15

The Effect of Causal Theories on Mental Disorder Diagnosis

Woo-kyoung Ahn and Nancy S. Kim

A seminal paper by Murphy and Medin (1985) produced a major shift in the categorization field. In recent years, the lion's share of research has been devoted to documenting the effects of people's existing beliefs, knowledge, and domain theories on categorization and concept learning (e.g., Lin & Murphy, 1997; Rehder & Hastie, 2001). The diagnosis of mental disorders poses a particularly interesting case because the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (*DSM–IV*; American Psychiatric Association, 1994) establishes a set of guidelines that are intended to be neutral with respect to varying theories on mental disorders. This chapter reviews a series of recent studies that we conducted (Ahn, Novick, & Kim, 2003; Kim & Ahn, 2002a) demonstrating that clinical psychologists are cognitively driven to apply their causal theories about mental disorders when reasoning about them, despite decades of training and practice with guidelines provided by various editions of the *DSM*.

Evolution of the *DSM*

In 1952 the American Psychiatric Association published the first edition of the *DSM*. At that time, psychoanalysis was a highly influential philosophy, with at least 10 out of 28 committee members who developed the *DSM–I* (American Psychiatric Association, 1952) being members of psychoanalytic organizations or sympathetic to that approach (Shorter, 1997). In view of this fact, it is not difficult to find influences of psychoanalytic theory in the *DSM–I*. In 1968 the second edition of the *DSM* appeared. Again, psychoanalysis heavily influenced the development of this edition. Freudian theoretical terms such as *neuroses*

The research reported in this article was supported in part by a National Institute of Mental Health grant (RO1 MH57737) to Woo-kyoung Ahn and a National Science Foundation Graduate Research Fellowship and Yale University Dissertation Fellowship to Nancy S. Kim.

and *hysteria* were widespread; *hysteria* referred to symptoms associated with "emotionally charged situations" that are "symbolic of the underlying conflicts" (American Psychiatric Association, 1968, p. 39). However, these theoretical constructs offered few operational criteria needed for reliable diagnoses (e.g., Hempel, 1965). Furthermore, by the early 1970s psychoanalysis was not the only predominant philosophy, making it difficult for the *DSM–II* (American Psychiatric Association, 1968) to be accepted by clinicians of different theoretical orientations. In addition to this complication, most mental disorders still lacked a single universally acknowledged pathogenesis.

In response to these problems, the later editions of the *DSM* (i.e., *DSM–III,* 1980; *DSM–III–R,* 1988; *DSM–IV,* 1994) took a radically different approach—namely, "a descriptive approach that attempted to be neutral with respect to theories of etiology" (American Psychiatric Association, 1994, pp. xvii–xviii). In accordance with this theory-neutral approach, most disorders are defined in terms of a set of surface symptoms or conditions that the patient must meet for diagnosis. For example, schizophrenia is characterized by five symptoms (hallucinations, delusions, disorganized speech, grossly disorganized or catatonic behavior, and negative symptoms). Clinicians are to search for symptoms in their patients that match the *DSM–IV* diagnostic criteria, without incorporating any additional notions regarding the way in which these symptoms may affect one another and, in most disorders, what caused these symptoms in the first place.[1]

Another significant shift in the assumptions underlying the representation of mental disorders is that, starting from the *DSM–III*, a probabilistic model has been explicitly adopted. According to this model, boundaries of categories are fuzzy and categories do not necessarily have defining features. For example, the prototypical patient with schizophrenia has five symptoms, but a presenting patient need have only two of those five symptoms for a diagnosis of the disorder. In some sense, this is a natural consequence of taking the theory-neutral, symptom-based approach because although patients with the same disorder might share a common underlying cause, the manifestations at the symptomatic level might vary depending on individual patients' genetic tendencies and life experiences.

Do Clinicians Represent Mental Disorders as Prototypes?

Over the past 20 years, beginning with the *DSM–III*, the *DSM* system has been positioned at the core of research, diagnosis, and treatment in psychopathology in the United States. As Joiner and Schmidt (2002) put it, "It is perhaps only a slight overstatement to say that one cannot get paid—either by insurance or granting agencies—unless *DSM* diagnoses are assigned" (p. 107).

[1]A notable exception is the step of ruling out general medical conditions as the primary cause of the presenting symptoms.

Given that the *DSM* is so important in clinicians' professional lives, has the *DSM's* theory-neutral stance been internalized by clinicians?

A series of earlier studies by Cantor and her colleagues presented evidence suggesting that clinicians appear to represent mental disorders in a way similar to the recent *DSM* systems—that is, in terms of prototypes. In Cantor, Smith, French, and Mezzich (1980), for instance, atypical patients sharing only a small number of symptoms with disorder prototypes were diagnosed less accurately and confidently than typical patients sharing a large number of symptoms with disorder prototypes. This phenomenon, known as the *typicality effect*, has been ubiquitously observed in natural categorization. For instance, people judge a robin as a bird more quickly than they judge a chicken as a bird and people generate typical members before atypical ones in generating exemplars of a category (for a review, see Smith & Medin, 1981). Given such evidence, the prototype approach to categorization posits that people store prototypes of categories that are usually represented in terms of surface features (e.g., "have wings," "fly," "build nests" for birds) and that the categorization of a new object is determined by its similarity to the stored prototypical representation. According to the prototype view, typicality effects arise because of variance in this similarity.

However, the fact that typicality effects occur does not necessarily mean that people represent categories only in terms of surface features or that categorization is based solely on surface similarity. For instance, our concept of birds is not just a list of bird features; most people know that birds can fly because they have wings, and birds can build nests in trees because they can fly. More generally speaking, even lay people's concepts, however naïve, are akin to scientific theories (e.g., Ahn, Marsh, Luhmann, & Lee, 2002; Carey, 1985; Keil, 1989; Murphy & Medin, 1985). In the domain of mental disorders, Kim and Ahn (2002b) found that lay people's concept of anorexia, for instance, 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.

In the following pages we will first review evidence showing that clinicians' representations of mental disorders are not lists of independent symptoms but rather theorylike structures connecting these symptoms. Afterwards, we will discuss the specific effects of these theories on clinicians' reasoning. First, we will examine how clinicians' theories determine which symptoms are perceived as more central or important than others are. Second, we will discuss how causal relations among symptoms influence clinicians' overall perception of how normal or abnormal patients are. Along the way, we will also contrast these findings with lay people's theories and the influence of those theories on reasoning. Finally, we will discuss whether being influenced by theories is rational.

Clinicians Represent Mental Disorders as Theories

Across three experiments, we (Kim & Ahn, 2002a) presented 35 expert clinical psychologists and 25 novice clinical trainees with lists of the symptoms of nine

mental disorders and asked them to specify the relations among the symptoms within each disorder. In these studies, we examined five mental disorders that were judged to be highly familiar even to undergraduates (anorexia nervosa, schizophrenia, major depressive episode, antisocial personality disorder, and specific phobia) in Experiments 1 and 2, and four personality disorders (avoidant, schizotypal, borderline, and obsessive–compulsive personality) in Experiment 4. In Experiment 1, participants were asked to describe their causal theory of each disorder by drawing an arrow between symptoms, pointing from cause to effect. In Experiments 2 and 4, participants were asked to draw any relations between symptoms as they saw fit, not limiting themselves to causal relations only.

Except for one novice, all participants opted to draw relations between symptoms on either all disorders or all but one disorder presented to them. Furthermore, participants drew fairly complex structures among symptoms (56.6 arrows per disorder per participant across the three experiments). Given these results, it would be difficult to argue that clinicians represent mental disorders as lists of independent symptoms as in the *DSM–IV*.

Are any special kinds of relations particularly prevalent in clinicians' theories? When participants were allowed to draw any relations among symptoms, we found that 76% of all relations that our participants drew were causal relations or relations that imply causality (e.g., "affects," "decreases," "determines," "enables," "leads to"). This result confirms previous suggestions that causality lies at the core of theory representations (Carey, 1985; Wellman, 1990). In addition, this result indicates that the relations between symptoms that participants reported were not simply statistical co-occurrences among symptoms.

How idiosyncratic are these theories? Using Kendall's *W* as a measurement of interjudge reliability, we determined the extent to which participants agreed on the relative causal centralities of symptoms.[2] Within each experiment, participants' agreement among themselves, as measured by Kendall's *W*, ranged from .10 to .38. Table 15.1 lists *W*s for each disorder. Because each value of *W* obtained in these studies was statistically greater than 0, it is difficult to assert that participants have theories that are completely or vastly different from one another. Instead, significant but moderate consensus on the causal centralities of symptoms was observed. This degree of consensus is particularly remarkable in that the participants in Experiments 2 and 4, from which these measures were taken, varied widely in their theoretical orientations (eight psychoanalytic–humanistic clinicians, 19 cognitive–behavioral clinicians, and 12 clinicians of other theoretical orientations).

[2]Consider one set of causal relations between symptoms reported by a hypothetical participant, shown in Figure 15.1. In this structure, Symptoms A and B have the highest causal centrality and are therefore assigned the rank of 1, Symptom C receives the causal centrality rank of 3, and so on. A different subject might draw different causal relations among symptoms, resulting in correspondingly different rank orderings of causal centrality. Kendall's W measures the degree of agreement among multiple judges on these rank orderings. W is related to the correlation coefficient and can range between 0 and 1.

Table 15.1. Consensus in Theories Among Clinical Psychologists

Disorder	Kendall's W (all $p < .05$)
Phobia	.38
Schizophrenia	.33
Borderline personality disorder	.33
Antisocial personality disorder	.32
Avoidant personality disorder	.27
Major depression	.26
Anorexia nervosa	.23
Obsessive–compulsive personality disorder	.17
Schizotypal personality disorder	.10

Effect of Theories on Weighting of Symptoms

Thus far, we have described evidence suggesting that clinical experts and trainees both hold complex causal theories about mental disorders. What, then, are the consequences of having these causal theories? Previous studies on categorization have shown that causal theories determine which features in concepts are perceived to be more central (Ahn, 1998; Ahn, Kim, Lassaline, & Dennis, 2000). In particular, it has been shown that features causally central to an individual's theory of that category are treated as more important in categorization than less causally central features (or the causal status effect). In Ahn et al. (2000), participants read about three characteristic features of a target category (e.g., animals called "roobans" tend to eat fruit, have sticky feet, and build nests in trees). When told that one feature tends to cause the second feature, which in turn tends to cause the third feature (e.g., eating fruit tends to cause roobans to have sticky feet, and having sticky feet tends to allow roobans to build nests in trees), the deepest cause in the causal chain influenced categorization most, whereas the terminal effect influenced categorization least. In another experiment, participants preferred to categorize objects on the basis of a common cause rather than a common effect. For instance, a woman who is depressed because she has low self-esteem was more likely to be categorized with a woman who is defensive because she has low self-esteem than with a woman who is depressed because she has been drinking.

The tendency to weigh causes more than effects in classification is rampant in real-life situations. DNA structure causes many other properties of plants and animals and is therefore considered the most important feature in their classification (e.g., a plant that is found to lack tulip DNA will never be classified as a true tulip). In law, the severity of the crime often depends more on the suspects' intentions than on their surface behaviors (e.g., killing someone by accident is a much less serious offense than intending to kill someone but inadvertently botching the plan).

One major reason to expect the causal status effect comes from the literature on psychological essentialism. Essentialism in the purely philosophical sense states that objects have essences that make them the objects they are

(Kripke, 1971; Locke, 1894/1975; Putnam, 1977). Whether or not this metaphysical claim is true, Medin and Ortony (1989) proposed that people act as if things have essences, a doctrine the authors called *psychological essentialism*. According to this view, essences in one's conceptual representations are believed to generate, cause, or constrain surface features of objects (Gelman & Wellman, 1991; Medin & Ortony, 1989). Hence the deepest known causal property of an entity might be a person's best guess as to its essence.

In the domain of mental disorders, Haslam and his colleagues have demonstrated that laypeople treat psychiatric categories as natural kinds and demonstrate essentialist thinking about mental disorders (e.g., Haslam, 2000; Haslam & Ernst, 2002). Kim and Ahn (2002b) also found that in the diagnosis of mental disorders undergraduate students weigh symptoms that cause other symptoms more heavily than symptoms that are caused by other symptoms.

The more critical question is whether professional mental health practitioners also demonstrate such a bias because, as mentioned earlier, the *DSM* system attempts to be theory-neutral. More specifically, the *DSM* system assumes, with a few explicit exceptions, that all symptoms in a given disorder are equally central to it, regardless of how symptoms might be structured in clinicians' theories. For instance, according to the *DSM–IV*, there are four symptoms that must all be present to warrant a diagnosis of anorexia nervosa, making these four symptoms equally central. However, according to the clinicians' data collected in our experiments, "distorted body image" was most causally central in the clinicians' theories. According to the causal status hypothesis, therefore, we would expect "distorted body image" to be the most central in diagnosis. Similarly, "absence of the period (in women) for more than three menstrual cycles" was rated by clinicians as the most causally peripheral and would therefore be expected to be the least central in diagnosis. We used several converging measures to test this hypothesis (Kim & Ahn, 2002a).

Weighting of Individual Symptoms

First, we asked participants to rate the diagnostic importance of each symptom for each *DSM–IV* disorder. For instance, participants in Experiment 2 considered the question "How important is the symptom of [*Y*] in diagnosing a person with [disorder *X*]?" on a scale of 0 to 100 (0 = *very unimportant*; 100 = *very important*). Then, we correlated these ratings with the symptoms' causal centrality ranks derived from the clinicians' theories. For example, as noted earlier, "distorted body image" in anorexia nervosa was causally central, causing many symptoms, including the fear of being fat and excessive exercise and dieting. We found that this symptom was considered to be diagnostically important as well (92.1). On the other hand, "absence of the period (in women) for more than three menstrual cycles," another diagnostic criterion for anorexia nervosa, was rarely judged to cause any other symptoms of that disorder and was also considered to be less diagnostically important (74.4). If clinicians strictly adhered to the *DSM–IV* guidelines, the correlation between these two measures should be zero. However, across three experiments, these correla-

tions were found to be significantly positive, suggesting that diagnosis could be reliably predicted by causal centrality.

Diagnosis of Hypothetical Patients

Second, hypothetical patient descriptions were developed, and participants judged the likelihood that each patient had a certain disorder (Experiment 1) or goodness of fit in the diagnostic category of a certain disorder (Experiments 2 and 4). To test the effect of causal theories, three types of patient descriptions were developed, all individually tailored to each participant's theory of the specific disorder. To illustrate this, consider the schematic of a participant's theory shown in Figure 15.1. One of the patient description types consisted of symptoms that were causally central in a participant's theory (e.g., symptoms A, B, and C), the second type consisted of symptoms that were causally peripheral in the participant's theory (e.g., symptoms E, F, and G), and the third type consisted of symptoms that neither cause nor are caused by other symptoms (e.g., symptoms H, I, and J). For each disorder, the number of *DSM–IV* diagnostic criteria symptoms was equated between patients. Thus diagnoses based strictly on the *DSM–IV* would not differentiate the three types of hypothetical patients. However, participants' mean ratings varied as a function of the causal centrality of the hypothetical patients' symptoms. The results, collapsed across the three experiments (1, 2, and 4), are presented in Figure 15.2. Patients with causally central symptoms were judged to be more likely to have a target disorder or to better fit the target disorder than were patients with causally peripheral symptoms, demonstrating the causal status effect. In addition, patients with isolated symptoms, or symptoms not thought to be causally connected to any other symptoms (e.g., Gentner, 1989), received the lowest ratings of all.

In the previous section, we described how clinicians appear to agree to some degree as to which symptoms are causally central in mental disorders. The hypothetical patient diagnosis task described in this section was used to assess whether clinicians' theories also concur with laypeople's theories. For each of nine mental disorders used, an average causal structure was first obtained by simply averaging causal strength ratings among all pairs of symptoms. Based

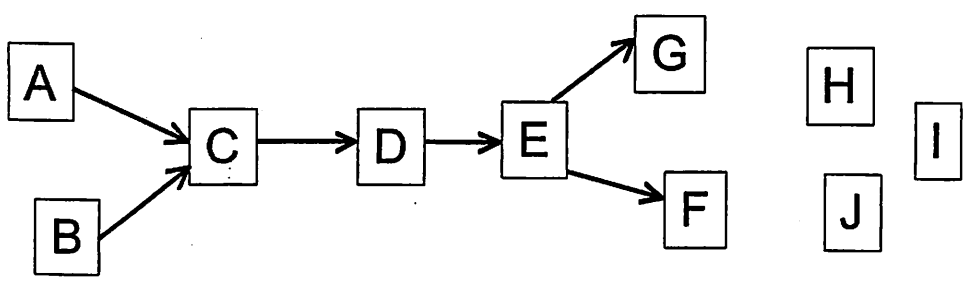

Figure 15.1. Hypothetical causal relations among symptoms.

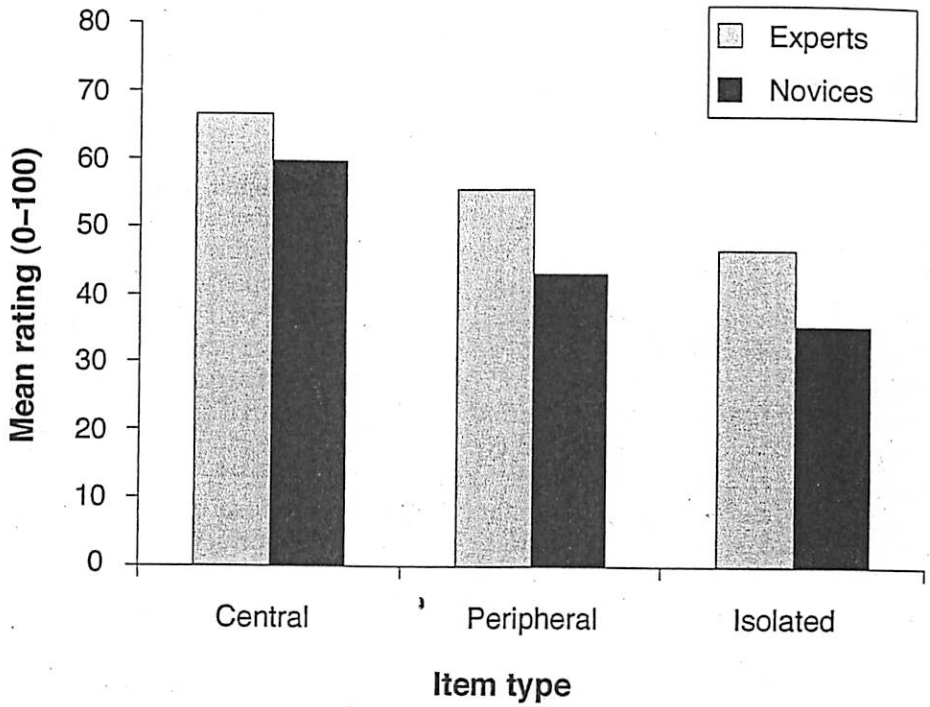

Figure 15.2. Clinical psychologists' and clinical psychology graduate students' mean ratings for hypothetical patients collapsed across Experiments 1, 2, and 4 (Kim & Ahn, 2002a).

on these averaged theories, hypothetical patients with causally central symptoms and hypothetical patients with causally peripheral symptoms were developed for each disorder. These descriptions were presented to undergraduate students, who were asked to judge how well each description fit with the corresponding disorder. The critical question was whether causal centrality as reported by clinical psychologists and trainees could predict typicality judgments made by undergraduate students. The results showed that undergraduate students rated causally central patients to be much more typical of the disorder in question than causally peripheral patients. This effect was present in all mental disorders used in the study except for borderline personality. In addition, the effect was greater for disorders that undergraduate students reported themselves as familiar with (mean ratings of 75.3 vs. 29.3 for causally central and causally peripheral patients, respectively) than for disorders with which they are unfamiliar (mean ratings of 75.9 vs. 64.8). It remains to be seen why experts' theories generally concur with commonsense theories of mental disorders. 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 the media or education.

Memory for Symptoms

About 1 to 2 hours after participants made these judgments concerning the hypothetical patients, they received a surprise memory task. The results showed that clinicians' memory for patients' symptoms also varied as a function of the causal centrality of symptoms. Figure 15.3 shows participants' recall data collapsed across Experiments 1 and 4 of Kim and Ahn (2002a). Again, causally central symptoms were most likely to be recalled, and isolated symptoms were least likely to be recalled. Furthermore, false alarm data from a recognition task also showed the influence of causal theories. Specifically, participants were more likely to falsely recognize causally central symptoms

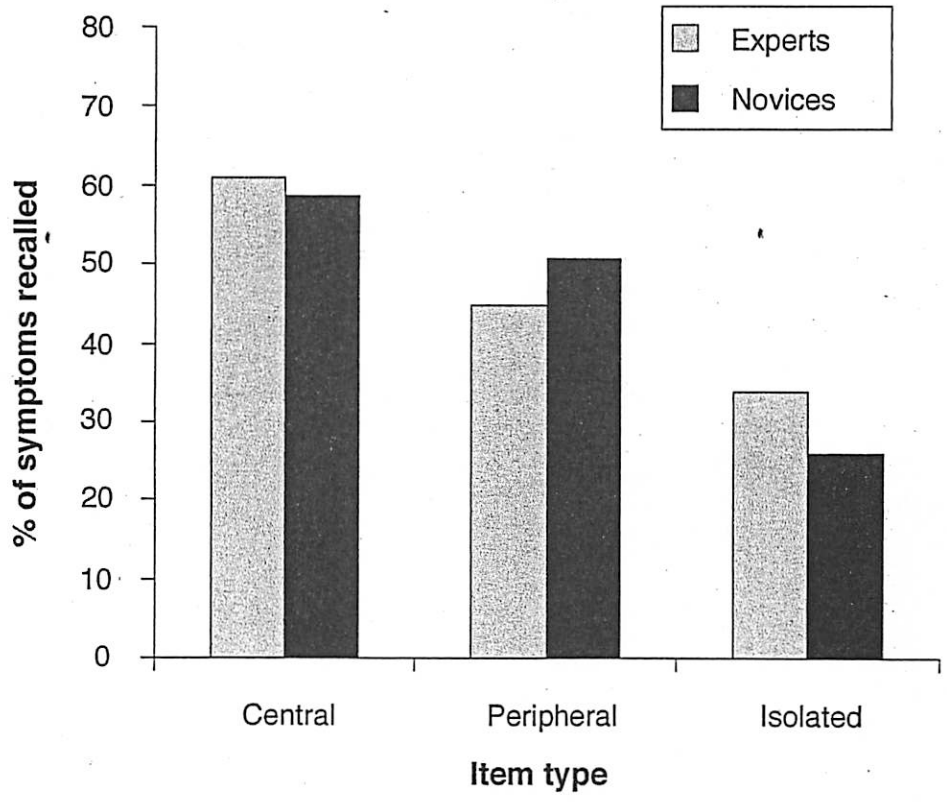

Figure 15.3. Percentages of symptoms correctly recalled from hypothetical patients with potential *DSM–IV* disorders seen before a time delay (Kim & Ahn, 2002a).

than causally peripheral or isolated symptoms as having been present earlier in the hypothetical patient's task.

Summary

It is amusing to note that the main results from our study (Kim & Ahn, 2002a) can be summarized by quoting Medin's 1989 *American Psychologist* article:

> The *DSM–III–R* guidebook (American Psychiatric Association, 1987) 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. Symptoms differ in the level of abstractness and the types and number of intersymptom relations in which they participate, and as a consequence, they differ in their centrality. (p. 1479)

This was a description of a preliminary study that the first author of this chapter was carrying out as a graduate student with Medin. Unfortunately, the project was never completed. We are glad to finally report back to Doug Medin 15 years later that his prediction was right all along.

Overall Perception of Normality

In our recent study (Ahn, Novick, & Kim, 2003) we also examined the degree to which causal explanations influence clinicians' overall perception of a person's normality. Meehl (1973) noted that when clinicians felt that they understood a patient, the patient seemed normal—that is, "understanding it makes it normal" (p. 244).

To empirically test this hypothesis, Ahn et al. (2003), in their Experiment 1, provided participants (all laypeople) with descriptions of people and asked them to judge how normal these people were. Each description identified three symptoms, taken from different disorders of the *DSM–IV*, that were manifested by the person. In the two experimental conditions, participants were further provided with explanations about how these three characteristics were causally related. These causal relations were either plausible (e.g., "Because Penny frequently suffers from insomnia and is in a habitual state of sleep deprivation, she has trouble remembering the names of objects. This memory problem, in turn, leads her to suffer from episodes of extreme anxiety because she fears that it will cause her to embarrass herself in front of others") or implausible (e.g., "Because Jarrod always chooses solitary activities, he requires excessive attention to make up for the lack of human contact. This need, in turn, causes him to be unable to remember new information because he relies on the attention of others to remember all the important information for him"). Participants in the control condition did not receive any causal explanations. The results showed that the presence of a causal explanation altered the perception of normality of the patients. Specifically, receiving plausible causal explanations made participants judge the people to be more

normal than they appeared to control participants, whereas implausible causal explanations made the same people look more abnormal than they appeared to control participants.

Often, explanations that people interject about symptoms are not simply about causal relations among the symptoms but also about what precipitated these symptoms in the first place. Ahn et al. (2003) also examined the effects of knowing the deeper cause of a group of symptoms in both undergraduates and clinical psychologists. One group of participants received the descriptions of three causally related symptoms (both plausible and implausible ones) used in Ahn et al.'s aforementioned Experiment 1. The other group of participants received additional causal explanations for the root symptom in each of these causal chains. For instance, the phrase "because she is very stressed out due to her workload" was added as an explanation for why "Penny frequently suffers from insomnia." Similarly, the phrase "because he was bullied a lot by his classmates when he was young" was added as an explanation for why "Jarrod always chooses solitary activities." As shown in Figure 15.4, participants who received the additional life-event root cause explanations judged that these people were more normal than did participants who did not receive such explanations.

Meehl (1973) treated this phenomenon as a reasoning fallacy, listing it as one reason that he does not attend psychiatric case conferences, where such behavior is common. Understanding why a man killed his wife, for instance, does not make the act a normal, excusable one. A clinical trainee participant in the pilot study of Ahn et al. (2003) shared the same sentiment. This pilot study was designed as a within-subject manipulation in which participants

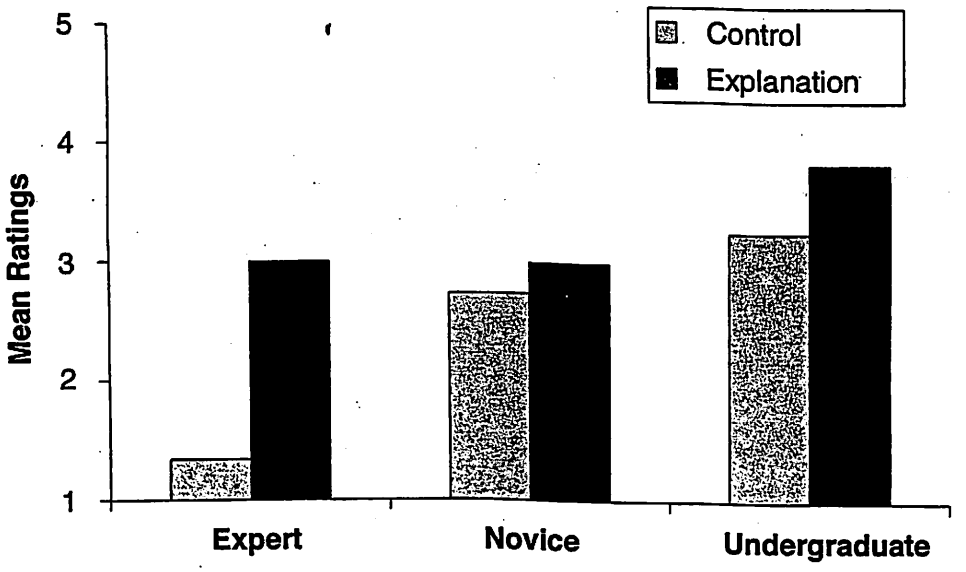

Figure 15.4. Clinical psychologists', clinical graduate students', and undergraduate students' mean judgments of people's normality, broken down by condition. Higher ratings indicate higher ratings of normality; lower ratings indicate greater abnormality.

were asked to rate the normality of patients described without the life-event explanations and then to rate them again with the life-event causes revealed. This participant refused to change any of her ratings in the second round (i.e., in the life-event cause condition) because she felt that it would be absurd to change her ratings of psychological normality simply because she now knew the underlying causes. In other words, when the two conditions were put before her, the fallacy of the "understanding it makes it normal" phenomenon became extremely clear.

To summarize, the results from Ahn et al. (2003) demonstrate that the plausibility of causal explanations or the mere presence of an underlying causal theory for symptoms can have a significant impact on perceptions of normal patients. Even expert clinicians are swayed by causal explanations in the same manner as undergraduates.

Implications for the *DSM*

The results reviewed in this paper suggest that the *DSM*'s attempt to discourage mental health practitioners from using their own theories has not been quite successful. One might argue that in real-life cases, clinicians are fairly guarded against the influence of their background theories because they must make formal *DSM* diagnoses using checklists. However, we believe that the effect of theory-based conceptual representations found in the current studies may still pervade critical aspects of clinical thinking. For instance, as shown in Kim and Ahn (2002a), clinicians are better at recalling symptoms that are central to their theories, and these biases may cause them 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 the way 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. Theory-based reasoning per se is not necessarily irrational, as we will explain later, but if the theories that clinicians use are invalid, dire consequences may result.

Do these results have implications for how the *DSM* should be revised? Although professional mental health practitioners' reasoning about mental disorders appears to be theory-based, it does not necessarily follow that the *DSM* system should be modified to be theory-based. Clinicians' mental representations of mental disorders may or may not have bearing on the *DSM* system depending on whether categories of mental disorders are thought of as natural or nominal kinds (e.g., Dawes, 1994).

One view would be that mental disorders are natural kinds—the kinds that exist out in the world, waiting to be discovered by humans. Under this assumption, a scientist's role would be to "carve nature at the joints," and thus the specific personal theories about mental disorders that practicing clinicians hold would be irrelevant to the truth yet to be discovered. To claim that it is relevant might sound as absurd as claiming that the periodic table should be modified to fit the way chemists reason about elements.

The opposite view would be that mental disorders are nominal kinds—the conventional kinds that are constructed by humans. According to this view, mental disorders are categories that our culture imposes upon the world. For instance, the past debate on whether homosexuality is a mental disorder had more to do with societal prejudices than with clinical psychology or psychiatry. Indeed, the American Psychiatric Association holds a copyright over the mental disorders described in the *DSM,* such that one must pay a copyright fee to publish descriptions of those criteria. By definition, this implies that mental disorders are invented. If this view is correct, then the way in which people, especially professionals, think about mental disorders is of great importance in determining how we classify and describe mental disorders. According to an extreme version of this view, all mental disorders should be classified by consensus.

Another logical possibility is a mixture of these two views: Some natural patterns characterize mental disorder symptoms, but it is up to humans to determine exactly where to draw the boundaries. Furthermore, we often lack a clear understanding of each disorder's etiology and underlying pathology, especially in the domain of mental disorders. Despite this paucity of information about etiology, clinicians must create a nosology of disorders to ensure diagnostic reliability and establish a common ground for discussion among clinicians and clinical researchers. The result is a classification system determined to a significant extent by votes among committee members. To the extent that culture currently plays a part in classification systems, recognizing how members of the culture understand the domain is vital.

Rationality

An inevitable question at this point is whether being influenced by causal theories is rational. Although it deviates from the atheoretical *DSM* model, theory-based reasoning per se is not necessarily irrational. In fact, scientific progress is often indicated by a shift from the amassing of experimental data and observations toward the eventual development of theories in terms of unobservable entities (Hempel, 1965). For instance, the early taxonomic systems of biological organisms were based on observable (largely morphological) characteristics, but with the development of the theory of evolution, the morphological basis was replaced by a phylogenetic basis.

Categorization based on theories rather than surface features is considered to be more scientifically fruitful—that is, it provides a basis for explanation, prediction, and generally scientific understanding (Hempel, 1965). A recent example in the medical domain clearly illustrates this point. Cancers have been categorized mainly by the area where they originate in the body—breast, colon, and so on. However, there has been a trend to reclassify cancer on the basis of genetic characteristics as scientists gain more understanding of their inner workings. Thus instead of saying a person has lung cancer, one might say that the person has cancer with particular genetic characteristics that just happen to be in the lung. The reason for this reclassification attempt is improved predictability. For instance, breast cancer patients at the same stage of disease can have markedly different treatment responses and overall

outcome, depending on breast cancer gene expression. A genetic analysis of breast tumors can be used to predict which patient should additionally receive chemotherapy after tumors have been removed (van't Veer et al., 2002).

Describing disorders at the symptom level is not necessarily invalid. Often, surface features are valid indicators of underlying deeper dimensions (e.g., Gelman & Medin, 1993). For instance, facial hair and a deep voice are good diagnostic cues for inferring the presence of male chromosomes. The symptom-level descriptions adopted in the recent *DSM* system are justified by Kraepelin (Kihlstrom, 2002), who argued:

> There is a fair assumption that similar disease processes will produce identical symptom pictures, identical pathological anatomy, and an identical etiology. If, therefore, we possessed a comprehensive knowledge of any one of these three fields—pathological anatomy, symptomatology, or etiology—we would at once have a uniform and standard classification of mental disease. (Kraepelin & Diefendorf, 1904/1907, p. 117)

However, without a clear understanding of underlying mechanisms, simple correlational structures have weak predictive power. For instance, we can redefine the early stages of pregnancy using the *DSM–IV* format as displaying 7 out of the following 10 symptoms: extreme fatigue, increased sense of smell, weight gain, missed period, weepiness, hunger, nausea and vomiting, heartburn, increased urination, constipation. Given this description, it is impossible to predict that the person will have a baby.

A number of theorists have discussed reasons to base taxonomies on theory. As previously mentioned, the *DSM*'s purpose in not specifying underlying theories is to avoid battles between different theoretical schools as to which theories should be included or emphasized in the manual. However, some argue that the advantage gained by such a solution is outweighed by the negative effects of an atheoretical taxonomy on clinical research (Follette & Houts, 1996). Although advancing research is ostensibly not the primary purpose of the *DSM* system, research will ultimately be the most important determinant of the way mental disorders are defined. Follette (1996) offers an additional criticism of the diagnostic criteria, pointing out that they never consider behavior within a context (i.e., both situational and biological). Moreover, in being simplified, the diagnostic criteria have become a much more crude representation of the disorders than their actual manifestation (Carson, 1996). Carson's 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 that 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.

References

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

Ahn, W.-k., Novick, L., & Kim, N. S. (2003). "Understanding it makes it normal": Causal explanations influence person perception. *Psychonomic Bulletin and Review, 10,* 746–752.

American Psychiatric Association. (1952). *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.

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.

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

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.

Gelman, S. A., & Medin, D. L. (1993). What's so essential about essentialism? A different perspective on the interaction of perception, language, and conceptual knowledge. *Cognitive Development, 8,* 157–167.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the nonobvious. *Cognition, 38,* 213–244.

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

Haslam, N. (2000). Psychiatric categories as natural kinds: Essentialist thinking about mental disorder. *Social Research, 67,* 1031–1058.

Haslam, N., & Ernst, D. (2002). Essentialist beliefs about mental disorders. *Journal of Social and Clinical Psychology, 21,* 628–644.

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

Joiner, T. E., & Schmidt, N. B. (2002). Taxometrics can "do diagnostics right" (and isn't quite as hard as you think). In L. E. Beutler & M. L. Malik (Eds.), *Rethinking the DSM: A psychological perspective* (pp. 107–120). Washington, DC: American Psychological Association.

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

Kihlstrom, J. (2002). To honor Kraepelin . . .: From symptoms to pathology in the diagnosis of mental illness. In L. E. Beutler & M. L. Malik (Eds.), *Rethinking the DSM: A psychological perspective* (pp. 279–303). Washington, DC: American Psychological Association.

Kim, N. S., & Ahn, W.-k. (2002a). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology: General, 131,* 451–476.

Kim, N. S., & Ahn, W.-k. (2002b). The influence of naive causal theories on lay concepts of mental illness. *American Journal of Psychology, 115,* 33–65.

Kraepelin, E., & Diefendorf, A. R. (1907). *Clinical psychiatry: A text-book for students and physicians* (7th ed.). New York: Macmillan. (Original work published 1904)

Kripke, S. (1971). Naming and necessity. In D. Davidson & Harman (Eds.), *Semantics of natural language* (pp. 253–355). Dordrecht, the Netherlands: D. Reidel.

Lin, E. L., & Murphy, G. L. (1997). Effects of background knowledge on object categorization and part detection. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1153–1169.

Locke, J. (1975). *An essay concerning human understanding.* London: Oxford University Press. (Original work published 1894)

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

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). New York: Cambridge University Press.

Meehl, P. E. (1973). Why I do not attend case conferences. In P. E. Meehl (Ed.), *Psychodiagnosis: Selected papers* (pp. 225–302). Minneapolis: University of Minnesota Press.

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

Putnam, H. (1977). Is semantics possible? In S. P. Schwartz (Ed.), *Naming, necessity, and natural kinds* (pp. 102–118). Ithaca, NY: Cornell University Press.

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.

Shorter, E. (1997). *A history of psychiatry: From the era of the asylum to the age of Prozac.* New York: Wiley.

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

van't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature, 415,* 530–536.

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

Categorization Inside and Outside the Laboratory

Essays in Honor of Douglas L. Medin

Edited by
Woo-kyoung Ahn, Robert L. Goldstone,
Bradley C. Love, Arthur B. Markman, and Phillip Wolff

American Psychological Association • Washington, DC