 67.0\%$ recalled) were more frequently recalled than both causally peripheral symptoms ($M = 51.0\%$ recalled); $t(20) = 2.88$; $p < .01$; $\eta^2 = .29$, and isolated symptoms ($M = 43.6\%$ recalled); $t(20) = 2.66$; $p < .02$. The difference in recall between causally peripheral and isolated symptoms, although in the predicted direction, was not significant, $t(20) = .52$; $p = .6$; $\eta^2 = .01$. No main effect for expertise ($p = .8$; $\eta^2 = .002$) or interaction of Expertise × Symptom Type ($p = .8$; $\eta^2 = .01$) was found, indicating that both clinicians and graduate students showed the causal status effect in the free-recall task. (See Figure 4.)

Conceptual centrality. Equation (1) was again used to test the hypothesis that the causal centrality of a symptom predicts its conceptual centrality (i.e., ease-of-imagining responses). Because the conceptual centrality questions were in the negated form of "how easily can you imagine a person with [disorder X] who does not have the symptom of [Y]?" these scores were each subtracted from 100. Thus, the higher the number is, the more conceptually central the symptom is to the disorder.

The analyses in this section deal with the diagnostic criteria symptom data only, because it is these that the *DSM–IV* treats as nondifferentially weighted in diagnosis (with the exception of depressed mood and anhedonia for major depressive episode). Spearman's rank-order correlations were run between the reported and model-predicted (see *Method* section for a description of how these predictions were obtained) conceptual centralities for each disorder. For instance, the rank orders generated from the model output shown in Figure 2 were correlated with the rank orders obtained from the conceptual centrality task for those symptoms. Because participants had been allowed to define the disorders for themselves in the initial disorder defining task, the analyses were run for each individual participant.

The overwhelming majority of participants showed positive correlations across disorders (17 of 20; 9 of the 11 clinicians and 9 of the 10 graduate students). The median individual overall correlation coefficients for clinical psychologists and graduate students were .41 (range: −.12–.50) and .27 (range: 0–.62), respectively.[11] Because only *DSM–IV* diagnostic criteria were used in this analysis, a *DSM* based model would predict correlation coefficients of 0. A one-sample t test of the Fisher-converted overall r_s scores revealed that the average correlation coefficient differed from 0, $t(20) = 6.62$; $p < .01$; $\eta^2 = .67$. Even after dropping major depressive episode from the analyses, because two of the symptoms are given different weight in diagnosis by the *DSM–IV* (APA, 1994) than the others, we found that this result holds up, $t(20) = 4.27$; $p < .001$; $\eta^2 = .48$.

[11] As in Sloman et al.'s (1998) original study, overall correlation coefficients across disorders were calculated by taking Fisher's z-transformation of each correlation, averaging across the z-transform scores, and converting the mean score back to units of correlation. Because this is a composite r based on different numbers of pairs of scores, no p value can be reported.

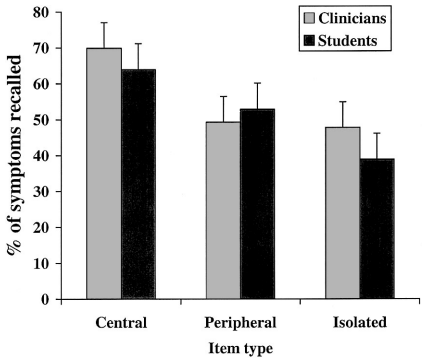

Figure 4. Percentages of symptoms correctly recalled from hypothetical patients seen prior to a time delay in Experiment 1. Error bars indicate standard errors.

Broken down by expertise, both clinical psychologists, $t(10) = 4.87$; $p < .01$; $\eta^2 = .69$, and graduate students, $t(9) = 4.26$; $p < .01$; $\eta^2 = .64$, showed significant differences. Broken down by disorder, a majority of participants showed the causal status effect for anorexia nervosa (14 of 19), major depressive episode (14 of 16), and specific phobia (13 of 16).[12] Only a minority showed the effect for antisocial personality disorder (9 of 20) and schizophrenia (5 of 12). This may have been due to lack of power or to too much error variance caused by running analyses on individual participants' data. As is shown in Experiment 2, when common symptom lists are used and analyses can be averaged across participants, these item differences disappear.

Familiarity analyses. The goal of the analyses in this section was to determine whether the degree of causal status effect exhibited by the participants changed as a function of how familiar they were with each particular disorder. Three measures of familiarity (i.e., confidence ratings for the causal diagrams drawn, familiarity with each of the disorders, and number of patients seen in the last year) were normalized and then averaged together for a composite familiarity score for each individual participant per disorder.[13] These composite scores were used to select the two most familiar and two least familiar disorders for each participant. Data were collapsed so that there was a single data set for high familiarity and another for low familiarity. To summarize the results first, familiarity marginally influenced the ease-of-imagining task and did not influence the results from the hypothetical patients and the recall tasks. The detailed results of each are described in the following paragraphs. (See also Kim, 2002, for descriptive statistics on familiarity ratings.)

For the ease-of-imagining familiarity analysis, a 2 (expertise; clinicians vs. graduate students) × 2 (familiarity; high vs. low) ANOVA was carried out on the correlation coefficients mapping the correlation between ease-of-imagining judgments and causal centrality. The critical main effect of familiarity was marginally significant, such that the causal status effect occurred somewhat more strongly for high-familiarity disorders than for low-familiarity disorders ($r_s = .45$ and $r_s = .25$, respectively); $F(1, 18) = 3.43$, $MSE = 0.12$; $p = .08$; $\eta^2 = .16$. There was no Expertise × Familiarity interaction ($p > .5$; $\eta^2 = .02$).

In the hypothetical patients familiarity analysis, a 2 (expertise; clinicians vs. graduate students) × 2 (familiarity; high vs. low) × 3 (item type; causally central vs. causally peripheral vs. isolated) ANOVA was conducted on the disorder membership likelihood ratings. The critical interaction of Familiarity × Item type was not significant ($p > .5$; $\eta^2 = .11$). There was a significant interaction of Expertise × Familiarity, $F(1, 19) = 6.39$, $MSE = 388.76$; $p = .02$; $\eta^2 = .25$. Specifically, clinicians gave higher likelihood ratings of disorder membership than graduate students did to patients with high-familiarity disorders (Ms = 53.3 and 30.4, respectively); $t(19) = 3.40$; $p < .01$; $\eta^2 = .38$. With low-familiarity disorders, on the other hand, there was no difference between the two expertise groups (Ms = 46.8 and 41.7, respectively); $t(19) = 0.69$; $p = .5$; $\eta^2 < .03$. There were no other effects or interactions concerning familiarity (all ps $> .1$; all η^2s $< .1$).

Finally, in the recall task familiarity analysis, a 2 (expertise; clinicians vs. graduate students) × 2 (familiarity; high vs. low) × 3 (item type; causally central, causally peripheral, and isolated) ANOVA was conducted on the recall frequency data. The critical interaction of Familiarity × Item type was not significant ($p > .1$; $\eta^2 = .006$). There was a marginally significant interaction of Expertise × Familiarity, $F(1, 19) = 3.62$, $MSE = 0.03$; $p = .07$; $\eta^2 = .16$. Mean percentages of symptoms recalled by clinicians were 56.2% for low-familiarity and 48.6% for high-familiarity disorders, and mean percentages of symptoms recalled by graduate students were 51.1% for low-familiarity and 55.3% for high-familiarity disorders (no contrasts were significant; all ps $> .1$; all η^2s $< .2$). There were no other effects or interactions concerning familiarity (all ps $> .4$; all η^2s $< .05$).

Disorder representation. The causal relations that participants drew were quite complex overall. As an indicator of the complexity of participants' theories, the number of causal links for each disorder and participant was counted. Participants often grouped symptoms and drew arrows between those groups, and these cases were coded such that each symptom in one group causes all of the symptoms in the other group. On average, 47.4 links were drawn for a single disorder (range: 0–344). There were no noticeable expertise effects (see Kim, 2002, for more detailed analyses).

Summary. Both clinicians and graduate students clearly showed a causal status effect for mental disorders in the hypothetical patients task, recall task, and ease-of-imagining task. The predicted effect for isolated symptoms was found in the hypothetical patients task, and although the effect was not significant in the recall task, it was in the expected direction. Familiarity was not found to be a strong moderator for the causal status effect in general. Finally, both experts' and trainees' theories were found to be quite complex.

[12] For individual disorders, a correlation coefficient could not be calculated in cases where either the participant elected not to draw a causal diagram, drew the diagram such that all diagnostic criteria symptoms were given the same causal centrality, or gave all diagnostic criteria symptoms the same conceptual centrality rating.

[13] Confidence and familiarity (a composite of familiarity ratings and the number of patients seen) were highly correlated in all three experiments: Experiment 1, $r(21) = .47$; $p = .03$; Experiment 2, $r(20) = .67$; $p < .01$; Experiment 4, $r(19) = .75$; $p < .01$. Thus, we felt reasonably comfortable collapsing the three types of ratings into a composite to represent overall familiarity.

Experiment 2: Converging Evidence for the Use of Causal Centrality in Clinical Reasoning

In this experiment, we expanded the generality of our findings using modified procedures. The theory-drawing task was changed to measure participants' theories about all symptom–symptom relations, not restricting the measure to causal relations alone. In Experiment 1, we asked participants to draw an arrow if they believed a symptom causes another symptom. However, relations such as "allow," "determine," "increase," or "lead to" appear to imply causality as well (Ahn et al., 2002), even though the word *cause* is not explicitly used in these cases. We would expect to find a causal status effect using these relations as well, but the instructions used in Experiment 1 may not have captured all these relations. In addition, causal, explanatory relations are only a small subset of the many possible types of relations, and a number of other noncausal relations, such as "precedes," "is an example of," or "is a cure for," might also be relevant to clinical theories of mental disorders. By allowing participants to draw all kinds of relations, we could measure how prevalent causal relations (or relations implying causality) were in clinicians' theories.

There were also several other notable modifications. The disorder-defining task was dropped and participants were provided with a list of standard symptoms defined by the participants in Experiment 1. This change was made to eliminate the possibility that expertise effects were not detected because all the participants were using different symptom lists. Standardized symptom lists allowed for direct comparisons between experts and trainees. Moreover, such lists allowed us to run consensus analyses determining the degree to which participants agreed on the relations they drew among the symptoms. In this way, we could test the *DSM–IV*'s assumption that clinicians' theories are very diverse.

The measures taken in the conceptual centrality, hypothetical patient, and memory tasks of Experiment 1 were also modified to broaden the scope of the findings (see Figure 1 for a summary). In the conceptual centrality task, the ease-of-imagining question was changed to a diagnostic importance question. In the hypothetical patients task, the diagnosis-likelihood question was changed to a typicality rating question as in Cantor et al.'s (1980) study. Finally, instead of a recall task, a symptom recognition task was implemented.

Method

Participants. Participants were another group of 14 experienced clinicians in private practice in the Nashville, Tennessee area and 6 clinical psychology interns at Vanderbilt or Yale University.[14]

Descriptive statistics were compiled from participants' responses to the therapist history questionnaire, identical to the one used in Experiment 1 (see Table 1). Thirteen clinicians held doctorates of philosophy in clinical psychology, and 1 clinician held a doctorate of education. All 14 experts were licensed psychologists. The interns included 5 current clinical psychology interns and 1 fourth-year graduate student about to begin an internship.

Materials and procedure. Predefined symptom lists for the same five disorders in Experiment 1 were compiled by dropping a symptom from the list if it was dropped by over 50% of the experts and over 50% of the trainee participants in Experiment 1. These predefined lists were used for all tasks and participants in Experiment 2.

Participants each engaged in two sessions, spaced 10–14 days apart, to complete eight tasks. In the first session, they completed the following: disorder familiarity ratings, a theory-drawing task and a diagnostic importance task. In the second session, they completed a hypothetical patient task, the study phase of a recognition task for everyday words, an everyday categories theory-drawing task, an everyday categories category importance task, the recognition phase of the recognition task for everyday words, and a symptom recognition task. The total time to complete both sessions ranged from approximately 2 to 4 hr. The counterbalancing and randomization procedures were the same as in Experiment 1.

The theory-drawing task was the same as in Experiment 1, except that participants were asked to draw any kind of relations between symptoms they saw fit. Participants were asked to consider using, but not to limit themselves to, the following relations: "is a subset of," "is an example of," "precedes," "co-occurs with," "is a precondition for," "causes," "jointly cause," "affects," "determines the extent of," "increases," "decreases," "is a catalyst for," "is used as a defense against," or "is a cure for." An analysis was conducted to determine the proportion of causal versus noncausal relations in participants' drawings. The distinction between causal and noncausal links was made according to data reported in Ahn et al. (2002). In that study, participants were asked to rate different types of links as to "the extent to which the term implies that there is a (positive or negative) causal link underlying the two items" (p. 113). In the current analysis, noncausal links were defined as those links with a mean rating lower than the midpoint rating in that study (i.e., "allows," "can support," "discourages," "enhances," "helps," "is a subset of," "is an example of," "minimizes," "precedes," "raises chances of," "uses"). Causal links were defined as those links with a mean rating higher than the midpoint rating (i.e., "affects," "causes," "decreases," "determines," "enables," "increases," "is a result of," "is dependent on," "is a precondition of," "leads to," "requires"). On the basis of this classification, we found that a surprising 97.3% of links drawn in the current study were actually causal or implied causal links. Therefore, instead of introducing a new term (i.e., *relational centrality*), we continue to use the term *causal centrality* in referring to the centrality of a feature in participants' theories. Most of our manipulations in the current study were based on all directional links the participants drew, but again, the vast majority of these links were causal. We also report additional analyses conducted only on relations classified as causal, demonstrating that the results are truly due to causal relations.

In addition to the instructions to draw any kinds of relations, there were other minor changes in the theory-drawing task compared with the instructions in Experiment 1. Because the task had become more complicated with the addition of different types of possible relations, we reduced the complexity of the strength-rating portion of the task by changing the 5-point scale to a 3-point scale (1 = *weak*, 2 = *moderate*, 3 = *strong*). Immediately after the participants had drawn their theories for the disorders, they were asked to go back and explain all the links. With the prior consent of the participants, the experimenter videotaped each diagram and the participant's hand pointing to the symptoms and links during these explanations. These videotapes were later used to assist in coding the data when anything about the drawings alone seemed unclear.

During the first session, the diagnostic importance of each symptom to the disorder was also measured. For each symptom, participants answered the question, "how important is the symptom of [Y] in diagnosing a person with [disorder X]?" with a rating on a scale of 0–100 (0 = *very unimportant*; 100 = *very important*).

For the hypothetical patient task, participants received hypothetical patient descriptions consisting of a set of either three causally central symptoms, three causally peripheral symptoms, or (if there were enough

[14] More interns could not be recruited because that population in general has a heavy workload, and few interns could spare the time to participate. The lack of power, however, is not problematic because we collapsed the data across all three major experiments later to address the possibility of expertise effects. This analysis is reported in Experiment 4.

isolated symptoms) three isolated symptoms, as in Experiment 1. Causal centralities were determined according to Equation (1) as in Experiment 1; however, these calculations were carried out by collapsing the data across all types of relations, treating them as a single measure of relational dependency as in Sloman et al. (1998).[15] Again, the number of diagnostic criteria symptoms was equated between patients so that diagnoses based strictly on the *DSM–IV* would not differentiate them. Following Cantor et al. (1980), for each patient, participants were asked, "how well, in your opinion, does a patient with the following characteristics fit in the diagnostic category of [X]?" on a scale of 0–100 (0 = *very poorly*, 100 = *very well*).

Several filler tasks followed. Participants studied word lists and took a yes–no recognition test of their memory for those words. They also completed the everyday categories theory-drawing task and everyday categories category importance task, as in Experiment 1; however, the modifications made to the mental disorder theory-drawing task and the category importance task in Experiment 2 were also made to these tasks. Once again, the dual role of these tasks was to prevent a ceiling effect in the subsequent symptom recognition task and to demonstrate that any potential expertise effects in the mental disorders domain did not indicate a domain-general expertise effect. No such expertise effects were found in these tasks, so the results will not be discussed in this article. (See Kim, 2002, for the results.)

Finally, participants received a recognition task for mental disorder symptoms in which they were asked to classify symptoms on a list as old or new on the basis of whether they had seen them earlier in the hypothetical patients task. The list included 30 relationally central and 30 relationally peripheral or isolated symptoms constructed individually according to each participant's theories. Causally peripheral symptoms and isolated symptoms had to be combined into one group because there were not enough symptoms in total to build a new set of each type of symptom alone. Of the symptoms in each group (central vs. peripheral or isolated), half were old and half new. As in the hypothetical patients task, the number of diagnostic criteria symptoms was balanced between all the different groups. The rest of the procedure for the recognition task was the same as in Roediger and McDermott (1995). The time lag between the hypothetical patients task and the symptom recognition task was a mean of 62.9 min (range: 50–120 min).

Results and Discussion

In this experiment, the primary goal was to replicate the causal status effect using different measures of theories, conceptual centrality, and memory. To summarize the large number of results in this section, we found that both experts and trainees showed a causal status effect in the hypothetical patient task, recognition task, and diagnostic importance task, and that there was a statistically significant degree of agreement on the theories. The results that follow are reported in approximate order of importance.

Hypothetical patients. A 2 (expertise; clinicians vs. interns) × 3 (item type; causally central, causally peripheral, or isolated) ANOVA revealed a main effect of item type, $F(2, 36) = 15.66$, $MSE = 172.84$; $p < .01$; $\eta^2 = .47$. Contrasts showed that patients with causally central symptoms ($M = 72.7$) were judged as more typical of the disorder than patients with causally peripheral symptoms ($M = 58.4$); $t(19) = 4.3$, $p < .01$; $\eta^2 = .50$, who in turn were judged as more typical than patients with isolated symptoms ($M = 47.2$); $t(19) = 2.7$, $p < .02$; $\eta^2 = .27$. The results were in the expected direction for all disorders and comparisons except for the causally peripheral versus isolated patients in schizophrenia only. Neither a significant main effect of expertise ($p > .9$; $\eta^2 < .001$) nor an interaction of Expertise × Item Type ($p = .9$; $\eta^2 < .01$) was found. (See Figure 5.)

Recognition: Mental disorder symptoms. The hit rates, or the percentage of old symptoms correctly identified as old, did not differ between item types (Ms = 85.4% and 89.3% for causally central and for causally peripheral or isolated symptoms, respectively), probably because of a ceiling effect. A 2 (expertise) × 2 (item type; causally central, causally peripheral or isolated) ANOVA revealed that neither of the main effects were significant, nor was the interaction (all ps > .1; all η^2s < .1). In contrast, as shown in Figure 6, participants were much more likely to falsely recognize new, causally central symptoms as symptoms that they had seen before (23.3%) than new, causally peripheral or isolated symptoms, 13.2%; $F(1, 18) = 4.74$, $MSE = 0.01$; $p = .04$; $\eta^2 = .21$. Neither the main effect of expertise ($p = .9$; $\eta^2 = .001$) nor the interaction of Expertise × Item type ($p > .1$; $\eta^2 = .12$) was significant. Additional analyses confirmed that participants showed greater sensitivity to causally peripheral or isolated symptoms ($d' = 2.62$) than to causally central symptoms, $d' = 2.04$; $F(1, 18) = 7.33$, $MSE = 0.33$; $p = .01$; $\eta^2 = .29$. Thus, participants were less able to distinguish between presented and nonpresented causally central symptoms. Just as false memory might be an issue for patients with psychological disorders or problems (Loftus & Ketcham, 1994), therapists may be biased to falsely remember having seen symptoms in their patients that are central to their personal theories about the disorder. Such findings are in accord with research showing that people often rely on their background knowledge to recall the source of a memory (e.g., Mather, Johnson, & DeLeonardis, 1999; see also Johnson, Hashtroudi, & Lindsay, 1993, for a general theoretical discussion of false memory) and with previous findings showing an effect of schema on false recognition (Bower, Black, & Turner, 1979).

Theory agreement. Kendall's coefficients of concordance were calculated to determine how much participants' rank-ordered lists of causal centrality (and, thereby, theories as operationalized in the current study) agreed with each other. Kendall's Ws were all significant, ranging from .23 to .38 across the five disorders (all ps < .0001). Thus, although the values of the Ws themselves were not particularly high, there was at least a significant level of agreement among these clinicians as to how the symptoms of the disorders interact with each other. This significant concordance of theories occurred regardless of the diversity of the participants' theoretical orientations, as reported in Table 1. Because of this statistically significant consistency in theories between participants, we were able to construct an average dependency structure

[15] Links in the form of "X causes Y," "X precedes Y," "X is a precondition for Y," "X affects Y," "X determines the extent of Y," "X increases Y," "X decreases Y," "X is a catalyst for Y," and "X is a cure for Y" were coded in a straightforward manner (i.e., as X → Y). Links in the form of "X is a subset of Y" and "X is an example of Y," in contrast, were reversed (i.e., as Y → X). Links marked "X co-occurs with Y" were ignored in these analyses because we interpreted these as atheoretical relations rather than explanatory relations. Ambiguous links, which included "X is used as a defense against Y" and any links not preconsidered by the authors, were decided on a case-by-case basis according to the independent votes of the three coders. Between each participant's first and second sessions, the data were coded by one trained coder and rechecked twice independently by the two authors. All other coding procedures adhered exactly to those used in Experiment 1.

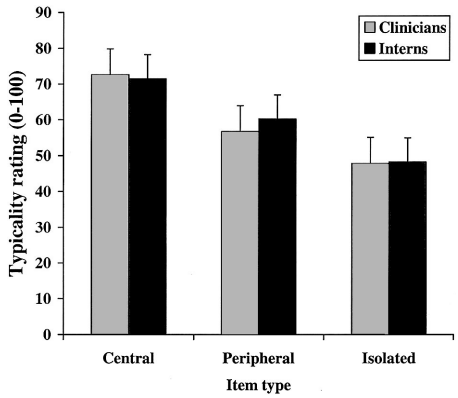

Figure 5. Clinical psychologists' and clinical psychology interns' typicality ratings for hypothetical patients in Experiment 2. Error bars indicate standard errors.

for each disorder that was used to run diagnostic importance analyses collapsed over participants, as seen in the next section.

Figures 7, 8, 9, 10, and 11 show average diagrams for each of the five disorders, with the averaged strengths for directional links between symptoms. (The absence of a link between two features was coded as 0.) For clarity of visual presentation, most strengths lower than 1.0 were omitted from the figures, but all strengths were included in the statistical analyses. The names of the symptoms are abbreviated because of space limitations (full names of the diagnostic criteria can be found in Table 2).

Diagnostic importance. Table 2 presents the mean diagnostic importance ratings of the diagnostic criteria for all five disorders. The main question in this section was whether causal centrality predicts diagnostic importance. For instance, participants on average believed that in anorexia nervosa (Figure 7) "disturbed experience of body shape or denial of the problem" causes many symptoms, including the fear of being fat and excessive exercise and dieting. Note that this symptom is also diagnostically important (92.1) as shown in Table 2. On the other hand, "absence of the period (in women) for 3+ menstrual cycles," another diagnostic criterion for anorexia nervosa, was rarely judged to cause any other symptoms of that disorder, as is shown in Figure 7, and is also less diagnostically important (74.4) as shown in Table 2. These symptoms are two of the four diagnostic criteria symptoms for anorexia

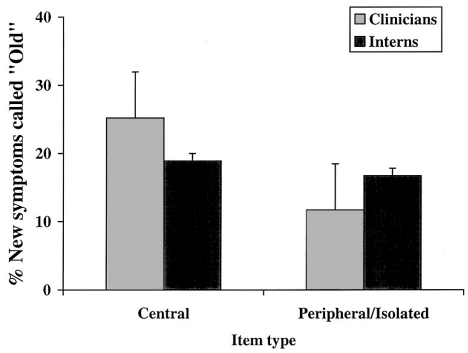

Figure 6. False alarms for the symptom recognition task in Experiment 2. Error bars indicate standard errors.

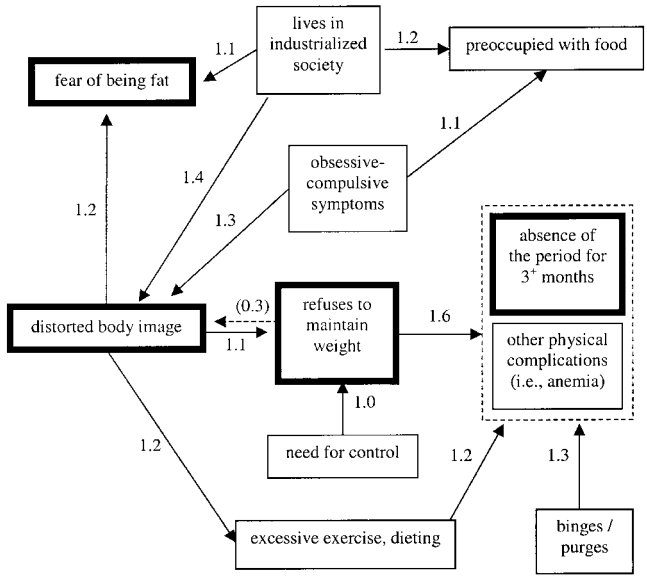

Figure 7. A composite of all participants' drawings of anorexia nervosa in Experiment 2. The dotted box encloses symptoms with identical causal links. Causal centrality predictions generated by Equation (1) after three iterations are shown in parentheses for each symptom group. Diagnostic criteria, shown in boldfaced boxes, are reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders,* Fourth Edition. Copyright 1994 American Psychiatric Association.

nervosa that the *DSM–IV* lists as necessary to a diagnosis of the disorder. This implies that our participants are deviating from the *DSM–IV*'s prescribed model of giving the diagnostic criteria equal weight in diagnosis and that diagnostic importance may instead be a function of causal centrality. The more formal analyses are detailed later in the article.

Again, the analyses in this section deal with only the diagnostic criteria data because of their implications for the *DSM–IV*. This time, because all participants were using the same symptom lists, analyses averaged across participants were entered into Equation (1) to obtain averaged causal rank orderings of the symptoms. These ranks (shown under "Centrality" in Table 2) were correlated with the ranks derived from the averaged diagnostic importance ratings for each disorder (shown under "Importance" in Table 2) using Spearman's rank-order correlation analyses. Collapsed over disorders, analyses showed there was a high positive correlation (r_s = .85, over all participants; r_s = .80, clinical psychologists alone; r_s = .66, interns alone)[16]. Broken down by disorder, analyses showed all correlations were positive (anorexia nervosa, r_s [4] = 1.00, $p < .01$; antisocial personality disorder, r_s [9] = .70, p = .04; major depressive episode, r_s [9] = .43, p = .2; specific phobia, r_s [5] = .70, p = .19; and schizophrenia, r_s [5] = .90, $p < .04$). Thus, causal centrality is predictive of participants' diagnostic importance ratings for both clinicians and interns.

[16] As in Sloman et al. (1998), we obtained average correlations across disorders by taking the mean of the Fisher's *z*-transformation of each disorder's *r* value and converting that mean back to units of correlation. Thus, sample sizes and *p* values cannot be reported for any correlation coefficients collapsed across disorders.

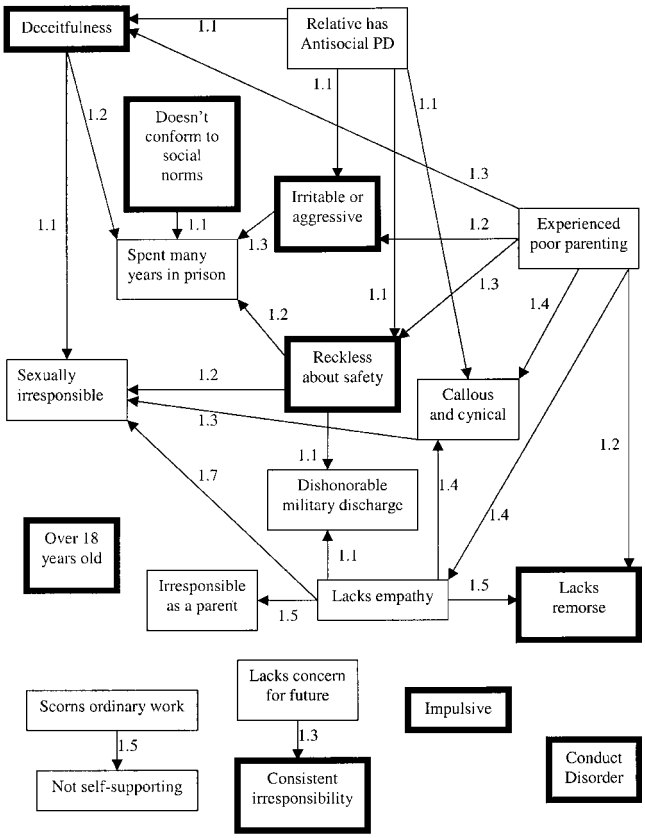

Figure 8. A composite of all participants' drawings of antisocial personality disorder (PD) in Experiment 2. Diagnostic criteria, shown in boldfaced boxes, are reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition. Copyright 1994 American Psychiatric Association.

We also examined data from only those participants who performed the diagnostic importance task followed by the theory-drawing task to rule out the possibility that priming of theories because of task ordering was responsible for driving the effect. Collapsed over disorders, data from this subgroup still showed a clearly positive correlation overall (r_s = .49). Compared with the correlation coefficient for data only from participants who completed the theory-drawing task first (r_s = .77), there was no significant difference between the two, $t(4) = 1.04$; $p = .4$; $\eta^2 = .21$ (at the $\alpha = .05$ level, observed power = .13).

The diagnostic importance analyses were rerun on only the causal or implied causal links to ensure that eliminating noncausal relations would not eliminate the causal status effect. Collapsed over disorders, analyses showed the averaged diagnostic importance ratings were not notably different from the previous analyses including noncausal directional links (r_s = .79, over all participants; r_s = .84, clinical psychologists alone; r_s = .62, interns alone).

Familiarity analyses. The goal of the analyses in this section was to determine whether the degree of causal status effect exhibited by the participants changed as a function of how familiar they were with each particular disorder. Familiarity analyses were run as in Experiment 1. Familiarity marginally influenced only the hits in the recognition task, such that hit rates were higher for symptoms of low-familiarity disorders than for high-familiarity disorders (88.8% and 83.8%, respectively); $F(1, 18) = 3.54$, $MSE = 0.04$; $p < .08$; $\eta^2 = .2$. However, familiarity did not influence the false alarms in the same task ($p > .1$; $\eta^2 = .1$), the results from the hypothetical patients task ($p > .9$; $\eta^2 = .002$), or the results from the diagnostic importance task ($p > .7$; $\eta^2 = .008$). (See Kim, 2002, for detailed results.)

Theory representation. All participants had theories about most or all of the disorders. Twelve of the 14 experts and all 6 novices opted to draw theories for all five disorders, and the other 2 experts drew theories for four of the five disorders. Participants' theories were quite complex, with an average of 81.1 links per disorder (range: 0–361; see Kim, 2002, for more detailed analyses).

Summary. Experiment 2 again showed evidence for a causal status effect for reasoning about mental disorders in both clinical psychologists and interns. Participants showed the causal status effect and the predicted effect for isolated versus causal symptoms when rating the typicality of hypothetical patients. They also showed a bias to falsely recognize more symptoms causally central to their theory of the disorder than symptoms causally peripheral or isolated to their theory. Finally, the diagnostic importance task

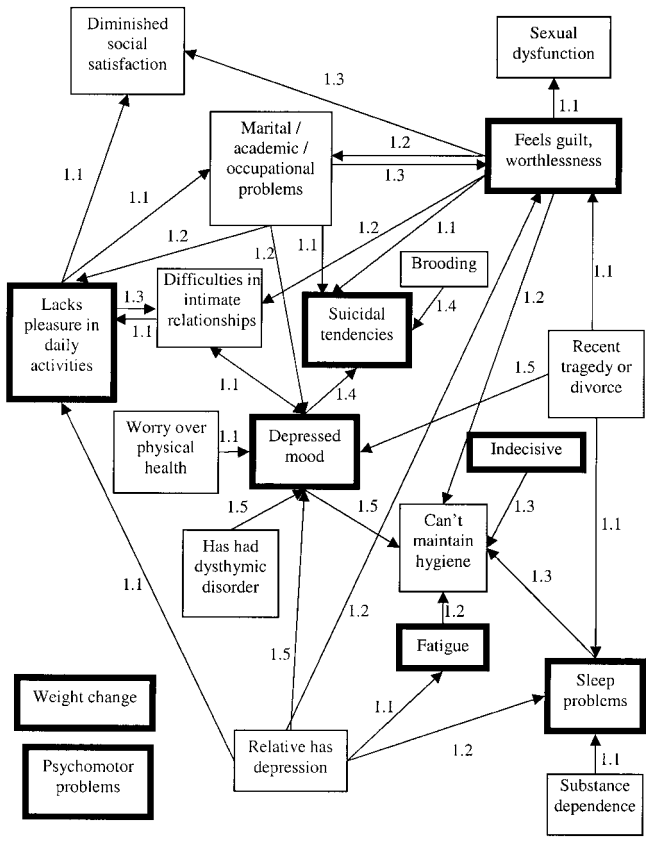

Figure 9. A composite of all participants' drawings of depression in Experiment 2. Diagnostic criteria, shown in boldfaced boxes, are reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition. Copyright 1994 American Psychiatric Association.

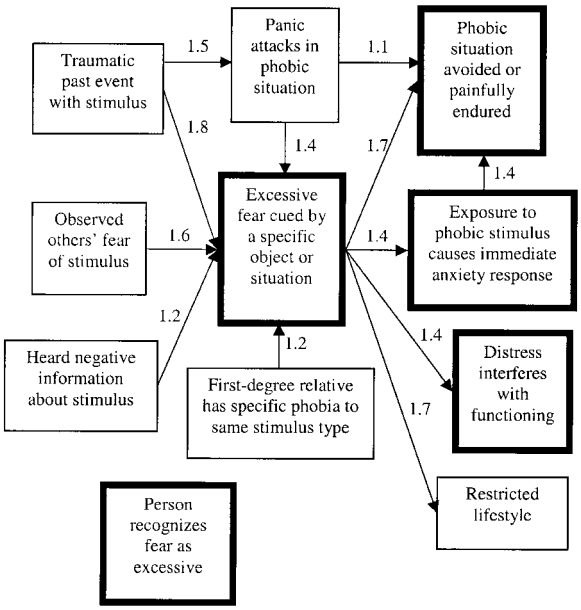

Figure 10. A composite of all participants' drawings of specific phobia in Experiment 2. Diagnostic criteria, shown in boldfaced boxes, are reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition. Copyright 1994 American Psychiatric Association.

showed a causal status effect for clinicians and interns. Participants showed statistically significant agreement on their theories about these five highly familiar disorders, despite their diversity in theoretical orientation. Most of the links drawn by participants were causal or implied causal links, suggesting that the effects found were driven specifically by a causal status effect. All participants had theories for all or most of the disorders, and these theories were quite complex. As in Experiment 1, familiarity with the disorders was not found to be a strong moderator for the causal status effect.

Experiment 3: Between-Group Consensus on Symptom Centrality

One particularly interesting finding in Experiment 2 was that participants' theories agreed with each other to a significant degree despite participants' diverse theoretical backgrounds. Experiment 3 explored the question of whether the causal status of symptoms in expert and novice theories also match up with how naive lay people assign weights to symptoms. A naive theory of psychology is thought to be present in young children (Carey, 1985) as well as in adults (e.g., Furnham, 1995; Matschinger & Angermeyer, 1996). It is possible that even longtime clinicians may find it difficult to adopt theories of behavioral phenomena that run strongly counter to the background knowledge of the culture at large. Alternatively (or simultaneously), experts' theories on mental disorders may become disseminated throughout lay culture by means of media or education. Indeed, for major depressive episode, the theories of lay people from our earlier work (Kim & Ahn, 2002) were highly consistent with those of the clinicians and clinical trainees in Experiment 2. The mean rank orders of diagnostic criteria symptoms for major depressive episode obtained in the lay people were highly correlated with those obtained from the clinicians and clinical trainees in Experiment 2 ($r_s = .93, p < .01$). Experiment 3 investigated whether the theories reported by our clinician and intern participants in Experiment 2 predict naive lay people's assignment of symptom weights.

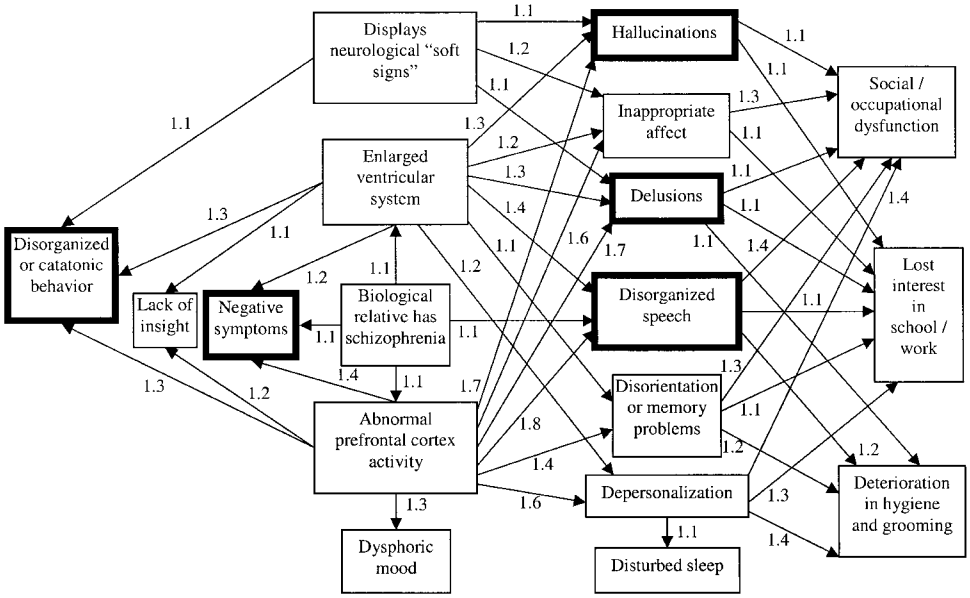

Figure 11. A composite of all participants' drawings of schizophrenia in Experiment 2. Diagnostic criteria, shown in boldfaced boxes, are reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition. Copyright 1994 American Psychiatric Association.

Table 2
Participants' Mean Diagnostic Importance Ratings and Causal Centrality Rank Orders as Calculated by Equation (1) for the DSM–IV *Diagnostic Criteria in Experiment 2*

Disorder and symptom	Centrality	Importance
Anorexia nervosa		
Refusal to maintain body weight at or above minimal levels	1	92.8
Fear of being fat even when underweight	3	91.4
Disturbed experience of body shape or denial of the problem	2	92.1
Absence of the period (in women) for 3+ menstrual cycles	4	74.4
Antisocial personality disorder		
Failure to conform to social norms with respect to lawful behaviors	3	89.3
Deceitfulness	1	89.8
Impulsivity or failure to plan ahead	5	56.3
Irritability or aggressiveness	7	73.5
Reckless disregard for safety of self or others	6	82.4
Consistent irresponsibility	8	82.0
Lack of remorse	4	91.5
Individual is at least age 18 years	9	61.1
Symptoms of conduct disorder occurred prior to age 15	2	82.8
Major depressive episode		
Depressed mood	1	93.0
Lack of pleasure in daily activities	3	86.7
Decrease or increase in weight	9	65.3
Sleep disturbances	4	86.8
Restlessness or unusual slowness	6	77.8
Fatigue or loss of energy	7	85.0
Feelings of worthlessness/excessive guilt	2	84.6
Indecisiveness or difficulty in concentrating	5	85.8
Recurrent thoughts of death/suicide or suicide plan/suicide attempt	8	90.3
Specific phobia		
Marked and persistent fear that is excessive or unreasonable, cued by the presence or anticipation of a specific object/situation	1	96.3
Exposure to the phobic stimulus provokes an immediate anxiety response	2	87.3
Person recognizes that the fear is excessive or unreasonable	5	75.5
The phobic situation is avoided/endured with intense anxiety	3	93.2
Marked distress about having the phobia, or interferes significantly with life functioning	4	88.8
Schizophrenia		
Delusions	2	91.7
Hallucinations	1	89.4
Disorganized speech	5	82.0
Grossly disorganized or catatonic behavior	3	87.2
Negative symptoms (i.e., affective flattening, alogia, or avolition)	4	82.7

Note. Higher numbers correspond to greater diagnostic importance. Rank orders shown for causal centrality were assigned within the diagnostic criteria for each disorder, although the centralities themselves were calculated using $c_{i,t+1} = \sum_j d_{ij} c_{j,t}$ (Equation [1]) based on directional relations with all symptoms (diagnostic criteria plus characteristic symptoms). Symptom descriptions are reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders,* Fourth Edition (*DSM–IV*). Copyright 1994 all American Psychiatric Association.

Method

Participants. For the purposes of this experiment, lay people were defined as any individuals without formal training in clinical psychology. Twenty-three undergraduate students attending either Vanderbilt University or Yale University, who met this criterion, completed the task for payment at the standard rate of $7 per hour.

Materials and procedure. Ten hypothetical patients were developed from the averaged theories of the clinical participants in Experiment 2. One causally central patient and 1 causally peripheral patient were created for each of the five disorders, in the same manner described in Experiment 2. No isolated patients could be created from the averaged theories because no symptoms were left isolated by all of the Experiment 2 participants.

As in Experiment 2, participants were presented with these patients and asked to rate how typical they were of the disorder. For each patient, participants answered the typicality rating question, "how well, in your opinion, does a patient with the following characteristics fit in the diagnostic category of [X]?" on a scale of 0–100 (0 = *very poorly,* 100 = *very well*). Patients were presented to participants in one of two counterbalanced orders.

Results and Discussion

A 2 (item type; causally central, causally peripheral) × 5 (disorder; the five disorders) ANOVA showed that undergraduates judged causally central patients ($M = 75.3$) to be much more typical of each disorder than causally peripheral patients ($M = 29.3$); $F(1, 22) = 154.65$, $MSE = 787.36$; $p < .01$; $\eta^2 = .88$. The results went in the expected direction for all five disorders; moreover, the interaction was not significant ($p > .1$; $\eta^2 < .08$), suggesting that the causal status effect occurred to a comparable extent for all five disorders. Because the undergraduates in this study were never shown the clinicians' theories, they were ostensibly drawing upon their commonsense knowledge to determine how the symptoms should be weighted. Thus, even untrained undergraduates weight causally central symptoms according to the averaged theories of clinicians in Experiment 2. This implies that some broadly held cultural knowledge about disorders seems to be synchronous with the knowledge demonstrated by the experts in our studies, regardless of their specific theoretical orientation.

Experiment 4: The Personality Disorders

Experiments 1 and 2 found evidence for theory-based reasoning in well-known disorders about which experts, trainees, and lay people shared relatively similar theories or causal structures. Would these results generalize to situations in which the consensus in theories is low? In Experiment 4 we first attempted to locate a disorder for which there is extremely low consensus in theories and examined whether theory-based reasoning occurs even under these conditions. We chose personality disorders because clinicians' diagnoses of them are notoriously unreliable (e.g., Hyler, Williams, & Spitzer, 1982; Spitzer, Forman, & Nee, 1979); there might, therefore, be a greater chance of finding a low-consensus disorder among them.

The second goal of Experiment 4 was to more systematically examine the possibility that familiarity with the disorder might moderate theory-based reasoning. The disorders for Experiments 1 and 2 were deliberately selected to be highly familiar even to lay people. Thus, it seems likely that any effects of familiarity—as well as potential expertise effects—may have been obscured be-

cause those disorders varied very little with respect to familiarity levels. Thus, in Experiment 4 we attempted to ensure, by means of informal pretesting, that both familiar and unfamiliar personality disorders were included. It was assumed that this would give more room for observing the possibility of familiarity as a moderator. Once again, because of the low reliability issue, personality disorders seemed particularly likely to include at least one or two unfamiliar disorders that would increase the range of the stimuli.

Experiment 4 investigated these issues by using procedures from Experiments 1 and 2. That is, we first measured the level of agreement between clinicians' theories of personality disorders and then confirmed that the causal status effect holds for these Axis II disorders.

Method

Participants. Participants were another group of 10 experienced clinicians (8 in the Nashville, TN area and 2 in the New Haven, CT area) and 9 clinical psychology graduate students at Yale University. Nine experts held doctorates of philosophy in clinical psychology, and 1 expert held a doctorate of psychology. All 10 experts were licensed psychologists. The graduate students included 4 first-years, 2 second-years, 2 third-years, and 1 clinical psychology intern. See Table 1 for additional demographic information.

Materials and procedure. The personality disorders used as stimuli in this experiment were avoidant and schizotypal personality (lower familiarity) and borderline and obsessive–compulsive personality (higher familiarity). These were selected on the basis of the familiarity ratings of 3 clinical psychology professors to ensure that a wide range of familiarity levels would be covered in this experiment. (See Appendix B for a detailed description of the disorder selection process.)

As in the previous experiments, the symptoms used included both the *DSM–IV* diagnostic criteria and characteristic symptoms from the manual that were not included in its criteria. These symptom lists were compiled by one of two experimenters and triple-checked against the *DSM–IV* by the other experimenter and an independent rater. All participants used the same symptom lists throughout the experiment so that we could examine between-participant consensus in theories afterwards, and select disorders with low consensus.

Participants each engaged in two sessions, spaced 10–14 days apart, to complete seven tasks. In the first session, they completed the following: disorder familiarity ratings, a theory-drawing task, and a conceptual importance task. In the second session, they completed the following: a hypothetical patient task, an everyday categories theory-drawing task, an everyday categories conceptual importance task, and a free-recall task. The total time to complete both sessions ranged from approximately 2 to 4 hr. Randomization and counterbalancing procedures were the same as in Experiments 1 and 2.

All tasks were similar to those completed in Experiment 2, except for the final memory task, which was similar to that completed in Experiment 1. (See Figure 1 for a summary of differences between these experiments.) There were only a few deviations. First, the set of personality disorders described at the beginning of this section was used. Second, in the hypothetical patients task, filler items based on the averaged structure of antisocial personality from Experiment 2 were included in all participants' patients. This was done to raise the cognitive load of the task so that performance on the subsequent recall task could be fairly compared with performance on its counterpart in Experiment 1. Finally, the diagnostic importance task from Experiment 2 was changed to a conceptual importance task to obtain one more converging measure. In the conceptual importance task, participants answered the question "How important is the symptom of [Y] to your concept of [disorder X]?" on a scale of 0–100 (0 = *very unimportant*, 100 = *very important*) for each symptom.

Results and Discussion

This experiment was concerned with two major investigations: Do clinicians rely on their own idiosyncratic theories in diagnosis and reasoning even when their theories do not agree with those of other clinicians and is this causal status effect mediated by familiarity with the disorder? To summarize the results detailed in the section, we found that participants once again showed a causal status effect in the hypothetical patient task, recall task, and conceptual importance task. Furthermore, this effect was found even in schizotypal personality disorder, for which there was no reliable consensus on a theory among the expert participants in this study. Familiarity was shown to mediate the results of both the hypothetical patients task and the conceptual importance task. The results that follow are presented in approximate order of importance: theory consensus, hypothetical patients, recall, conceptual importance, familiarity, and theory representation.

Theory consensus. The data were coded exactly as in Experiment 2, and rank-ordered lists of causal centralities were compiled for each participant and disorder using Equation (1). Of primary interest was whether participants' rank-orderings of symptom centralities (in other words, theories) agreed or disagreed with each other. The clinical psychologists did not significantly agree among themselves as to their theory of schizotypal personality ($W = .09$; $p = .6$), whereas graduate students' agreement only reached marginal significance ($W = .20$; $p = .07$). See Figure 12 for an example of two clinical psychologists' conflicting theories of schizotypal personality disorder. Particularly noteworthy is the central versus peripheral positioning of "excessive social anxiety." The two groups combined were statistically significantly in agreement for schizotypal personality, likely because of the increased sample size, but it should be noted that this was the lowest overall coefficient of all nine disorders tested in the current report ($W = .10$; $p = .04$).

Clinicians, graduate students, and all participants combined had significantly concordant theories for borderline personality (Kendall's $Ws = .30$, .46, and .33, respectively; all ps $< .01$), obsessive–compulsive personality ($Ws = .28$, .21, and .17, respectively; all ps $< .04$), and avoidant personality ($Ws = .26$, .39, and .27, respectively; all ps $< .01$). See Figures 13, 14, 15, and 16 for a summary diagram of participants' theories for each of the four disorders.

Hypothetical patients. Even though participants had more idiosyncratic theories about schizotypal personality disorder, they still relied on their own theories in diagnosis. Analyses run on the schizotypal hypothetical patient data alone showed a significant main effect of item type, $F(2, 34) = 6.74$, $MSE = 269.97$; $p < .01$; $\eta^2 = .28$. Causally central patients were rated as marginally more typical than causally peripheral patients (Ms = 59.1 and 47.2, respectively); $t(18) = 2.05$, $p < .06$; $\eta^2 = .19$, and as significantly more typical than isolated patients ($M = 40.0$); $t(18) = 4.17$; $p < .01$; $\eta^2 = .49$. Causally peripheral patients were not significantly more typical than isolated patients, although the direction of results was in the predicted direction, $t(18) = 1.21$; $p = .2$; $\eta^2 = .08$.

The results were also replicated across all four disorders. A 2 (expertise; clinicians vs. graduate students) × 3 (item type; causally central, causally peripheral, and isolated) ANOVA revealed a main effect of item type, $F(2, 34) = 8.57$, $MSE = 114.89$; $p < .01$; $\eta^2 = .34$. Contrasts showed that patients with causally central

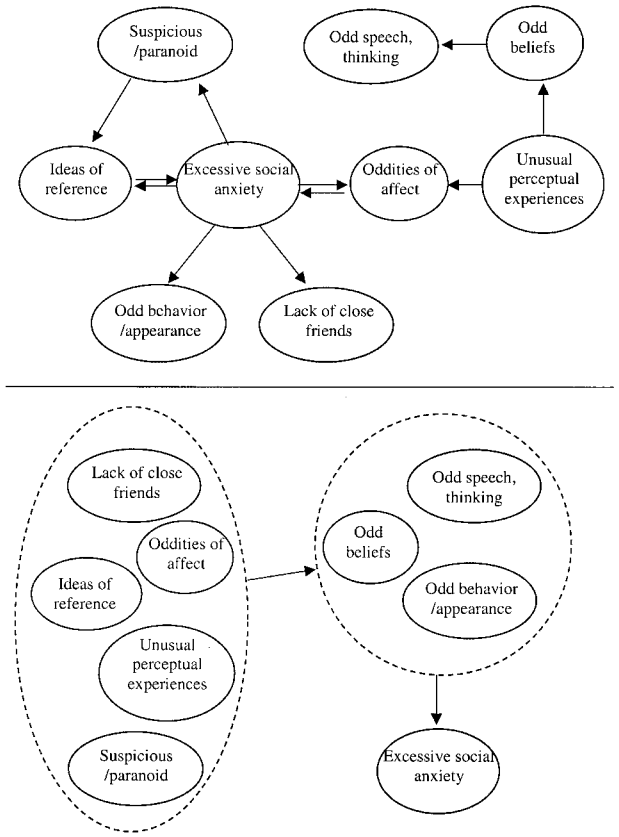

Figure 12. Sample data showing disagreement in theories for schizotypal personality disorder in Experiment 4. Only the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.) diagnostic criteria are pictured here. Dotted circles indicate groupings drawn in the diagram by the participant.

symptoms ($M = 57.8$) were judged as more typical of the disorder than patients with causally peripheral symptoms ($M = 52.5$); $t(18) = 2.25$; $p < .04$; $\eta^2 = .22$, which in turn were judged as more typical than patients with isolated symptoms ($M = 43.5$); $t(18) = 2.48$; $p = .02$; $\eta^2 = .26$. Figure 17 shows the results broken down by experts and trainees. Neither a significant main effect of expertise ($p < .2$; $\eta^2 = .11$) nor an interaction of Expertise × Item Type ($p = .6$; $\eta^2 = .03$) was found. Because antisocial personality disorder was a filler item, it was not included in these analyses (or in Figure 17). A separate analysis revealed that the causal status effect found in Experiments 2 and 3 for antisocial personality disorder was replicated, $t(18) = 2.65$; $p < .02$; $\eta^2 = .28$.

Participants in Experiments 1, 2, and 4 completed a similarly structured hypothetical patients task, so an analysis was run on the collapsed data to test whether theoretical orientation was a moderator of the causal status effect. Collapsing the data across the three experiments allowed for sufficient numbers of participants in each theoretical orientation cell to run the analysis. The different types of theoretical orientation were collapsed into three categories, to give the analysis sufficient power to be a fair test. These categories included psychoanalytic–humanistic (including participants identifying themselves as either psychoanalytic, psychodynamic, or humanistic; $n = 14$), cognitive–behavioral (including cognitive, behavioral, and cognitive–behavioral; $n = 29$), and other (including eclectic, family systems, and any other types of orientation; $n = 17$). A 2 (expertise; experts, novices) × 3 (item type; central, peripheral, isolated) × 3 (orientation; psychoanalytic–humanistic, cognitive–behavioral, other) ANOVA revealed no significant effects or interactions involving orientation (all ps $> .1$; all $\eta^2 < .06$; for the critical interaction of Item Type × Orientation, at the $\alpha = .05$ level, observed power = .51). To rule out the possibility that including the less well-defined group, other, inflated the error term, we ran the same ANOVA again excluding those data. The results were the same, with no significant effects or interactions involving orientation (all ps $> .1$; all $\eta^2 < .03$). Similarly, no effects or interactions involving expertise were significant (all ps $> .2$; all $\eta^2 < .05$; for the critical interaction of Item type × Expertise, at the $\alpha = .05$ level, observed power = .25), even with a much larger sample size. Thus, overall, participants across the three experiments showed the causal status effect in this task regardless of their theoretical bent and expertise.

Free recall. A 2 (expertise; clinicians, graduate students) × 3 (symptom type; causally central, causally peripheral, isolated) ANOVA conducted on the four main disorders revealed a main effect of symptom type, $F(2, 340) = 47.26$, $MSE = 0.02$; $p < .01$; $\eta^2 = .74$. Specifically, causally central symptoms were correctly recalled more frequently than causally peripheral symptoms (Ms = 52.0% and 44.1% recalled, respectively); $t(18) = 2.26$; $p < .04$; $\eta^2 = .22$, which in turn were correctly recalled more frequently than isolated symptoms ($M = 15.4\%$ recalled); $t(18) = 6.22$; $p < .01$; $\eta^2 = .68$. Figure 18 shows the results broken down by level of expertise (again, the filler antisocial personality disorder is not included in this figure). No main effect of expertise ($p = .8$; $\eta^2 = .001$) or interaction of Expertise × Symptom Type ($p > .1$; $\eta^2 = .10$) was found.

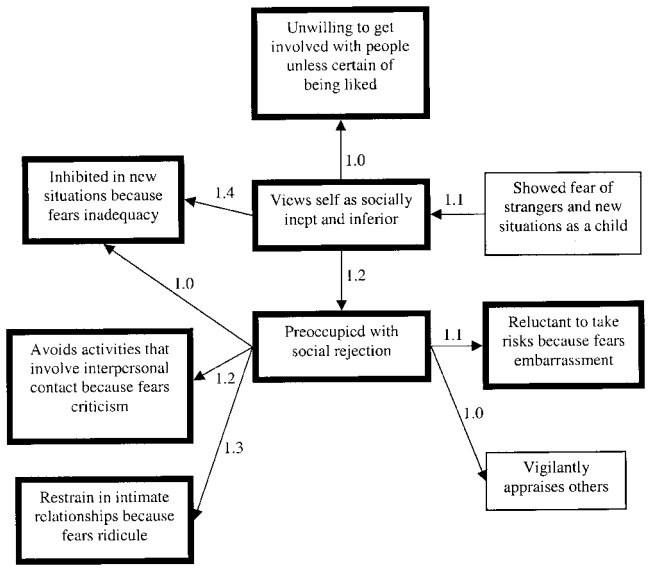

Figure 13. A composite of all participants' drawings of avoidant personality disorder in Experiment 4. Diagnostic criteria, shown in boldfaced boxes, are reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition. Copyright 1994 American Psychiatric Association.

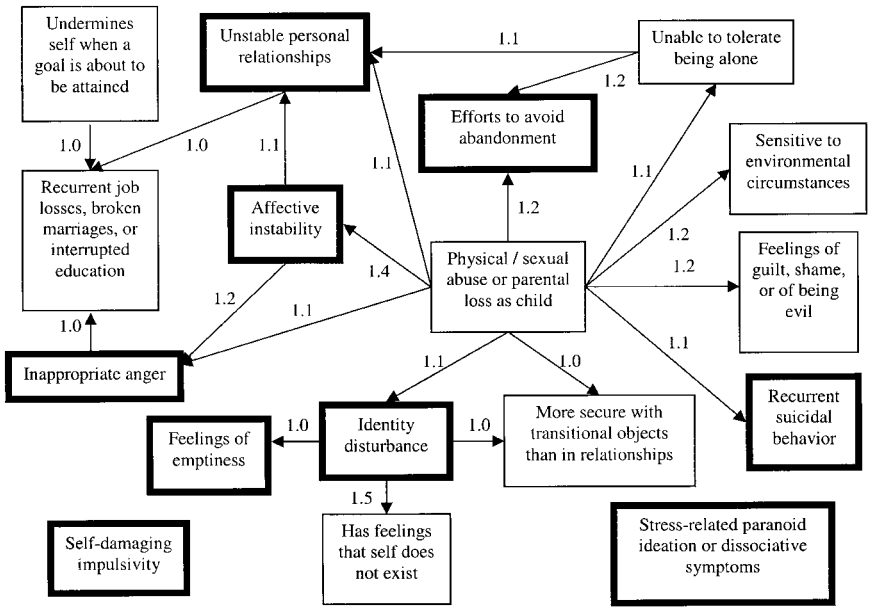

Figure 14. A composite of all participants' drawings of borderline personality disorder in Experiment 4. Diagnostic criteria, shown in boldfaced boxes, are reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition. Copyright 1994 American Psychiatric Association.

The same analysis was conducted on schizotypal personality disorder alone because of the low consensus on causal theories for this disorder. There was a significant main effect of item type, $F(2, 340) = 23.41$, $MSE = 0.05$; $p < .01$; $\eta^2 = .58$. Causally central symptoms were more likely to be recalled correctly than causally peripheral symptoms (Ms = 64.0% and 43.9%, respectively); $t(18) = 2.48$; $p = .02$; $\eta^2 = .26$, which, in turn, were more likely to be recalled correctly than isolated symptoms (16.7%); $t(18) = 3.60$; $p < .01$; $\eta^2 = .42$.

Not including antisocial personality disorder in either experiment, we found levels of correct recall for symptoms were significantly higher in Experiment 1 than in Experiment 4 (overall Ms = 53.4% and 37.1%, respectively); $t(38) = 3.40$, $p < .01$; $\eta^2 = .23$. Thus, participants were better able to correctly recall symptoms of Axis I disorders than symptoms of Axis II disorders, across the different symptom types.

Conceptual importance. Averaged relational rank orderings of the symptoms were obtained as in Experiment 2 (shown under Centrality in Table 3) and were correlated with the averaged conceptual importance ratings for each disorder (shown under Importance in Table 3). Collapsed over disorders, we found a positive correlation ($r_s = .46$, over all participants; $r_s = .46$, clinical psychologists alone; $r_s = .48$, graduate students alone). As in Experiment 2, we also ran this analysis after removing all the noncausal directional data. The correlations remained essentially unchanged ($r_s = .43$, over all participants; $r_s = .46$, clinical psychologists alone; $r_s = .45$, graduate students alone). Thus, these effects are most likely because of a causal status effect rather than a general relational effect per se. Broken down by disorder, three of the four disorders showed positive correlations, avoidant, $r_s (7) = .54$, $p = .2$; borderline, $r_s (9) = .13$, $p = .7$; and obsessive–compulsive, $r_s (8) = .88$, $p < .01$. Schizotypal did not show a positive correlation, $r_s (9) = -.13$, $p = .7$, most likely because there were no significantly homogeneous theories among clinicians as reported in the previous analyses under the *Theory consensus* section. Individually, however, a majority of participants showed the causal status effect for schizotypal personality (12 of 19).

We also examined data from only those participants who performed the conceptual centrality task followed by the theory-drawing task to examine the possibility that priming of theories because of task ordering was responsible for driving the effect. Collapsed over disorders, we found this subgroup still showed a clearly positive correlation overall ($r_s = .21$). This rank correlation coefficient is marginally significantly smaller than the correlation coefficient for data from only those participants who completed the theory drawing task first ($r_s = .60$); $t(3) = 2.83$; $p < .07$; $\eta^2 = .73$; at the $\alpha = .05$ level, observed power = .49. Thus, in Experiment 4 (but not in Experiment 2), having people remember their theories before completing conceptual importance judgments marginally enhanced the causal status effect.

Familiarity analyses. The goal of the analyses in this section was to determine whether the degree of causal status effect exhibited by the participants changed as a function of how familiar they were with each particular personality disorder. As in Experiments 1 and 2, a single data set for high familiarity and another for low familiarity were produced based on composite familiarity scores. To summarize the results in this section, we found familiarity with disorders to be a significant moderator for the effect of background theories in the diagnosis-related tasks (the hypothetical patients task and conceptual importance task), although not in the recall task. Specifically, the causal status effect occurred only for high-familiarity disorders in the hypothetical patients task, and clinicians showed the causal status effect only for high-familiarity

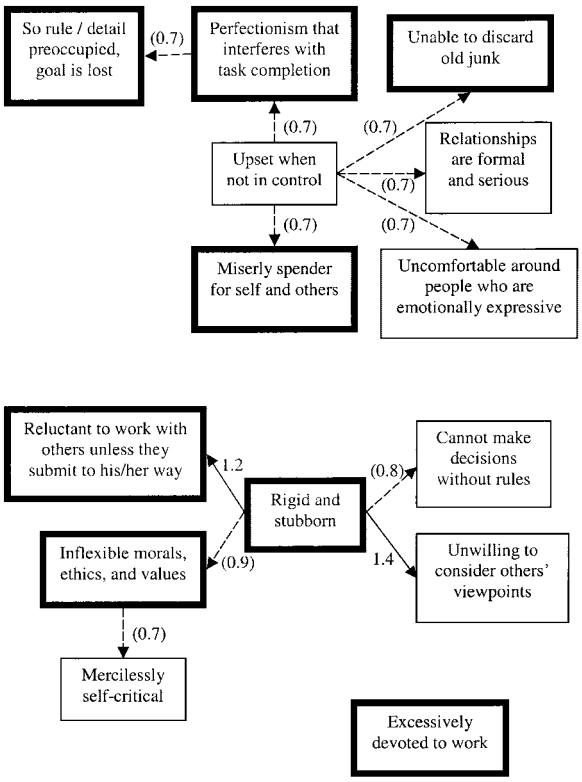

Figure 15. A composite of all participants' drawings of obsessive–compulsive personality disorder in Experiment 4. Because there was more room for detail here, links from 0.7 to < 1.0 are included in the diagram in parentheses (all links, as for the disorders, were included in the analyses). Links with an average strength less than 1.0 are indicated by dotted arrows. Diagnostic criteria, shown in boldfaced boxes, are reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition. Copyright 1994 American Psychiatric Association.

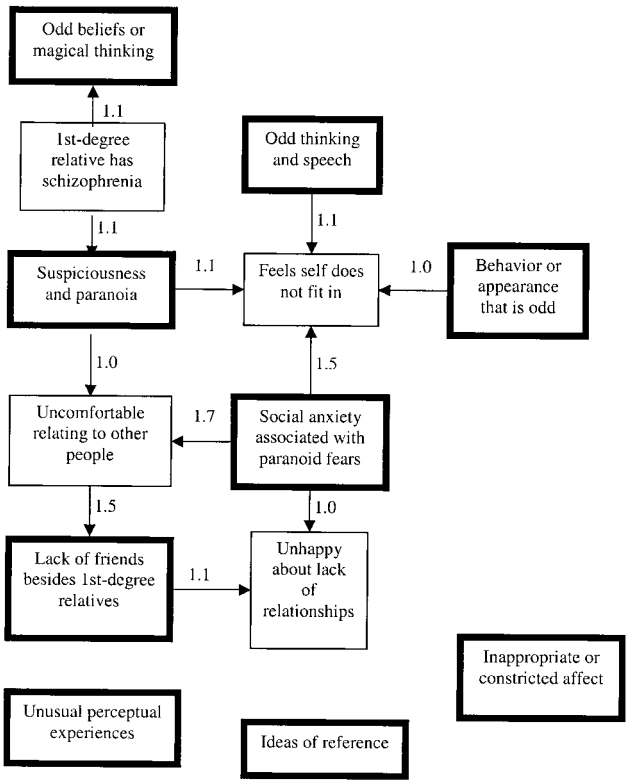

Figure 16. A composite of all participants' drawings of schizotypal personality disorder in Experiment 4. Diagnostic criteria, shown in boldfaced boxes, are reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition. Copyright 1994 American Psychiatric Association.

disorders in the conceptual importance task. The detailed results of each are described in the following paragraphs.

In the hypothetical patients familiarity analysis, a 2 (expertise; clinicians vs. graduate students) × 2 (familiarity; high vs. low) × 3 (item type; causally central vs. causally peripheral vs. isolated) ANOVA was conducted on the typicality ratings. The critical interaction of Familiarity × Item type was significant, $F(2, 34) = 3.95$, $MSE = 520.19$; $p < .03$; $\eta^2 = .19$. Most notably, causally central patients were rated as more typical than causally peripheral patients for the high-familiarity disorders (Ms = 62.2 and 49.3, respectively); $t(18) = 3.16$; $p < .01$; $\eta^2 = .36$ but not for the low-familiarity disorders (Ms = 53.7 and 54.8, respectively); $t(18) = -0.33$; $p = .7$; $\eta^2 = .006$. Isolated patients received comparable ratings regardless of how familiar the disorder was (Ms = 42.1 and 44.8 for high- and low-familiarity disorders, respectively); $t(18) = -0.74$; $p = .5$; $\eta^2 = .03$. No other effects or interactions concerning familiarity were significant (all ps > .2; all $\eta^2 < .07$).

In the conceptual importance familiarity analysis, a 2 (expertise; clinicians vs. interns) × 2 (familiarity; high vs. low) ANOVA was carried out on the correlation coefficients mapping the correlation between conceptual importance judgments and causal centrality. There was a significant interaction of Familiarity × Expertise, $F(1, 16) = 7.0$, $MSE = 0.47$; $p < .02$; $\eta^2 = .30$, such that experts showed the causal status effect only for high-familiarity disorders (mean $r = .39$, as opposed to .01 for low-familiarity disorders); $t(8) = 3.06$; $p < .02$; $\eta^2 = .54$, whereas novices showed the effect to an equal degree for high- and low- familiarity disorders (mean

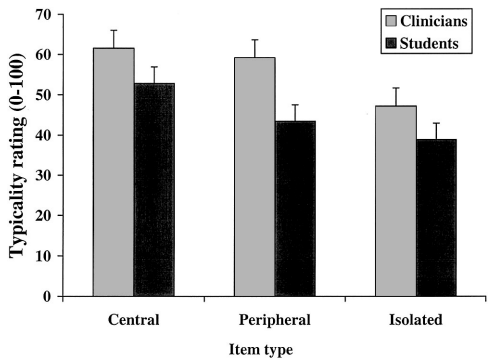

Figure 17. Clinical psychologists' and clinical psychology graduate students' typicality ratings for hypothetical patients with potential personality disorders in Experiment 4. Error bars indicate standard errors.

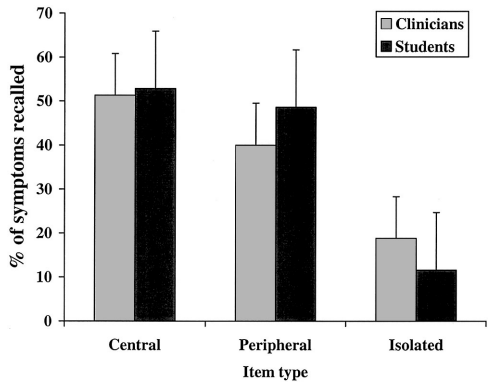

Figure 18. Percentages of symptoms correctly recalled from hypothetical patients with potential personality disorders seen prior to a time delay in Experiment 4. Error bars indicate standard errors.

*r*s = .34 and .42, respectively); $t(8) = -0.70$; $p = .5$; $\eta^2 = .06$. There were no other significant effects (all *p*s > .1; all $\eta^2 = .1$).

Finally, in the recall task familiarity analysis, a 2 (expertise; clinicians vs. graduate students) × 2 (familiarity; high vs. low) × 3 (item type; causally central, causally peripheral, and isolated) ANOVA was conducted on the rates of correct recall. There were no significant effects or interactions at all concerning familiarity (all *p*s > .2; all $\eta^2 < .1$).

In sum, Experiment 4 increased the range of familiarity of disorders by using disorders that were considered to be unfamiliar to the pretest participants (see Appendix B). When this range was increased, a stronger effect of familiarity was found than in the previous experiments, such that the causal status effect tended to be stronger with familiar disorders than with unfamiliar disorders. This could be due to a confidence effect, in which clinicians rely more on their theories when they are more familiar with the disorders being considered.

Theory representation. Nine of the 10 experts and 8 of 9 novices opted to draw causal theories for all personality disorders. The 10th expert and the 9th novice drew causal theories for three of the four disorders. Participants' theories were quite complex with an average of 39.3 links per disorder (range: 0–177; see Kim, 2002, for more detailed analyses).

The number of total links for these disorders was generally lower than the number of total links in Experiment 2, presumably because there were also fewer symptoms in the Experiment 4 disorders ($M = 18$ symptoms per disorder vs. 25 symptoms per disorder in Experiment 2). In addition, participants reported having seen significantly fewer patients with the Experiment 4 disorders than with the Experiment 2 disorders (*M*s = 2.8 vs. 13.1 patients in the past year, respectively); $t(37) = 3.07$; $p < .01$; $\eta^2 = .20$. Participants were also marginally significantly less confident of the theories they drew in Experiment 4 than in Experiment 2 (*M*s = 5.3 and 6.4, respectively, on a 9-point scale); $t(37) = 2.02$; $p = .05$; $\eta^2 = .10$. Participants' familiarity ratings for the disorders went in the same direction, although the difference was not significant ($p > .1$; $\eta^2 = .06$).

Proportion of causal links. Just as in Experiment 2, we conducted an analysis to determine what percentage of the relational links drawn were actually causal links. Again, causal and noncausal links were defined according to Ahn et al. (2002). The results showed that 87.4% of links drawn were causal links. Although lower than the percentage of causal links in Experiment 2, the proportion of causal links is nonetheless large enough to suggest that the effect found in Experiment 4 was driven by a causal status effect.

Summary. Participants showed a strong causal status effect and the predicted isolated versus causal symptoms effect in both the hypothetical patients task and the recall task. This was even the case for schizotypal personality disorder, a disorder in which clinicians' theories were found to differ from each other radically. This suggests that clinicians may rely on their theories in clinical reasoning even when those theories may be extremely idiosyncratic. A causal status effect in the conceptual importance measure was also found in graduate students and clinical psychologists. After increasing the range of familiarity in the stimuli, familiarity in general was found to be a moderator for the causal status effect in diagnosis-related tasks only (i.e., hypothetical patients and conceptual importance, but not recall). In the hypothetical patients task, the causal status effect was found only in high-familiarity disorders for both expertise groups. In the conceptual importance task, there was an effect of expertise such that the causal status effect was found in high- but not low-familiarity disorders for clinicians, whereas graduate students showed the effect for both types of disorders. Most of the links drawn by participants were causal or implied causal links, again suggesting that the effect was driven specifically by a causal status effect. Finally, all participants had causal theories about most or all of the disorders, and these theories were quite complex.

Experiment 5: Between-Group Consensus on Symptom Centrality in Personality Disorders

In Experiments 2 and 4, most participants' theories agreed with each other to a significant degree despite the fact that these participants were diverse as to theoretical orientation. The results of Experiments 2 and 3 demonstrated that the clinicians, interns, and lay people in these studies all weighted symptoms in the same way. However, the disorders used in Experiment 4 were all classified as less familiar to lay people than the Experiment 2 disorders (see Appendix A). Are clinicians' theories also in agreement with commonsense knowledge for these less familiar disorders? Experiment 5 sought to determine whether this is the case.

Method

Participants. As in Experiment 3, lay people were defined as any individuals without formal training in clinical psychology. Twenty-three undergraduate students at Vanderbilt University, who met this criterion, completed the task along with other unrelated tasks in exchange for partial fulfillment of requirements for an introductory psychology course.

Materials and procedure. Some of the disorders used in Experiment 4 were purposely selected as those unfamiliar even to experts. Thus, we expected that lay people might not have even heard of those disorders (e.g., schizotypal personality). However, it was possible that these lay participants could have concepts of these unfamiliar personality disorders without necessarily knowing the labels. If this was true, then examining their concepts of these disorders simply by presenting them with clinical labels might lead to an underestimation of their performance. Thus, following the

Table 3
Participants' Mean Conceptual Importance Ratings and Causal Centrality Rank Orders Calculated by Equation (1) for the DSM–IV *Diagnostic Criteria in Experiment 4*

Disorder and symptom	Centrality	Importance
Avoidant personality disorder		
Avoids occupational activities that involve significant interpersonal contact, because of fears of criticism, disapproval, or rejection	5	79.5
Is unwilling to get involved with people unless certain of being liked	4	68.7
Shows restraint within intimate relationships because of the fear of being shamed or ridiculed	6	70.3
Is preoccupied with being criticized or rejected in social situations	2	77.9
Is inhibited in new interpersonal situations because of feelings of inadequacy	3	78.7
Views self as socially inept, personally unappealing, or inferior to others	1	81.1
Is unusually reluctant to take personal risks or to engage in any new activities because they may prove embarrassing	7	74.7
Borderline personality disorder		
Frantic efforts to avoid real or imagined abandonment	4	80.8
A pattern of unstable and intense interpersonal relationships in which others are alternately idealized and devalued	5	84.3
Identity disturbance: markedly and persistently unstable sense of self	1	86.5
Impulsivity in two or more potentially self-damaging areas	8	84.7
Recurrent suicidal behavior, gestures, or threats, or self-mutilating behavior	9	84.2
Affective instability due to a marked reactivity of mood	2	80.8
Chronic feelings of emptiness	6	69.5
Inappropriate, intense anger or difficulty controlling anger	3	80.0
Transient stress-related paranoid ideation or severe dissociative symptoms	7	57.1
Obsessive–compulsive personality disorder		
Unable to discard old junk, even that which has no sentimental value	8	58.7
Shows perfectionism that interferes with task completion	2	86.4
Preoccupied to such a degree with details, rules, etc. that the goal is lost	4	84.1
Shows rigidity and stubbornness	1	86.0
Is reluctant to work with others unless they submit to exactly his or her way of doing things	6	63.3
Is scrupulous and inflexible about matters of morality, ethics, or values	3	69.5
Adopts a miserly spending style toward both self and others	7	47.9
Is excessively devoted to work and productivity	5	75.5
Schizotypal personality disorder		
Ideas of reference (excluding delusions of reference)	8	72.4
Odd beliefs or magical thinking that influences behavior and is inconsistent with subcultural norms	3	80.8
Unusual perceptual experiences, including bodily illusions	7	72.1
Odd thinking and speech	4	84.5
Suspiciousness or paranoid ideation	2	67.6
Inappropriate or constricted affect	5	85.5
Behavior or appearance that is odd, eccentric, or peculiar	6	80.0
Lack of close friends or confidants other than first-degree relatives	9	68.5
Excessive social anxiety that does not diminish with familiarity and tends to be associated with paranoid fears rather than negative judgments about self	1	67.9

Note. Higher numbers correspond to greater conceptual importance. Rank orders shown for causal centrality were assigned within the diagnostic criteria for each disorder, although the centralities themselves were calculated using $c_{i,t+1} = \Sigma_j\, d_{ij} c_{j,t}$ (Equation [1]) based on directional relations with all symptoms (diagnostic criteria plus characteristic symptoms). Symptom descriptions are reprinted with permission from the *Diagnostic and Statistical Manual of Mental Disorders,* Fourth Edition (*DSM–IV*). Copyright 1994 American Psychiatric Association.

procedure used by Kim and Ahn (2002), all participants were first presented with lists of the *DSM–IV* diagnostic criteria and characteristic symptoms for the four disorders from Experiment 4. Participants were asked to simply read through the lists so that they would have an idea of what the disorder names meant. The order of symptoms within each disorder list was scrambled to create two different versions that were counterbalanced between participants. The order of the disorders was randomized for each participant.

The remainder of the materials exactly mirrored those used for lay participants in Experiment 3. Eight hypothetical patients were developed on the basis of averaged causal theories of the clinical participants in Experiment 4. One causally central patient and one causally peripheral patient were created for each of the four personality disorders, in the same manner described in Experiment 3. No isolated patients could be created from the averaged theories, because no symptoms were left isolated by all of the Experiment 4 participants.

As in Experiment 3, participants were then presented with these patients and asked to rate how typical they were of the disorder. Patients were presented to participants in one of two counterbalanced orders.

Results and Discussion

A 2 (item type; causally central, causally peripheral) ×4 (disorder, the four disorders) ANOVA demonstrated that undergraduates judged causally central patients ($M = 75.9$) to be much more typical of each disorder than causally peripheral patients ($M = 64.8$); $F(1, 22) = 31.65$, $MSE = 178.29$; $p < .01$; $\eta^2 = .59$. The results went in the expected direction for all four disorders, but there was a significant interaction, $F(3, 66) = 3.92$, $MSE = 214.80$; $p = .01$; $\eta^2 = .15$. Multiple comparisons showed that the causal status effect occurred to a significant extent in avoidant, $t(22) = 3.01$; $p < .01$; $\eta^2 = .29$; obsessive–compulsive, $t(22) = 2.51$; $p = .02$; $\eta^2 = .22$; and schizotypal, $t(22) = 4.33$; $p < .01$; $\eta^2 = .46$, personality disorders, but not in borderline personality disorder, $t(22) = 0.53$; $p = .6$; $\eta^2 = .01$. Borderline personality is thereby the only demonstration of a disorder in Experiments 3 and 5 for which lay and clinician theories did not agree. Interestingly, this suggests that not all clinical theories are necessarily in concordance with commonsense background knowledge. For the majority of the disorders tested, however, despite the fact that undergraduates are less familiar with these personality disorders than the ones used in Experiment 2, they still made diagnostic decisions in accord with those made by clinicians.

General Discussion

Summary of Results

In this study, a number of similar tasks and analyses were used over the three major experiments (1, 2, and 4). To facilitate discussion of the data, we summarize the major results by task across experiments. (See Figure 1 for a summary diagram.)

For the hypothetical patients task (Task III in Figure 1), patients with causally central symptoms were more likely to be diagnosed with the target disorder (Experiment 1), were judged to be more typical of the target disorder (Experiment 2), and were more important in participants' concepts of the target disorder (Experiment 4) than patients with causally peripheral symptoms. In addition, patients with causally central and patients with causally peripheral symptoms received significantly higher ratings on all three measures than did patients with isolated symptoms.

The two types of memory tasks (Task IV in Figure 1) showed a similar pattern of results. In the recall task, Experiments 1 and 4 showed that causally central symptoms were more frequently recalled correctly than causally peripheral symptoms and that these, in turn, were more frequently recalled than isolated symptoms. In the recognition task (Experiment 2), participants were more likely to falsely recognize symptoms in hypothetical patients that were central to their theory of the disorder than symptoms that were peripheral to or isolated from their theory.

Additional measures supplied more converging evidence for the use of theories in reasoning about mental disorders. In analyses correlating individual participants' rated symptom centralities (Task II in Figure 1) with the centralities of those symptoms in their theory drawings, the correlation coefficients were greater than 0 in all three major experiments. Overall, 52 of the 59 participants in these experiments showed some degree of the causal status effect in this measure (30 of 35 experts; 23 of 25 novices).

In addition, Experiments 3 and 5 showed that the causal theories of clinical psychologist experts and trainees concurred with the lay people's symptom weightings for all Axis I and Axis II disorders tested except for borderline personality disorder. Notably, the clinician participants in Experiment 4 were shown to agree significantly among themselves on their theory of borderline personality.

Implications for Theories of Conceptual Thought

The current study extends previous demonstrations of theory-based reasoning (e.g., Wisniewski & Medin, 1994) by showing theory-based reasoning in the unique field of mental disorders, in which clinical psychologists are provided with a model of diagnosis that sidesteps the use of theory (Spitzer et al., 1994). The current research demonstrated strong converging evidence that despite the fact that they have a well-known atheoretical manual, clinicians nonetheless use their own theories when reasoning about mental disorders. It is unlikely that these effects are solely because of short-term priming of their theories during the task. We found positive correlations between causal centralities and conceptual centralities even among those who drew their theories after they rated conceptual centralities. In addition, we found the causal status effect in the hypothetical patients task, which was carried out about 2 weeks after the participants drew their theories.

This work also details, for the first time, the use of a specific mechanism of theory-based reasoning by experts in a domain. Previous work showing the effects of top–down expectations in experts (e.g., Chapman & Chapman, 1967, 1969; Chi, Feltovich, & Glaser, 1981; Gaultney, Bjorklund, & Schneider, 1992; Gobet & Simon, 1998; Simon & Chase, 1973) did not investigate those effects to the same degree of specificity as has been carried out here. Thus, in this sense also, the current research breaks new ground. (See also Rehder & Hastie, 2001, and Strevens, 2000, for other examples of specifying theory-based categorization, and Ahn & Kim, 2000, and Ahn et al., 2001, for more discussion of these mechanisms.)

In addition, the current study (Experiments 2 and 4) found that the vast majority of relations in clinical psychologists' representations of mental disorders could be classified as either causal or as implying causality (Ahn et al., 2002). These results confirm previous researchers' claims that the core component of theory representations is causality (Carey, 1985; Hickling & Wellman, 2001).

However, we do not intend to argue that clinicians do not use similarity-based prototype representations (e.g., Cantor, et al., 1980) or exemplar- or case-based reasoning. Indeed, in accord with arguments that both theory-based and similarity-based reasoning play a role in conceptual thought (Keil et al., 1998; Wisniewski & Medin, 1994), the type of theory-based reasoning shown here could easily coexist with, or conceivably even arise from, a similarity-based representation. It is also possible that one could develop a prototype-based or exemplar-based model of clinical reasoning where features are interconnected and the features' causal status determines their weights.

Furthermore, we do not intend to argue that feature weights are determined only by domain theories (see Ahn & Kim, 2000; Rehder & Hastie, 2001, for further discussion on this issue). Other

known determinants for feature weights include category validity (the probability that an object has a certain feature given that it belongs to a certain category; e.g., the probability that a patient has feelings of worthlessness given that the patient has a major depressive episode) and cue validity, or diagnosticity (the probability that an object belongs to a certain category given that it has a certain feature; e.g., the probability that a patient has a major depressive episode given that the patient has feelings of worthlessness).

The current study did not strictly control for cue and category validities because we used real-life disorders, rather than artificially controlled disorders, to increase the study's external validity and to use the *DSM–IV* as the alternative model of clinical reasoning. We acknowledge that our data did not conclusively rule out the possibility of a purely statistical interpretation. For the following reasons, however, we believe it to be reasonably unlikely that the causal status effect we observed in the current study was confounded with the effect of probabilities associated with features.

First, in a separate investigation, we controlled for probabilities of symptoms in a study investigating lay people's use of the causal status effect in artificial everyday categories and found that the effect exists robustly above and beyond the effects of probability alone (Ahn et al., 2000). Similarly, previous documentation has shown that people's theories predict their responses on the measure of conceptual centrality used in Experiment 1, but their theories do not predict their estimates of cue and category validity (Sloman et al., 1998). In another study with lay people, we manipulated which real-life symptoms (embedded in artificial mental disorders) were thought to be causally central and still got a robust causal status effect (Kim & Ahn, 2002). This manipulation effectively controlled for any prior notions participants might have had about how frequently one would expect each symptom to occur in the world.

Furthermore, the fact that clinicians showed a strong causal status effect for schizotypal personality in Experiment 4, despite having extremely different theories for that disorder, demonstrates that they did not rely on the objective probabilities of the symptoms. For instance, suppose that Symptom A has a higher probability of occurrence than Symptom B. Some of the participants weighted A more than B when making judgments, whereas others weighted B more than A, depending on their own idiosyncratic theories. If feature weighting was based on cue or category validities, the same symptoms should have been given the same weights in diagnosis across all participants. Moreover, none of the disorders could be said to have shown strong agreement per se, although the effects were statistically significant.

Finally, the way in which the *DSM–IV* is structured could potentially lead a clinician to assume that the category validities of symptoms within a disorder are equal to each other. That is, schizophrenia's "any 2 out of 5 symptoms" specification carries the implication that a patient with schizophrenia should be equally likely to have any of the diagnostic criteria symptoms listed in the manual. Thus, in this sense the *DSM–IV* itself serves as a control for category validity. Thus, we can be reasonably confident that the effects shown here cannot be explained away by the probabilities of symptoms.

Lack of Expert–Novice Differences

Expertise was never found to be a strong moderator of the effect in any measures used in the current study. Indeed, even undergraduates rely on their own lay theories when reasoning about mental disorders (Kim & Ahn, 2002). Such results are consistent with the well-documented finding that professionals and nonprofessionals in therapy differ little, if at all, as to their patients' treatment outcomes (Dawes, 1994; Durlak, 1979; Faust & Zlotnick, 1995; Meehl, 1954; but see Atkins & Christensen, 2001; Burlingame, Fuhriman, Paul, & Ogles, 1989; and Karon & VandenBos, 1970, for discussions of possible circumstances under which expertise may matter for therapy outcomes).

Why were no expertise effects found in this study? The possibility of low power can be discounted because a collapsed analysis of the hypothetical patient data across Experiments 1, 2, and 4 also came up with nonsignificant expertise effects. One speculative, and certainly more interesting, possibility is that because of the atheoretical emphasis in the field of clinical psychology, clinicians are left to develop their own specific theories and have no way of verifying whether their theories are correct or not even as their expertise grows (Dawes, 1994). It remains to be investigated whether this is truly the case.

The other possibility is that the critical factor is not years of practice per se (which was our criterion for determining expertise) but rather familiarity with specific disorders. Indeed, in Experiment 4, where the range of familiarity with specific disorders was expanded, familiarity was found to be a significant moderator of the effect in the hypothetical patients and conceptual importance tasks (though not for the memory task). Specifically, in these two tasks, the causal status effect was stronger for familiar than for unfamiliar disorders, suggesting that clinicians may rely more on their theories when they feel familiar with the disorder.

What, then, have experts learned of value to diagnosis over their training and years of practice? Elstein (1988) has suggested that what may be gained with expertise in clinical reasoning is the speed with which a clinician can come to a diagnosis. There may also be other expert–novice differences in aspects of theory-based clinical reasoning that were not measured in the current experiments. For instance, experts may be better than novices at recognizing a symptom for what it is. In addition, accuracy is a critical component of diagnosis that was not addressed at all in the current study.

Furthermore, experts may be better at differential diagnosis, a process that can be divided very roughly into two parts. The *DSM–IV* casebook (Spitzer et al., 1994) provides examples of how to diagnose numerous case studies. In all of these, the process involves, first, determining which diagnostic criteria are present and whether there are enough to justify a diagnosis and, second, ruling out medical conditions and other mental disorder possibilities. The current study suggests that clinicians might incorporate theory-based reasoning into the first part of the process. There may be expert–novice differences, however, in the second part of the process, which may require more experience to master. In addition, there are other cues to feature weighting that were not measured in the current study. Base-rate cues including category validity and cue validity are two of the most important. It is possible that expert–novice differences would be present for such feature weighting cues, in that novices have not yet acquired their own

sense of how frequently symptoms occur in the world. Such possibilities await further investigation.

The Relationship Between This Research and the DSM System

It has not been the intent here to imply that these data show how the *DSM* system should be modified. Rather, this study seeks to provide a descriptive, not normative, model of clinical reasoning. Ideally, to revise the *DSM* itself on the basis of a theory-like model, it would clearly be necessary to discover a single correct etiology for each disorder and then use that framework for the manual, as is often done for medical diseases. However, current mental disorder etiologies have not yet reached that point. It is possible that these data may be useful by providing researchers who study the etiology of disorders with potential theories that most clinicians seem to agree on. In other words, these data may serve as starting hypotheses for these researchers.

A number of theorists have discussed reasons to base taxonomies on theory. One of the *DSM*'s goals in not specifying underlying theories was to avoid battles between different theoretical schools as to which theories should be included in or focused on in the manual. However, some have argued that the advantage gained by such a solution is outweighed by disadvantages concerning the negative effect a taxonomy that is silent with respect to theory has on clinical research (Follette & Houts, 1996). Although advancing research is ostensibly not the primary purpose of the *DSM* system, these authors argue that diagnosis and research are inextricably linked in that being able to conduct research on a particular disorder depends on being able to identify people who have it. Follette (1996), for instance, criticized the diagnostic criteria, pointing out that they never consider behavior within a context (i.e., both situational and biological). Carson (1996) also criticized the diagnostic criteria, pointing out that in being simplified, they have become much more crude than the actual manifestation of the disorders. His argument is that such a trend runs counter to the fact that advances in science historically involve taking measures of the studied phenomenon with increasing levels of precision. Thus, there may be a number of conceptual reasons why an atheoretical taxonomy of disorders is far from ideal. It will be the task of other studies, however, to determine whether this is the case.

Implications for Theories of Clinical Reasoning

The question of how clinicians think and reason has been primarily concerned with fallacies in clinical reasoning and ways in which such errors can be circumvented (i.e., Dawes, 1994; Dawes, Faust, & Meehl, 1989; Einhorn, 1988; Elstein, 1988; Garb, 1996; Meehl, 1954). Much of this work to date has investigated clinical reasoning with respect to principles of decision making, including cognitive heuristics and biases (i.e., Eva & Brooks, 2000; Eva, Brooks, & Norman, 2001; Garb, 1996; Rabinowitz, 1993; Turk & Salovey, 1985). In contrast, the current article examined clinical reasoning as an example of categorization.

One argument against the data presented here could be that they are not meaningful because in most real-life cases clinicians must make formal *DSM* diagnoses using checklists and are therefore unlikely to be influenced by their causal theories to the degree documented here. However, we believe that the effect of theory-based conceptual representations found in the current studies may still pervade critical aspects of clinical work. As shown in the current study, clinicians are better at recalling symptoms central to their theories, and they may be biased to falsely remember theory-central symptoms of patients they have already seen. These tendencies may influence clinicians' informal initial diagnoses, which may, in turn, markedly affect how clinicians subsequently perceive and interact with their patients. For instance, symptoms of mental disorders are often ambiguous, and clinicians may focus their attention on detecting symptoms central to their theories.

It should be noted, however, that theory-based reasoning in itself is not a reasoning fallacy, provided that clinicians' theories are valid (Dawes et al., 1989). Experts do seem to have idiosyncratic theories in some cases, as for schizotypal personality disorder. When theories are idiosyncratic and also happen to be invalid, relying on them may constitute a fallacy in clinical judgment. In general, however, categorization based on valid theories conforms to the higher levels of taxonomy that scientists should strive for (Hempel, 1965). As philosophers of science have argued over the decades, the goal of scientific research is to eventually develop a theory that explains a set of observations, not just to collect more and more observations. It seems to follow, then, that clinicians are also justified in developing theories that make sense of the knowledge they have amassed about mental disorders. Indeed, symptoms that explain and cause other symptoms may be the most important ones to attend to and remember, because they may be the more useful predictors for prognosis and treatment. In the current study, most clinicians' theories were found to be in general agreement with each other's and with lay people's theories, at least in disorders that are also familiar to lay people. This suggests that experts' theories of these socioculturally familiar disorders are not highly idiosyncratic but rather seem to concur with commonsense notions and may, therefore, be worthy of careful consideration in seeking to understand the process of clinical reasoning.

References

Ahn, W. (1998). Why are different features central for natural kinds and artifacts? The role of causal status in determining feature centrality. *Cognition, 69,* 135–178.

Ahn, W., Kalish, C., Gelman, S. A., Medin, D. L., Luhmann, C., Atran, S., et al. (2001). Why essences are essential in the psychology of concepts. *Cognition, 82,* 59–69.

Ahn, W., & Kim, N. S. (2000). The causal status effect in categorization: An overview. In D. L. Medin (Ed.), *The psychology of learning and motivation* (pp. 23–65). San Diego, CA: Academic Press.

Ahn, W., Kim, N. S., Lassaline, M. E., & Dennis, M. J. (2000). Causal status as a determinant of feature centrality. *Cognitive Psychology, 41,* 361–416.

Ahn, W., Marsh, J. K., Luhmann, C. C., & Lee, K. (2002). Effect of theory-based feature correlations on typicality judgments. *Memory & Cognition, 30,* 107–118

American Psychiatric Association. (1958). *Diagnostic and statistical manual of mental disorders* (1st ed.). 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. (1988). *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.

Atkins, D. C., & Christensen, A. (2001). Is professional training worth the bother? A review of the impact of psychotherapy training on client outcome. *Australian Psychologist, 36,* 122–131.

Barlow, D. H., & Durand, V. M. (1999). *Abnormal psychology* (2nd ed.). Pacific Grove, CA: Brooks/Cole.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology, 11,* 177–220.

Burlingame, G. M., Fuhriman, A., Paul, S., & Ogles, B. M. (1989). Implementing a time-limited therapy program: Differential effects of training and experience. *Psychotherapy, 26,* 303–313.

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

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: Plenum.

Carson, R. C. (1996). Aristotle, Galileo, and the *DSM* taxonomy: The case of schizophrenia. *Journal of Consulting and Clinical Psychology, 64,* 1133–1139.

Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous psycho-diagnostic observations. *Journal of Abnormal Psychology, 72,* 193–204.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74,* 271–280.

Chi, M., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121–152.

Clarkin, J. F., Widiger, T. A., Frances, A., Hurt, S. W., & Gilmore, M. (1983). Prototypic typology and the borderline personality disorder. *Journal of Abnormal Psychology, 92,* 263–275.

Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York: Free Press.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243,* 1668–1674.

Durlak, J. A. (1979). Comparative effectiveness of paraprofessional and professional helpers. *Psychological Bulletin, 86,* 80–92.

Einhorn, H. J. (1988). Diagnosis and causality in clinical and statistical prediction. In D. C. Turk & P. Salovey (Eds.), *Reasoning, inference, and judgment in clinical psychology* (pp. 51–70). New York: Free Press.

Elstein, A. S. (1988). Cognitive processes in clinical inference and decision making. In D. C. Turk & P. Salovey (Eds.), *Reasoning, inference, and judgment in clinical psychology* (pp. 17–50). New York: Free Press.

Eva, K. W., & Brooks, L. R. (2000). The under-weighting of implicitly generated diagnoses. *Academic Medicine, 75,* S81-S83.

Eva, K. W., Brooks, L. R., & Norman, G. R. (2001). Does "shortness of breath" = "dyspnea"?: The biasing effect of feature instantiation in medical diagnosis. *Academic Medicine, 76,* S11-S13.

Faust, D., & Zlotnick, C. (1995). Another Dodo bird verdict—Revisiting the comparative effectiveness of professional and paraprofessional therapists. *Clinical Psychology & Psychotherapy, 3,* 157–167.

Follette, W. C. (1996). Introduction to the special section on the development of theoretically coherent alternatives to the *DSM* system. *Journal of Consulting and Clinical Psychology, 64,* 1117–1119.

Follette, W. C., & Houts, A. C. (1996). Models of scientific progress and the role of theory in taxonomy development: A case study of the *DSM*. *Journal of Consulting and Clinical Psychology, 64,* 1120–1132.

Furnham, A. (1995). Lay beliefs about phobia. *Journal of Clinical Psychology, 51,* 518–525.

Garb, H. N. (1996). The representativeness and past-behavior heuristics in clinical judgment. *Professional Psychology: Research and Practice, 27,* 272–277.

Gaultney, J. F., Bjorklund, D. F., & Schneider, W. (1992). The role of children's expertise in a strategic memory task. *Contemporary Educational Psychology, 17,* 244–257.

Gelman, S. A. (2000). The role of essentialism in children's concepts. In H. W. Reese (Ed.), *Advances in child development and behavior* (pp. 55–98). San Diego, CA: Academic Press.

Gelman, S. A., & Kalish, C. W. (1993). Categories and causality. In R. Pasnak & M. L. Howe (Eds.), *Emerging themes in cognitive development* (Vol. 2, 3–32). New York: Springer-Verlag.

Genero, N., & Cantor, N. (1987). Exemplar prototypes and clinical diagnosis: Toward a cognitive economy. *Journal of Social and Clinical Psychology, 5,* 59–78.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155–170.

Gentner, D. (1989). The mechanisms of analogical reasoning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199–241). Cambridge, England: Cambridge University Press.

Gobet, F., & Simon, H. A. (1998). Expert chess memory: Revisiting the chunking hypothesis. *Memory, 6,* 225–255.

Hempel, C. G. (1965). *Aspects of scientific explanation*. New York: Free Press.

Hickling, A. K., & Wellman, H. M. (2001). The emergence of children's causal explanations and theories: Evidence from everyday conversation. *Developmental Psychology,* 37, 668–683.

Hill, D. (1983). *The politics of schizophrenia: Psychiatric oppression in the United States*. Lanham, MD: University Press of America.

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

Horowitz, L. M., Wright, J. C., Lowenstein, E., & Parad, H. W. (1981). The prototype as a construct in abnormal psychology: I. A method for deriving prototypes. *Journal of Abnormal Psychology, 90,* 568–574.

Hyler, S. E., Williams, J. B. W., & Spitzer, R. L. (1982). Reliability in the *DSM–III* field trials: Interview vs. case summary. *Archives of General Psychiatry, 39,* 1275–1278.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28.

Karon, B. P., & VandenBos, G. R. (1970). Experience, medication, and the effectiveness of psychotherapy with schizophrenics. *British Journal of Psychiatry, 116,* 427–428.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Keil, F. C., Smith, W. C., Simons, D. J., & Levin, D. T. (1998). Two dogmas of conceptual empiricism: Implications for hybrid models of the structure of knowledge. *Cognition, 65,* 103–135.

Kim, N. S. (2002). *Clinical psychologists' theory-based representations of mental disorders affect their diagnostic reasoning and memory*. Unpublished doctoral dissertation, Yale University, New Haven, CT.

Kim, N. S., & Ahn, W. (2002). The influence of naïve causal theories on lay diagnoses of mental illnesses. *American Journal of Psychology, 115,* 33–65.

Kraepelin, E. (1913). *Psychiatry: A textbook*. Leipzig, Germany: Barth.

Lassaline, M. E. (1996). Structural alignment in induction and similarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 754–770.

Lehman, A. K. (1992). The content and organization of therapists' mental representations of their patients and the psychotherapy process (Doctoral dissertation, Yale University, 1991). *Dissertation Abstracts International, 53,* 1067.

Loftus, E. F., & Ketcham, K. (1994). *The myth of repressed memory*. New York: St. Martin's Press.

Mather, M., Johnson, M. K., & DeLeonardis, D. M. (1999). Stereotype reliance in source monitoring: Age differences and neuropsychological test correlates. *Cognitive Neuropsychology, 16,* 437–458.

Matschinger, H., & Angermeyer, M. C. (1996). Lay beliefs about the

causes of mental disorders: A new methodological approach. *Social Psychiatry and Psychiatric Epidemiology, 31,* 309–315.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 12,* 1469–1481.

Medin, D. L., Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology, 20,* 158–190.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–316.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363.

Rabinowitz, J. (1993). Diagnostic reasoning and reliability: A review of the literature and a model of decision-making. *Journal of Mind & Behavior, 14,* 297–315.

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General, 130,* 323–360.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21,* 803–814.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605.

Ross, B. H. (1997). The use of categories affects classification. *Journal of Memory and Language, 37,* 240–267.

Russell, J. A. (1991). In defense of a prototype approach to emotion concepts. *Journal of Personality and Social Psychology, 60,* 37–47.

Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist, 61,* 394–403.

Sloman, S. A., & Ahn, W. (1999). Feature centrality: Naming versus imagining. *Memory & Cognition, 27,* 526–537.

Sloman, S. A., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence. *Cognitive Science, 22,* 189–228.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Spanos, N. P. (1996). *Multiple identities and false memories: A sociocognitive perspective*. Washington, DC: American Psychological Association.

Spitzer, R. L., Forman, J. B. W., & Nee, J. (1979). *DSM–III* field trials: I. Initial interrater diagnostic reliability. *American Journal of Psychiatry, 136,* 815–817.

Spitzer, R. L., Gibbon, M., Skodol, A. E., Williams, J. B. W., & First, M. B. (Eds.). (1994). DSM–IV *casebook: A learning companion to the Diagnostic and statistical manual of mental disorders, fourth edition*. Washington, DC: American Psychiatric Press.

Strevens, M. (2000). The essentialist aspect of naïve theories. *Cognition, 74,* 149–175.

Turk, D. C., & Salovey, P. (1985). Cognitive structure, cognitive processes, and cognitive–behavior modification: II. Judgments and inferences of the clinician. *Cognitive Therapy and Research, 9,* 19–33.

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

Widiger, T. A. (1982). Prototypic typology and borderline diagnoses. *Clinical Psychology Review, 2,* 115–135.

Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science, 18,* 221–281.

Wittgenstein, L. (1953). *Philosophical investigations*. New York: Macmillan.

(*Appendixes follow*)

Appendix A

Selection of Stimuli for Experiments 1, 2, and 3

In a preexperiment to select the mental disorders used as stimuli in Experiments 1, 2, and 3, 23 undergraduate students from Yale University were recruited in exchange for payment of $7 per hr to complete a survey on how familiar they were with each of about 40 *DSM–IV* Axis I and personality disorders. For each disorder, participants answered seven yes–no questions in the following format: "Have you ever heard of [disorder X]?" "Have you ever thought about [disorder X] in the last three years?" "Have you ever heard about research related to [disorder X]?" "Have you ever heard about a patient with [disorder X]?" "Have you ever heard or thought about what might cause [disorder X]?" "Do you have some idea of what symptoms are associated with [disorder X]?" "Have you ever thought about how symptoms of [disorder X] might be related to each other?"

An index of overall familiarity with each disorder was computed by calculating the percentage of "yes" responses to each disorder across participants. The four Axis I disorders with the highest scores were selected for use in the current study, with the following exceptions. Only one disorder from each class of disorders was included (i.e., anorexia nervosa had the highest score and bulimia the third highest score; only anorexia nervosa was included because both are classified broadly as eating disorders). In addition, dissociative identity disorder was excluded on the grounds that there is still much debate over the disorder's existence (Spanos, 1996). More importantly, even if we assume the disorder's existence, the disorder would nonetheless be very rare. If no clinical trainee or expert is likely to have seen a patient with it, the disorder may be insensitive to any possible expert-novice differences. Using these criteria, we selected anorexia nervosa, major depressive disorder, specific phobia, and schizophrenia for use in the current experiment.

In addition, one Axis II personality disorder was selected. Obsessive–compulsive personality disorder had the highest score but was not included on the grounds that it is very similar to obsessive–compulsive disorder (indeed, some of the clinical graduate student pilot participants in this study voiced the opinion that it is the same thing, except that obsessive–compulsive disorder patients recognize that they have a problem). Thus, antisocial personality disorder, the personality disorder with the second-highest score, was included instead.

Appendix B

Selection of Stimuli for Experiments 4 and 5

In a preexperiment to determine which *DSM–IV* personality disorders were most familiar and which were least familiar to clinicians, familiarity ratings were obtained from a separate group of 3 clinical psychology professors at Vanderbilt University. Each participant rated his or her own familiarity with each of the 10 personality disorders. They were also asked to rate how familiar they thought that the "average" clinician would be with each of the disorders. All ratings were made on a 7-point scale (1 = *very unfamiliar*; 7 = *very familiar*).

For each disorder, we calculated mean ratings separately for the "self" and "average" responses. A third set of ratings was also considered—those that were obtained from undergraduates for Experiment 1 (undergraduates gave familiarity ratings for all the disorders in the *DSM–IV* from both Axis I and II). The two highest overall and two lowest overall rated disorders were selected for use in this study, with one modification. Schizotypal and schizoid personality disorder were, strictly speaking, the two least familiar personality disorders. However, on the grounds that they might be too conceptually similar, we replaced schizoid with the next least familiar disorder, avoidant personality disorder. The two highest familiarity disorders were borderline personality and obsessive–compulsive personality.

Received December 18, 2001
Revision received July 29, 2002
Accepted July 31, 2002 ■