Theory-Based Reasoning in Clinical Psychologists

Nancy S. Kim (nancy.s.kim@yale.edu)
Department of Psychology, Yale University; 2 Hillhouse Avenue
New Haven, CT 06520 USA

Woo-kyoung Ahn (woo-kyoung.ahn@vanderbilt.edu)
Department of Psychology, Vanderbilt University; 534 Wilson Hall
Nashville, TN 37240 USA

Abstract

Progress in science is marked by the formation of theories that explain a body of observations. Contrary to this guiding philosophy, clinical psychologists have prescribed to an atheoretical system of diagnostic reasoning since 1980. We report two studies demonstrating that clinicians have not internalized atheoretical reasoning despite the system's widespread acceptance. The results show that clinicians' own theories about disorders significantly affect their diagnoses of hypothetical patients and memory for symptoms. Clinicians are cognitively driven to form and apply theories to observations despite decades of deliberate training, practice, and pressure to reason atheoretically.

Introduction

Philosophers of science have argued for decades that scientific progress is delineated not merely by the amassing of observations, but rather by the formation of broad principles that organize and explain these observations in a cohesive manner (Hempel, 1965). Evolutionary theory, for example, is regarded as a revolutionary advance in organismic biology in large part because it provides a deeper structure for a scientific taxonomy of living things, moving away from classification based on superficial features. Data are measured and considered within the larger structure of this overarching theory. Similarly, the human mind constantly seeks out underlying rules and principles that make sense of incoming data concerning the surrounding world. Adults (Murphy & Medin, 1985), children (Gelman, 2000; Keil, 1989), and even infants (Marcus, Vijayan, Rao, & Vishton, 1999) spontaneously extract and apply underlying organizing principles and abstract rules that go beyond surface features. In this way, the human mind forms categories and concepts based on its theories about the surrounding world (Carey, 1985).

Atheoretical versus theory-based reasoning

In contrast, the current *Diagnostic and Statistical Manual of Mental Disorders* (APA, 1994), prescribes an atheoretical approach to diagnosing 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 *DSM-IV*'s (APA, 1994) widely acclaimed solution is to eliminate theory use altogether when making a diagnosis, incorporating instead checklists of symptoms compiled by a panel of experts. In doing so, it represents each disorder as a list of unrelated symptoms, ignoring the causal relations between symptoms that are a fundamental aspect of theory representations (Carey, 1985). For most disorders, the *DSM-IV* (APA, 1994) states that a subset of the list is sufficient for a diagnosis regardless of which combination of symptoms appears, thereby assuming that all symptoms in the list are equally central to the disorder. For example, any 2 of the following 5 symptoms warrant a diagnosis of schizophrenia, according to the *DSM-IV*: hallucinations, delusions, disorganized speech, grossly disorganized or catatonic behavior, and negative symptoms. Since eliminating any overt mention of an underlying theory for the taxonomy two decades ago (APA, 1980), the *DSM* system has become widely accepted in the U.S., 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* (APA, 1994).

Has the *DSM* system succeeded in internalizing atheoretical reasoning in clinicians? An atheoretical approach would suggest that experienced clinicians, after years of emphasizing use of the *DSM* system, will come to embody its prescription of atheoretical reasoning (APA, 1994). In contrast, the theory-based approach would suggest that clinicians, despite such emphasis on the elimination of theory, are still influenced by their own idiosyncratic theories about disorders when reasoning about them (Medin, 1989).

The two approaches were differentiated by testing for the presence or absence of the causal status effect, a specific mechanism by which theory-based reasoning occurs (Ahn, 1998; Ahn, Kim, Lassaline, & Dennis,

2000). The causal status effect is said to occur when category 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. To derive the causal centralities of individual symptoms imbedded in a complex theory, the following formula[1] can be used:

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

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 upon feature i. Thus, in the current studies we operationalized the theory-based view as a systematic effect of relational structures on conceptual representation and use.[2]

A theory-based view would further predict that any features relationally connected to other features would be treated as more important than isolated features in reasoning (Gentner, 1983). That is, if symptom A causes symptom B, but symptom C is isolated (it does not and is not caused by any other symptoms in a clinician's theory), C would be the least central symptom of the three.

Study 1

We measured expert and trainee clinical psychologists' causal theories in the first session. Then we examined whether the causal centralities of symptoms in their theories predict how important these symptoms are in diagnosis (in the first and second sessions), and how well they are remembered (in the second session).

Participants

Participants were 11 experienced clinical psychologists, and 10 clinical psychology graduate students. The experienced clinical psychologists had been in practice for a minimum of 15 years (ranging from 15 - 52 years with a median of 28 years). Ten were licensed psychologists with Ph.D.'s, and 1 was a board-certified psychiatrist with an M.D.

Session 1: Measurement of causal theories and conceptual centrality

We measured participants' individual causal theories for each of 5 disorders that were judged to be highly familiar by undergraduate students. The 5 disorders were Anorexia Nervosa, Antisocial Personality Disorder, Major Depressive Episode, Specific Phobia, and Schizophrenia.

In an initial disorder defining task, participants viewed a list of symptoms for each disorder. Symptoms included both the *DSM-IV* (APA, 1994) diagnostic criteria and the non-criterial, characteristic symptoms from the manual's disorder description. Participants were asked to define each disorder for themselves by adding new symptoms, crossing out symptoms, combining two or more symptoms, and / or dividing a single symptom into two or more symptoms.[3] All subsequent tasks in both sessions incorporated these individually tailored lists.

Participants' causal theories were then measured for each disorder. Participants received slips of paper, each bearing the name of a symptom. They were first asked to arrange the symptoms around the corresponding disorder name. Next, participants drew arrows between symptoms to indicate causal relations as they thought was appropriate. Finally, they rated the strength of each causal relation on a scale of 1-5 (1=very weak; 5=very strong). From these causal drawings, we determined the causal centrality of each symptom using Equation (1). Isolated features were always assigned the lowest causal centrality.

During this session, we also measured the conceptual centrality of each symptom to the disorder. Clinicians were asked, "how easily can you imagine a person with [disorder X] who does not have the symptom of [Y]?" for each symptom on a scale of 0-100 (0=very difficult to imagine; 100=very easy to imagine). The order of the two tasks, conceptual centrality and causal theory measurement, was counterbalanced between participants. The results demonstrated that conceptual

[1] 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 analyses on causal centrality reported below are based on rank orders of causal centrality derived from this formula, and different formulas do not produce radically different rank orders.

[2] We do not intend to claim here that theory-based categorization is limited to the effect of relational structures. Categorization may also be affected by the content of relations, an issue that was not the focus of the current studies.

[3] For instance, a participant might choose to divide the single symptom "disturbed experience of body shape or denial of the problem" into separate symptoms ("disturbed experience of body shape;" "denial of the problem").

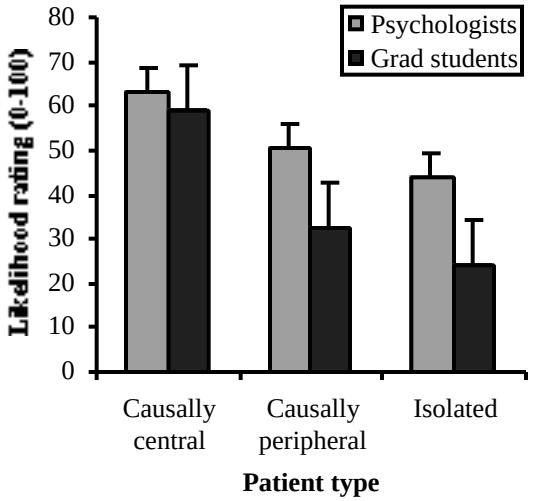

Figure 1. Clinical psychologists' and clinical graduate students' likelihood ratings of mental disorder category membership for hypothetical patients in Study 1. Error bars indicate standard errors.

centrality was positively correlated with causal centrality (as determined by Equation [1]) for 18 out of 20 participants,[4] such that the more causal a symptom was within each participant's theories, the more central the symptom was to that participant's concept of the disorder. The median overall correlation coefficients for clinical psychologists and graduate students were .41 (range: -.12 to .50) and .27 (range: .11 to .62), respectively.

Session 2: Hypothetical patient diagnosis and free recall of symptoms

Participants were brought back to the lab about 10 - 14 days after the first session. For the second session, we created a unique set of hypothetical patients for each participant based on that participant's own theories as reported in the causal centrality task. Two to three hypothetical patients were constructed for each of the 5 mental disorder categories. Each patient was composed of three symptoms that were either causally central or causally peripheral as determined by Equation (1), or isolated[5]. For example, causally central patients consisted of three symptoms causally central to the participant's theory of the disorder (i.e., they were thought to cause more symptoms more strongly than other symptoms did). Participants were told that these patients did not exhibit any other symptoms. The number of *DSM-IV* (APA, 1994) diagnostic criteria was equated over the types of patients. Therefore, according to strict *DSM-IV* (APA, 1994) criteria, each of the three different types of hypothetical patients should be equally likely to have the disorder in question.

We asked participants to rate the likelihood that each of these hypothetical patients actually had the associated disorder. Specifically, for each hypothetical patient, participants answered the question, "what is the likelihood, in your opinion, that a patient with the following characteristics has [disorder X]?" on a scale of 0-100 (0=very unlikely; 100=very likely).

Although the number of criterial symptoms according to the *DSM-IV* (APA, 1994) was equated across the three types of patients, participants judged patients with causally central symptoms (mean of 61.0) as nearly 20% more likely to have the disorder than patients with causally peripheral symptoms (mean of 42.0; $t = 4.5$, $p < .001$). Patients with isolated symptoms were judged as least likely to have the disorder (mean of 34.5; $t = 3.3$, $p = .003$). Figure 1 shows the results broken down by expert and trainee participant groups. There was neither a significant main effect of expertise nor any significant interaction involving expertise.

Approximately one hour after they completed the hypothetical patient diagnosis task, participants were asked to recall the symptoms of those hypothetical patients. Participants recalled significantly more causally central (67%) than causally peripheral (51%; $t = 2.9$, $p = .009$) or isolated (44%; $t = 2.7$, $p < .02$) symptoms.

Study 2

Study 2 expanded the generality of these findings using modified procedures. There were two principal changes. First, the causal centrality task was modified to measure participants' theories about all kinds of symptom-symptom relations, not restricting the measure to causal relations only. Second, another aspect of memory for symptoms was examined by using a recognition task instead of the recall task. These changes will be described in detail in the following sections.

Participants

Participants were another group of 14 experienced clinicians and 6 clinical psychology interns. The expert clinicians had been in practice for a minimum of 15 years (ranging from 17-43 years with a median of 26 years). All 14 expert clinicians were licensed psychologists; 13 had Ph.D.'s and 1 had an Ed.D.

[4] The data of one graduate student participant could not be included in this particular analysis because there was no variance among that participant's conceptual centrality responses within each disorder.

[5] Some participants left no symptoms isolated in their theory of a disorder. In these cases, a patient description composed of isolated symptoms could not be created.

Stimulus materials

The same 5 disorders in Study 1 were also utilized in this study. Unlike in Study 1, however, participants were provided with a list of "standard" symptoms. These symptoms were defined by the participants in Study 1. Namely, a symptom was dropped from the list of symptoms for Study 2 if it was dropped by over 50% of the experts and over 50% of the trainee participants in Study 1. Using these standard lists of symptoms allowed us to make direct comparisons between participants' theories, especially between those of experts and trainees.

Session 1: Measurement of relational theories and conceptual centrality

We measured participants' individual theories for each disorder using the same procedure as before, except that this time we asked participants to draw any kind of relations between symptoms they saw fit, not limiting the measure to causal relations. Participants rated the strength of each relation on a scale of 1-3 (1= weak; 2=moderate; 3=strong). They 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," "is a cure for." The relational centrality of each symptom was then determined using Equation (1). For instance, in applying the formula, "A is a precondition for B" and "A precedes B" are treated as "B depends on A." We used these relational centralities to further investigate clinicians' and clinical trainees' use of theories in reasoning with three modified measures of diagnostic importance and memory. During the first session, we also measured the diagnostic importance of each symptom to the disorder. Participants were asked, "how important is the symptom of [Y] in diagnosing a person with [disorder X]?" on a scale of 0-100 (0=very unimportant; 100=very important). As in Study 1, the order of the two tasks was counterbalanced between participants. The results demonstrated that diagnostic importance was positively correlated with relational centrality (as measured by Equation [1]) for 18 out of 20 participants (average r = .77, clinical psychologists; r = .66, interns). That is, the more symptoms depended on a symptom, the more central that symptom was to participants' concepts of the disorder. The median overall correlation coefficients for psychologists and graduate students were .25 (range: .02 to .61) and .44 (range: -.94 to .68), respectively.

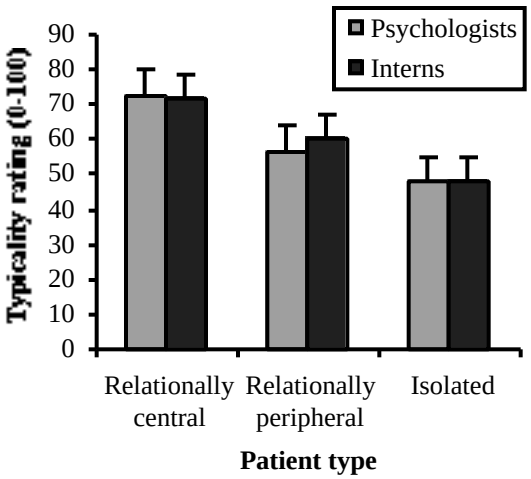

Figure 2. Clinical psychologists' and clinical psychology interns' typicality ratings for hypothetical patients in Study 2. Error bars indicate standard errors.

Session 2: Hypothetical patient typicality and recognition of symptoms

As in Study 1, we constructed 2 to 3 hypothetical patients for each of 5 mental disorder categories. Each patient consisted of a set of three symptoms that were either relationally central, relationally peripheral, or isolated, for each disorder and each participant. As in Study 1, a patient composed of isolated features was not included if a participant did not leave any symptoms isolated in their theory. Again, the number of criterial symptoms was equated between patients so that diagnoses based strictly on the *DSM-IV* (APA, 1994) would not differentiate them. Participants were asked to assess how typical hypothetical patients were of the disorder (following Cantor, Smith, French, & Mezzich, 1980). Specifically, participants answered the standard 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) for each patient. Patients with relationally central symptoms (mean of 72.3) were judged as more typical of the disorder than patients with relationally peripheral symptoms (mean of 57.8; $t = 4.3$, $p < .001$), which in turn were judged as more typical than patients with isolated symptoms (mean of 47.9; $t = 2.7$, $p < .02$). Figure 2 shows the results broken down by experts and trainees. Neither a significant main effect of expertise nor an interaction effect involving expertise was found.

Following an approximately one-hour delay, participants received a standard recognition task (following Roediger & McDermott, 1995), in which they were asked to classify symptoms on a list as "old" or new" based on whether they had seen them earlier in the hypothetical patients task. The list included 30

relationally central and 30 relationally peripheral symptoms. Half of the symptoms in each group were old and half new. Consistent with previous findings showing an effect of schema on false recognition (Bower, Black, & Turner, 1979), participants were much more likely to falsely recognize new, relationally central symptoms as symptoms that they had seen before (23.3%) than new, relationally peripheral symptoms (13.2%; t = 3.0, p = .008). Participants showed greater sensitivity to relationally peripheral symptoms (d' = 2.62) than to relationally central symptoms (d' = 2.04; F[1, 18] = 2.43; p = .01). Thus, participants were less able to distinguish between presented and non-presented relationally central symptoms. False memory has generally been thought to be an issue for patients with psychological disorders or problems (Loftus & Ketcham, 1994). Interestingly, we found that therapists are biased to falsely remember having seen symptoms in their patients that are central to their personal theories about the disorder.

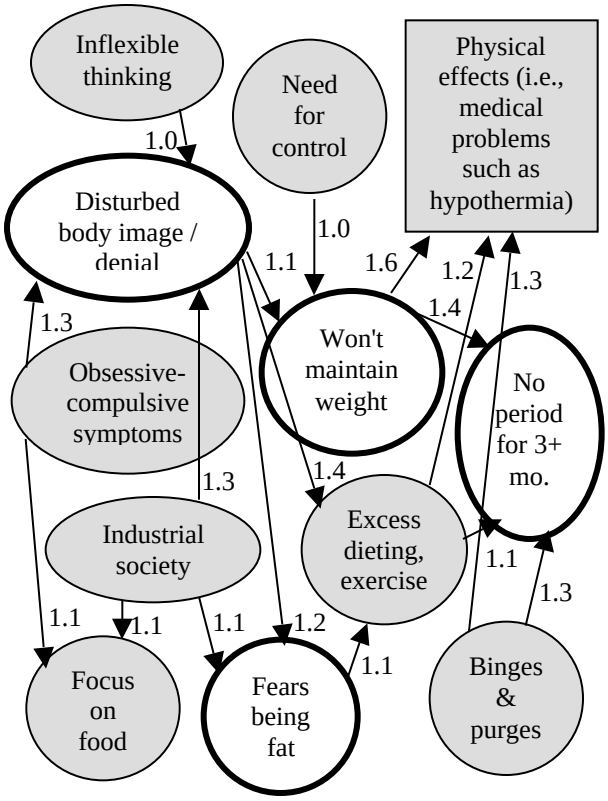

Figure 3. Participants' average relational theory of Anorexia Nervosa in Study 2. (Note: An arrow from A to B indicates that B depends on A [or A affects B]. The symptom descriptions in the figure were truncated and sometimes combined to keep the figure readable. Only dependency relations that received mean strength ratings greater than 1 on a 3-point scale are shown in the figure. Symptoms in white circles with boldface borders are *DSM-IV* (APA, 1994) diagnostic criteria. The *DSM-IV* implies that these are not weighted differently.)

Consensus on theories

Unlike in Study 1, all participants received the same set of symptoms, allowing a direct comparison of their theories. Participants' theories, as measured by relational centrality rank orders, were highly consistent with each other (Kendall's coefficients of concordance ranging from .34 to .59 across the 5 disorders, all p's < .0005). Because of this, we were able to construct an average dependency structure for each disorder, such as the one shown in Figure 3 for Anorexia Nervosa.

These average theories of experts generally agreed with those of lay people. For instance, the mean rank orders of diagnostic criteria symptoms for major depression obtained in an earlier study from undergraduate students (Kim & Ahn, 2001) were highly correlated with those obtained from clinicians and clinical trainees in Study 2 (r = .93, p < .001). We also developed hypothetical patient descriptions based on clinicians' and clinical trainees' averaged theories of the 5 disorders and gave them to 23 undergraduates. Relationally central patients (mean of 75.3) were judged to be more typical of a disorder than relationally peripheral patients (mean of 29.3; t = 12.44, p < .001).

Discussion

In sum, we found that symptoms playing a central role in clinicians' theories were considered to be more important in diagnosis, were more likely to be recalled later, and were more likely to be falsely recognized as having been present in patient descriptions. Despite the fact that the modern *DSM* system, which has become engrained into the conscious practice of clinical psychology, avoids the use of theory, clinicians prefer to base their reasoning on their own organizing theories. This was shown at a much more specific level of analysis than previous work documenting top-down effects of theory on reasoning (i.e., Chapman & Chapman, 1967; Wisniewski & Medin, 1994). Furthermore, in all six measures, there were no differences between levels of expertise, suggesting that years of training and long-term use of the modern *DSM* system do not diminish the effect of theory on reasoning.

Implications

When making formal *DSM* diagnoses using checklists, it is possible that clinicians may not be as strongly affected by their theories. However, the effect of theory-based conceptual representations found in the current studies may still pervade critical aspects of clinical work. As shown in our study, clinicians are

better at recalling symptoms central to their theories, and 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.

We also note that theory-based reasoning in itself is not a reasoning fallacy, provided that clinicians' theories are valid (Dawes, Faust, & Meehl, 1989). However, in the case of less well-known disorders such as personality disorders, experts may have more idiosyncratic theories. This, if true, may account in part for the notoriously low reliability between clinicians in diagnosing the personality disorders. We are currently conducting a study to examine this issue. In cases such as these, reliance on invalid theories may perhaps 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). 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, we found clinicians' theories 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 socio-culturally familiar disorders are not highly idiosyncratic, but rather seem to be based on commonsense notions, and may therefore be worthy of careful consideration in revising the *DSM*.

Acknowledgments

We thank Marvin Chun, Frank Keil, Donna Lutz, Laura Novick, Peter Salovey, and Andrew Tomarken for helpful comments, and Jessecae Marsh and Judy Choi for help in running participants. This research was supported in part by a National Science Foundation Graduate Research Fellowship to Nancy S. Kim and a National Institute of Mental Health Grant (RO1 MH57737) to Woo-kyoung Ahn.

References

Ahn, W. (1998). Why are different features central for natural kinds and artifacts? *Cognition, 69*, 135-178.

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.

American Psychiatric Association (1980). *Diagnostic and Statistical Manual of Mental Disorders*, 3rd Edition. Washington, DC: Author.

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition. Washington, DC: Author.

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

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.

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

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

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. (2000). In H. W. Reese (Ed.), *Advances in child development and behavior*. San Diego, CA: Academic Press.

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

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

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

Kim, N. S., & Ahn, W. (in press). The influence of naive causal theories on lay concepts of mental illness. *American Journal of Psychology*.

Loftus, E., & Ketcham, K. (1994). *The Myth of Repressed Memory*. New York: St. Martin's Press.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by 7-month-old infants. *Science, 283*, 77-80.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist, 44*, 1469-1481.

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

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

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 803-814.

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