

DIAGNOSES AND THE BEHAVIORS THEY DENOTE

A Critical Evaluation of the Labeling Theory of Mental Illness

John Ruscio
Elizabethtown College

The stigma of mental illness is a profound social problem with a long history, and it is widely believed that diagnostic labels cause or contribute to such stigmatization. In an evaluation of labeling theory and the research that it prompted, special attention is devoted to a close examination of 3 widely cited studies (Langer & Abelson, 1974; Rosenhan, 1973a; Temerlin, 1968). Despite a pervasive confounding of diagnostic labels with the behaviors they denote, which increases the apparent influence of “mere labels,” the empirical literature does not support the putative negative effect. To more productively combat the stigma of mental illness, it is suggested that psychologists pursue community-based educational and contact-oriented programs, recognize the unavailability and value of diagnoses, improve diagnostic reliability and validity, and compassionately convey diagnoses in the context of humane and effective treatments.

The societal rejection of individuals with severe mental illness has long been recognized as a significant social problem (e.g., Goodyear & Parish, 1978; Phillips, 1966). Though this stigmatization may have evolved over time to become less negative in some regards, it remains specific to the role of the mentally ill (Fracchia, Canale, Cambria, Ruest, & Sheppard, 1976; Skinner, Berry, Griffith, & Byers, 1995). Given this unfortunate state of affairs, it is important to address questions regarding the stigmatization process. One of the most disputed components of this process involves how mentally ill individuals become targets of negative attitudes: Does their social rejection stem from diagnostic labels or aberrant behaviors?

Because particular labels and behaviors frequently co-occur, our everyday experiences and other informal observations do not necessarily allow us to reliably dis-

tinguish their unique causal influence. Diagnostic labels—whether made on the basis of a formal system such as the *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*; American Psychiatric Association, 1994) or any other observations that allow professionals to differentiate characteristics shared by some, but not all, patients—convey behavioral information, albeit imperfectly, and this information is subject to substantial distortion by those unfamiliar with the true correlates of psychological disorders. Teasing apart the impact of diagnostic labels and the behaviors they denote has serious implications for understanding and alleviating the stigma of mental illness. Thus, it is imperative to carefully examine scientific research on their differentiation, and a wealth of relevant data exist in the psychological, psychiatric, and sociological literatures.

I begin this examination with 3 influential studies that are still frequently cited as support for putative labeling effects (Langer & Abelson, 1974; Rosenhan, 1973a; Temerlin, 1968). Critics of diagnosis argue that studies such as these demonstrate the damaging effects that diagnostic labels can have on patients' lives. A closer look, however, shows that these studies only indirectly or weakly address a labeling effect and, especially when considered from a Bayesian perspective, provide little empirical support for it. Next, I turn to the larger

John Ruscio, Department of Psychology, Elizabethtown College.

The author would like to thank Scott O. Lilienfeld and Thomas A. Widiger for their many thoughtful reactions to earlier drafts of this paper, and Antonee R. Stern for her research assistance.

Correspondence concerning this article should be addressed to John Ruscio, Department of Psychology, Elizabethtown College, One Alpha Drive, Elizabethtown, PA 17022. E-mail: rusciojp@etown.edu.

literature on labeling theory for more systematic tests of diagnostic labels' potentially stigmatizing effects. Despite the pervasive confounding of labels and the behavioral information they implicitly convey—which serves to inflate the apparent impact of “mere label” manipulations—this body of research nonetheless fails to substantiate claims of labeling effects. Thus, though there are a number of important and unresolved issues surrounding the diagnosis of mental illness (see Frances et al., 1991; Widiger & Clark, 2000), the evidence suggests that deviant behaviors are the primary cause of stigmatization, whereas the alleged stigmatizing effect of diagnoses and related labels is considerably less influential. In concluding sections, I discuss the necessity and value of diagnosis, offer reasons for the persistence of beliefs regarding labeling effects, suggest avenues for future labeling-related research, and consider more effective educational messages and means of relieving the stigma of mental illness.

PSEUDOPATIENTS AND PSEUDOSCIENCE

Perhaps the best place to begin is with Rosenhan's (1973a) well-known and provocative “On Being Sane in Insane Places.” Eight mentally healthy individuals—including Rosenhan himself—presented themselves to different mental hospitals, exhibiting anxiety and requesting admission based on a complaint of distressing auditory hallucinations. All of these “pseudopatients” were admitted to the hospitals. One individual was diagnosed with manic depression, whereas the other 11 were diagnosed with schizophrenia.¹ Once admitted, the pseudopatients stopped faking their symptoms. Aside from extensive note-taking for the purpose of observational data collection, the pseudopatients were instructed to act normally to determine whether the staff would discover their “sanity” and release them. Pseudopatients were discharged after an average stay of 19 days, each receiving the same diagnostic revision: The original condition was now classified as “in remission.”

The systematic observations made by the pseudopatients painted an unflattering picture of the quality of care provided in these mental hospitals. For example, staff members spent relatively little time with patients, and this was especially true of senior staff.

Likewise, a number of unethical and abusive staff behaviors were documented. Whereas ample data were gathered and reported to describe staff conduct and raise a number of important administrative concerns, Rosenhan (1973a) reached a number of logically and evidentially questionable conclusions on substantive issues concerning diagnostic reliability and labeling effects. For example, Rosenhan claimed that “psychiatric diagnoses . . . carry with them personal, legal, and social stigmas” (p. 252). He supposed that “the data speak to the massive role of labeling in psychiatric assessment. Having once been labeled schizophrenic, there is nothing the pseudopatient can do to overcome the tag. The tag profoundly colors others' perceptions of him and his behavior” (pp. 252–253). Because “the sane are not ‘sane’ all the time . . . the insane are not always insane,” Rosenhan reasoned by analogy that “it makes no sense to label ourselves permanently depressed on the basis of an occasional depression” (p. 254). Patients are cast in a hopeless light, as “the label sticks, a mask of inadequacy forever” (p. 257). The final sentence ties together sentiments expressed throughout the paper, implying that diagnostic labels are responsible for abusive practices: “In a more benign environment, one that was less attached to global diagnosis, [the staff's] behaviors and judgments might have been more benign and effective” (p. 257).

In a flurry of responses to these bold contentions (e.g., a series of letters published in the April 27, 1973, issue of *Science*; a 1975 special section of the *Journal of Abnormal Psychology*; and an elaborated critique by Spitzer, 1976), a number of researchers judged that Rosenhan had used seriously flawed methodology, ignored relevant data, and reached unsound conclusions. For example, Rosenhan seldom made appropriate comparative judgments (Dawes, 2001), drawing instead on a smattering of anecdotes selected from the wealth of observational data (e.g., concluding that bias alone led one staff member to perceive one patient as having a history of emotional ambivalence in close relationships); frequently called upon speculative theory, presumed consensus of opinion, or other relatively weak sources of evidence to support strong empirical claims (e.g., “the view has grown that psychological categorization of mental illness is useless at best and downright harmful, misleading, and pejorative at worst,” [Rosenhan, 1973a, p. 251], no citations to supportive scientific data provided); made a number of questionable inferences regarding others' perceptions with no independent corroboration (e.g., claiming that nurses' factual observation that “patient engaged in writing behavior” evi-

1. Rosenhan (1973a) does not describe how the 8 pseudopatients were distributed across the 12 hospitals. Either 1 pseudopatient successively entered 5 hospitals, 5 pseudopatients each entered 2 hospitals apiece, or some other configuration of repeated admissions took place.

denced their label-biased perception of psychopathology, though no pathology was referenced); and appealed to hypothetical counterfactuals that clearly constitute what Dawes (1994) calls "arguments from a vacuum" (e.g., in the case of emotional ambivalence noted above, stating that "an entirely different meaning would have been ascribed if it were known that the man was 'normal,'" [Rosenhan, 1973a, p. 253]).

Perhaps the greatest difficulty in accepting Rosenhan's conclusions stems from the pseudopatients' discharge diagnoses. Eleven pseudopatients were diagnosed with "schizophrenia, in remission" and one with "manic depression, in remission." Spitzer (1976) gathered data that suggest these classifications were used extremely rarely in psychiatric hospitals. The impressive agreement that Rosenhan reports across diagnosticians working in widely varying settings and evaluating a number of different pseudopatients contradicts the assertion that diagnoses are unreliable. Moreover, near-perfect agreement on such an unusual diagnosis proves just how attentive professionals were to these individuals' behaviors. Initial diagnoses of psychosis appear not to have significantly influenced perceptions, for in every case the staff correctly observed the absence of signs or symptoms of psychopathology at discharge. Thus, Rosenhan's own observations suggest that important clinical decisions were based more on pseudopatients' behaviors than their diagnoses. The shaky foundations of Rosenhan's case should give one pause in drawing upon it as support for allegations that diagnostic judgments are made unreliably or that labels are, on balance, more harmful than helpful.

PATIENTS, JOB APPLICANTS, AND PSYCHOLOGICAL DISTURBANCE

I now turn to what is perhaps the second most widely known psychological study of labeling: Langer and Abelson's "A Patient By Any Other Name . . ." (1974). In this experiment, behavioral and psychoanalytic clinicians watched a videotape of a job interview with the sound removed. Half the therapists of each orientation had been told that the interviewee was a *patient*; the other half, that he was a *job applicant*. After viewing the tape, participants responded to a series of open-ended questions about the interviewee that blind raters subsequently quantified along a 10-point scale of psychological adjustment. Psychoanalytic therapists' ratings were more negative for patients than for job applicants, whereas behavioral therapists' ratings were comparable across experimental conditions. The authors

conclude that psychoanalytic therapists were biased by a mere label whereas behavioral therapists were apparently immune to this biasing effect.

Davis (1979) challenged these conclusions from a Bayesian perspective. The label—either patient or job applicant—provided base rate information on the individual's psychological adjustment, and the muted videotape provided behavioral information with which a clinician could revise his or her initial expectations. By referring to "mere labels," Langer and Abelson (1974) appear to have presumed, without compelling support, that patients and job applicants experience equivalent levels of adjustment and therefore that the base rate information was wholly irrelevant. However, though extremely heterogeneous and largely overlapping populations, patients and job applicants probably do differ to some extent in their psychological adjustment. Thus, to disregard the label would be to ignore relevant base rate information, a common judgmental error (Kahneman & Tversky, 1973; Nisbett, Borgida, Crandall, & Reed, 1976; Nisbett & Ross, 1980) that would likely have reduced the validity of clinicians' final judgments.

For this reason, Davis (1979) argued that one can condemn the judgments of Langer and Abelson's (1974) psychoanalytic therapists—or exonerate those of their behavioral therapists—only if psychological adjustment is equivalent across patient and job applicant populations. Without more detailed information on the actual adjustment of patients and job applicants—which Langer and Abelson did not provide—there is no defensible criterion against which to evaluate the clinicians' judgments, and one cannot determine whether either group of therapists was, on the whole, unduly sensitive or insensitive to any real difference between groups. Likewise, it is impossible to interpret the observed difference in ratings across the two schools of psychotherapy. For example, it may be that psychoanalytic therapists made more accurate judgments by heeding relevant base rates of psychological disturbance and behavioral therapists made less accurate judgments by ignoring them. From the information provided in the original report, there is no way to tell whether this interpretation, Langer and Abelson's speculation, or any other view is warranted. In this way, the Bayesian perspective calls attention to the importance of considering the validity of a label for the judgment at hand. Just as invalid labels can be dangerous because they may be given undue consideration in reaching decisions, it is important to carefully and appropriately weigh valid base rate information; either too great or too small an emphasis can adversely affect judgmental accuracy.

To illustrate the fact that the behavioral information implicitly denoted by valid labels can play a crucial role in reaching sound judgments, consider the following hypothetical situation. Imagine that two new families, Smith and Jones, move into houses on either side of yours. You invite your new neighbors over for dinner and drinks, and everyone enjoys a splendid evening. Not only do you share many interests and values, but it turns out that each family has children close in age to your 11-year-old daughter. As you all spend more time getting to know one another in a variety of social settings, parents and children alike get along marvelously. You could not be more pleased. You then learn from a reliable source that Mr. Smith has been diagnosed with pedophilia. How would you react to this information? Would you feel equally comfortable encouraging your daughter to spend time playing in both neighbors' homes?

I suspect that few readers can honestly report that they would. This scenario is directly analogous to what was done not only in the Langer and Abelson (1974) experiment, but in most studies that ostensibly investigate labeling effects. A label—here, the label of pedophile—provides (imperfect) base rate information that is relevant and useful to judgments about your daughter's safety. Participants, however, are often castigated for taking similar evidence into account. To put it another way, labels communicate critically important behavioral information, yet participants are nonetheless expected to ignore them. Such an expectation is reminiscent of Rosenhan's contention that one should not diagnose individuals as depressed merely because they have occasional depressive episodes. By this logic, one should not label Mr. Smith a pedophile, because he may not abuse children all of the time. This fallacious conclusion stems from the unrealistic requirement that labels must summarize or predict behavior with perfect precision. In fact, even highly reliable, valid, and useful labels will only imperfectly describe the past and only probabilistically predict the future. For example, epilepsy is not diagnosed with perfect reliability or validity, nor does it cause people to experience seizures constantly (or, in many cases, even frequently), yet it is extremely important to document this condition when it exists. Diagnostic labels represent inferences, and as such they are fallible guides for judgments and decisions. This fallibility, however, is an irrational basis for the refusal to make inferences.

In labeling research, the fundamentally Bayesian issue of how to integrate multiple sources of information that vary in their validity (e.g., base rates denoted by labels and subsequently gathered behavioral data) is

almost entirely ignored. In the case of the Smith and Jones families, how much weight should be assigned to pleasant social interactions, and how much to the label of pedophile? Getting to know the two families in social settings is analogous to the vignettes participants read or the videotapes they watch in labeling studies. The utility of such information may be quite limited. Your informal observations in social settings are likely of poor validity in assessing potential threats to your daughter. How many people volunteer to others the fact that they are pedophiles? What behaviors might you observe, in a friendly get-together, that would suggest pedophilia? Considered in a similar way, a careful reading of many labeling studies casts doubt on the usefulness of the extralabel information that is presented to participants.

On the other hand, the label of pedophile denotes (albeit with imperfect precision) the potential for future sexual abuse of children. Just as Langer and Abelson's (1974) participants were not told how the patient came to earn that label, you were not told how Mr. Smith came to earn the label of pedophile; all that you know is that he has been diagnosed by a mental health professional. Despite the possibility that the patient was misidentified (i.e., that he is actually not a patient) or that Mr. Smith was misdiagnosed, it is reasonable to make tentative behavioral inferences in both cases. The patient may well be psychologically maladjusted to some extent, and Mr. Smith may well have sexually abused one or more children. Meehl's maxim is a reminder that the best predictor of future behavior is past behavior, which you recognize intuitively if you feel less comfortable with your daughter playing at the Smiths' home. Even though previous misdeeds are only inferred from the label—which itself constitutes neither direct behavioral observation nor a description of behavior—the possibility that Mr. Smith will repeat his past behavior is nonetheless alarming. The question of how strongly to weight this information is a more thorny issue. Dawes (1991) qualified Meehl's maxim to further note that although past behavior is the best predictor of future behavior, it typically isn't very good. Thus, the label of pedophile will only partially predict future behavior. Balancing concern for the welfare of one's daughter with the competing desire not to set in motion a potentially destructive self-fulfilling prophecy (i.e., treating neighbors with suspicion may lead them to act in ways that confirm your fears) is not simple. Although the amount of weight to be assigned to the label presents a dilemma, the fact that it provides relevant information is undeniable.

Precisely this problem pervades most all of the research on labeling. For example, when Fryer and

Cohen (1988) report that *psychiatric* patients were rated as more irresponsible, less dependable, and less clear-thinking than *medical* patients, one cannot determine whether these judgments are warranted. Psychiatric and medical patients may differ on some or all of these dimensions, in which case the Bayesian question arises once again: Were raters insensitive, appropriately sensitive, or overly sensitive to the label? Fryer and Cohen's experiment is typical of those in this literature: Most investigators appear to presume that a label does not denote any empirically relevant behavioral information (but see Wood & Valdez-Menchaca, 1996, who take note of this issue).

In sum, one can draw conclusions from Langer and Abelson's (1974) experiment, but it is also important to recognize an important qualification. The patient was evaluated more negatively than was the job applicant, a discrepancy evident only in the judgments of psychoanalytic clinicians, not those of behavioral clinicians. Unfortunately, in the absence of a valid criterion variable, the meaning of differences in judgments across labels and schools of thought are unclear. It is not possible to determine the influence that either label had on the validity of the assessments made by either type of clinician.

DISCOUNTING QUALIFIED EXPERTS TO IMPROVE JUDGMENTS?

A detailed look at one final study is in order by virtue of its frequent citation as support for alleged labeling effects. Temerlin (1968) showed psychiatrists, clinical psychologists, and clinical psychology graduate students a videotape in which an actor portrayed an ordinary, mentally healthy physical scientist and mathematician who had read a book about psychotherapy and wanted to discuss it with a psychologist. Before watching the tape, clinicians were informed by a prestigious diagnostician with many professional honors that the individual on the tape was "a very interesting man because he looked neurotic, but actually was quite psychotic." After viewing the tape, participants selected their best-guess diagnosis from a list of 30 choices: 10 psychotic disorders, 10 neurotic disorders, and 10 miscellaneous personality types, including "normal or healthy personality." A majority (60%) of the psychiatrists, along with 28% of the clinical psychologists and 11% of the graduate students, diagnosed the individual as psychotic. In contrast, none of the 78 participants in four control groups (e.g., no suggestion, suggestion of mental health) diagnosed this individual as psychotic.

Temerlin (1968) concluded that clinicians' diagnostic decisions are readily influenced by the suggestions of prestigious colleagues. Other researchers (e.g., Langer & Abelson, 1974) have cited this study as evidence of labeling effects, and Sushinsky and Wener (1975) proclaimed that labeling effects are highly generalizable on the basis of three experiments in which a prestigious expert offered precisely the same suggestion as in Temerlin's (1968) original study. Clinicians' diagnoses plainly were skewed in the direction of the expert's suggestion, but does this constitute cause for alarm?

Consider a scenario that presents you with choices analogous to those put to Temerlin's (1968) participants. Imagine that it is the end of the first day in a civil trial, and you are a member of the jury. The plaintiff, crossing the street while a "walk" light flashed, was struck by the defendant's car. The defendant, an uninsured motorist, had failed to stop for a red light. A prestigious neuropsychologist testifies, based on a careful examination that included an appropriate battery of cognitive tests, that although the middle-aged plaintiff may seem reasonably well-adjusted, he suffers from profound memory loss as a result of the head injury sustained in the accident. The neuropsychologist explains that the plaintiff's memory for the accident itself, along with subsequent events, is largely intact, yet he cannot recall many events prior to the accident and therefore qualifies for a diagnosis of retrograde amnesia. You keep a watchful eye on the plaintiff, who mostly observes the day's testimony, participating only to contribute a brief, untested, factual account of the accident. Throughout the day, he appears perfectly normal to you. If asked to select your best-guess diagnosis of the plaintiff from a list of 10 cognitive impairments, 10 psychotic disorders, and 10 personality disorders (including normal mental health), what would you say?

Understandably, you would want more information. How might a defense expert challenge the diagnosis of retrograde amnesia? How reliable was the plaintiff's memory prior to the accident? Might his memory have been poor or deteriorating independent of the alleged head injury? Is there any physical evidence of actual damage caused by the accident, and if so is it consistent with what is known about the etiology of retrograde amnesia? Is the plaintiff suing for punitive damages, in which case there is a motive for malingering? To maintain comparability with Temerlin's (1968) protocol, you are not allowed to ask any such questions. You simply have to make a diagnostic inference from this extremely limited supply of data.

Under the circumstances, it may be imprudent to second-guess a qualified expert's informed opinion on the basis of your own informal observations. An expert neuropsychologist's assessment of cognitive functioning provides fallible but diagnostically relevant information, whereas behavioral observations of a participant during one day of a court proceeding are of much more limited value. Bearing in mind the nature of the alleged damage, would individuals with and without retrograde amnesia observably differ during the first day of a court proceeding? Because the plaintiff's memory is supposedly impaired for events prior to those about which he testified, you could argue that there simply are no diagnostically relevant observational data available. In precisely the same way as in this hypothetical scenario, Temerlin's (1968) task presents the familiar problem of combining multiple pieces of information of differential validity. Participants were expected to pay due attention to a videotaped interview of questionable import and ignore the pertinent—though, unknown to them, false—identification of psychosis advanced by a reputable diagnostician.

Upon reflection, it should be clear that it is wise to factor the judgment of a well-qualified expert into one's decision making unless there are sufficient grounds to completely ignore it. To do so is not to automatically fall prey to the logical fallacy of an "argument from authority," as it does not mean that one ought to blindly defer to expert judgment without applying a healthy dose of skeptical scrutiny. From a Bayesian perspective, one ought to weigh the opinion of an expert according to his or her true level of expertise and the quality of the data from which his or her conclusions were drawn. For example, one should attempt to ascertain whether an alleged expert is in fact well qualified to render an opinion on the issue at hand, as well as whether there is sufficiently valid information for doing so. Judgments, decisions, and testimony are not vitiated by the mere trappings of psychological expertise, but should be supported by a track record of sound judgments in similar cases (for insightful treatments of expert qualifications, see Faust & Ziskin, 1988; Grove & Barden, 1999; and Meehl, 1997). Though Temerlin's (1968) data demonstrate a differential propensity to give weight to expert opinion across 3 groups of professionals with varying training and experience, this nonetheless illustrates only suggestibility: Many clinicians' diagnoses were influenced by some combination of experimental demand characteristics and potentially relevant expert opinion. As in Langer and Abelson's (1974) experiment, one would need a more defensible criterion

against which to evaluate participants' diagnostic judgments to argue that one group was more or less accurate than the others.

THE CONTINUING INFLUENCE OF THIS EARLY WORK

The three investigations examined above are frequently employed to bolster critiques of diagnostic labels. For example, many popular textbooks continue to draw on these studies. In a section on labeling in the chapter on psychological disorders, Myers's (1998) introductory psychology text presents Rosenhan (1973a) as a "demonstration of the biasing power of diagnostic labels. . . . That these normal people were misdiagnosed is not surprising . . . furthermore, before being released (an average of 19 days later), the 'patients' normal behaviors, such as taking notes, were often misinterpreted as symptoms" (p. 459). As noted earlier, there is no evidence that normal behaviors were mistaken for symptoms, and the uniform use of rare "in remission" discharge diagnoses suggests that normalcy was indeed recognized. A brief discussion of the Langer and Abelson (1974) experiment is then used to argue that "other studies confirm that labels affect how we perceive one another," with this discussion leading to the point that "labels can also stigmatize people in others' eyes" (p. 459). It is only at the end of the section, after further noting that "labels not only bias perceptions, they can also change reality . . . labels can serve as self-fulfilling prophecies" (pp. 459–460), that Spitzer (1975) is cited in a brief reminder that there can be benefits to diagnostic labels. The naive reader could not be faulted for forming the impression that diagnoses are largely responsible for the stigma of mental illness.

Nairne's (2000) introductory psychology text gives similar prominence to Rosenhan's (1973a) demonstration, contending that "its lessons about the hazards of labeling are clear" (p. 551). Beside the section on the Rosenhan study, Nairne presents a definition of diagnostic labeling effects and a figure showing the main effect across the patient and job applicant conditions in the Langer and Abelson (1974) experiment; a brief caption provides little methodological information. Davison and Neale's (2001) abnormal psychology text devotes much more attention to Langer and Abelson. Without mentioning the Bayesian critique articulated by Davis (1979), Davison and Neale ultimately conclude that this experiment shows the potentially biasing effects of one's theoretical orientation, with the clear implication that it was the psychoanalytic therapists whose judgments were ren-

dered less valid by such a bias. Comer's (1998) abnormal psychology text discusses labeling theory in an early chapter on etiology—"abnormal functioning is influenced greatly by the diagnostic labels given to troubled people and by the ways other people react to those labels . . . a famous and controversial study by the clinical investigator David Rosenhan (1973) supports this position" (p. 104)—and in a later chapter on schizophrenia—"we have already seen the very real dangers of diagnostic labeling. The famous Rosenhan (1973) study . . . is a particularly influential demonstration of these dangers" (p. 502). After briefly reviewing some findings and noting that this study is controversial, Comer concludes that "the investigation does demonstrate, however, that the label of 'schizophrenia' can have a negative effect on people—not just on how they are perceived, but on how they themselves feel and behave" (p. 503). The latter claim goes well beyond the data in that the alleged effects on the pseudopatients' feelings and behaviors were not even discussed by Rosenhan. Textbooks on clinical psychology, too, echo the enduring impact of Rosenhan's demonstration: "As Rosenhan (1973) clearly demonstrated in his famous study, we often see what we expect to see" (Cullari, 1998, p. 258); "Rosenhan's (1973) study of 'sane' graduate students admitted to mental hospitals demonstrated how others then construed all their behavior as evidence of abnormality" (Todd & Bohart, 1999, p. 73). This is but a sampling of the claims that are still repeated, or even sharpened and leveled into more dramatic forms (Gilovich, 1991).

Textbooks are by no means the only places in which the early labeling studies are presented frequently and favorably. A search of the Social Science Citation Index conducted on December 4, 2001, documents relevant trends in the scholarly research literature. Because the search was conducted within the most recent electronic database—1986 to the present, which begins more than a decade after each of the three original studies was published—it illustrates significant and lasting impacts on the field. Rosenhan (1973a) garnered 728 citations, whereas the detailed discussions and critiques of Rosenhan's demonstration published in the special section of the *Journal of Abnormal Psychology* have been cited far less often. The citation counts were as follows: Crown (1975), 16; Farber (1975), 15; Millon (1975), 20; Spitzer (1975), 1; Weiner (1975), 24. Spitzer's (1976) elaborated critique, published in the *Archives of General Psychiatry*, was cited 22 times. Whereas Langer and Abelson (1974) was cited 146 times, Davis's (1979) critique was cited just 9 times; both appeared in the *Journal of Consulting and Clinical Psychology*. I am not aware of any published critique of

Temerlin's (1968) experiment, which received 75 citations. In total, this search uncovered nearly 1,000 citations of the three early studies and just over 100 citations of published critiques. The influence of these studies persists in a relatively unchallenged manner.

AN EVALUATION OF LABELING THEORY

Fortunately, in addition to three oft-cited yet largely uninformative studies there is a substantial—and more recent—literature on labeling. Thomas Scheff (1966) published a landmark book in which he outlined a labeling theory of mental illness, which has served as the impetus for a vigorous debate and sparked many of the empirical studies of labeling. Whereas Scheff's (1966) sociological theory attributes violations of explicit rules to the actions of criminals or delinquents, it refers to psychopathology or aberrant behaviors that violate implicit rules as *primary deviance*. Labeling theory posits that if this primary deviance leads an individual to acquire a diagnostic label, society members' reactions to this label will produce *secondary deviance*, additional pathology or behavioral disturbance that causes or exacerbates mental illness.

Researchers have studied the putative effects of labeling by assessing the experiences of differently labeled groups of patients—or, more commonly, others' perceptions of them—for evidence of the adverse consequences of secondary deviance. The tacit belief has been that if experiences or evaluations differ across groups of differently labeled individuals, this supports the labeling theory. Beyond this basic framework, sociologists and others working in this vein have been vague in deriving predictions from labeling theory, though one can sometimes work through the theoretical constraints imposed by an empirical claim or a proposed reform. For example, is it not clear whether labeling theory predicts different effects for general labels, such as "mentally ill," as opposed to more specific labels, such as "schizophrenic." But those who have drawn upon labeling theory in their arguments to abandon diagnosis or revise diagnostic labels should, to be consistent, believe that it is specific labels that cause negative effects. Otherwise, their suggested reforms would have no effect, as changing labels will not alter the overarching label of "mentally ill," which one can infer with substantial validity from any evidence of treatment-seeking or referral to a mental health professional. Thus, although it is not always clear how strongly data tend to support or refute hypotheses drawn from labeling theory, often one can evaluate the internal consistency of beliefs and evidence.

In his review of the evidence, Scheff (1974) evaluated a total of 18 studies that he believed were explicitly related to labeling theory. He judged 13 studies to be consistent with the theory and 5 to be inconsistent, thus concluding that his theory was supported by the evidence. This "box score" approach to reviewing the literature has many drawbacks and limitations (Meehl, 1990), chief among them the fact that disconfirmations should ordinarily carry more epistemic weight than confirmations. Granting equal weight to confirmations and disconfirmations is especially problematic in light of the well-known tendencies for researchers not to submit for publication studies yielding null results (the file drawer problem; Rosenthal, 1979) and for editors to reject studies yielding results that conflict with popular theories (publication bias; Meehl, 1990). Perhaps even more important is that much of the research allegedly supportive of labeling theory is methodologically flawed in a variety of ways. Without providing a detailed critique, it should suffice to prove the point that Scheff (1974) singled out 2 studies as the strongest sources of support for his theory: Rosenhan (1973a) and Temerlin (1968). Given how weakly, if at all, these studies bear witness to labeling effects of any kind, let alone a causal role of labels in producing, intensifying, or stigmatizing mental illness, one can safely conclude that this body of studies provides equivocal support, at best, for labeling theory.

Critics of labeling theory, on the other hand, have argued that the role of secondary deviance is greatly overstated (Gove, 1970, 1982) and have assembled various lines of evidence that tend to refute the theory. For example, Gove (1982) notes that mental hospitals have a rigorous screening process to admit patients, most often on a voluntary basis, who need professional help: "The vast majority of persons labeled mentally ill are seriously impaired and their impairment is the major reason for labeling . . . labeling is not a major factor in a chronic career of mental illness but, in fact, labeling tends to initiate processes that minimize the length and severity of a person's disorder" (p. 291). Indeed, as argued below, unless a professional is prepared to offer precisely the same treatment to all patients, a classification scheme is *required* to connect patients with appropriate treatments. This, of course, is the primary pragmatic purpose of the diagnostic enterprise. Gove also reviews evidence on the temporal ordering of former patients' manifestations of symptoms and significant others' expectations. Whereas labeling theory contends that the expectations of others produce, or at least shape, symptomatic behavior, research suggests that the reverse is in fact true: Former patients' behaviors determined the expectations of family

members (Angrist, Lefton, Dinitz, & Pasamanick, 1968; Freeman & Simmons, 1963).

Whereas labeling theory predicts that individuals with sufficient resources to forgo hospitalization for mental illness should do so to avoid the secondary deviance caused by labeling, Gove and Howell (1974) found that just the opposite tends to occur. For example, when severity of disorder was controlled, married and upper-class individuals were more likely to receive treatment than unmarried or lower-class individuals. Even more telling is a study (Gove & Fain, 1973) in which extensive interviews with 429 former mental patients revealed improvement in their social relationships, positive evaluations of their hospital experiences, improved assessments of their situations, and increased capacity to deal with their problems. Of course, some of these positive comments may be the result of biases in retrospective reporting, such as the "effort after meaning" that can occur when individuals attempt to make sense of their previous experiences (Dawes, 1994) or the fact that the nature of memories recalled at a given time can be influenced by one's current emotional state (Lewinsohn & Rosenbaum, 1987). A small minority of the former patients (19) reported exclusively negative outcomes. However, there is no way to tell whether reactions were related to diagnostic labels. Many of these individuals may have had psychological disorders (e.g., paranoid schizophrenia) or prior experiences—in or out of therapeutic contexts—that predisposed them to hold unfavorable attitudes toward clinicians or the entire mental health enterprise. Only seven of these individuals mentioned the stigma of hospitalization as a problem, and it is unclear whether these negative reactions were prompted by diagnostic labels or other aspects of the hospitalization experience. Although Gove and Fain's correlational data do not suffice to establish any causal relationships, they are nonetheless difficult to reconcile with labeling theory.

To explain the discrepancy between patients' experiences and the handful of studies that allegedly show negative effects of labels, Gove and Fain (1973) note that disparaging individuals with mental illnesses in an abstract, impersonal way (e.g., reading a vignette and circling a response on a social rejection rating scale) is far different from actually perpetrating discrimination against a mentally ill person. As social psychologists have noted, attitudes do not always predict behaviors very well (Wicker, 1969). For example, in 1930 Richard LaPiere studied racial prejudice by traveling around the United States with a Chinese couple. Over the course of two years, they visited a total of 251 hotels, restaurants, and other business establishments, encountering racial discrimination just once. Six months after visiting,

LaPiere sent a questionnaire asking each proprietor whether Chinese individuals would be allowed as guests. Of the 128 responses he received, 118 said they would not, 9 gave conditional responses, and 1 said yes; prejudiced attitudes seldom translated into overt discrimination (LaPiere, 1934). Even with more reliable measurements of attitudes and aggregate measures of behavior, which provide more accurate estimates of attitude-behavior correspondence (Ajzen & Fishbein, 1977), attitudes still account for a relatively small percentage of behavioral variance. Thus, societal rejection of the mentally ill may not result in significant secondary deviance because few people act on their prejudice.

In his final analysis, Gove (1982) maintained that "a careful review of the evidence demonstrates that the labeling theory of mental illness is substantially invalid" (p. 295). More recently, Link, Cullen, Frank, and Wozniak (1987) reviewed empirical studies in which diagnostic labels and deviant behaviors varied independently of one another (e.g., they were manipulated orthogonally), thereby affording a comparison of their relative influence on stigma-related judgments. Studies were excluded if labels varied, but not behaviors, or if the study involved a subjective labeling of a behavior description. Among the 12 remaining studies, 10 failed to support labeling theory in that behaviors were shown to be more important determinants of social reactions than labels. As noted earlier, greater epistemic weight should be assigned to disconfirmations than to confirmations of a theory. Thus, despite the limitations of a simplistic box-score methodology, 10 out of 12 strikes *against* labeling theory is a particularly poor empirical showing. As an illustrative example, consider one of the largest, most thorough studies of its kind (Kirk, 1974). Three factors were systematically manipulated—labels (mentally ill, wicked, under stress), labelers (self, family, some people, psychiatrist), and behavior (normal, moderate, severe)—and 864 college students made ratings. Only behaviors influenced social rejection; neither labels nor labelers had a discernible effect. More recent studies (e.g., Boisvert & Faust, 1999; Cornez-Ruiz & Hendricks, 1993; Schwartz, Weiss, & Lennon, 2000; Wood & Valdez-Manchaca, 1996) offer no additional support for labeling theory.

In addition to the empirical results summarized above, there are further difficulties for labeling theory. First, research has demonstrated the cross-cultural generality of some important diagnostic constructs, which is inconsistent with the marked divergence that one would expect under labeling theory. For example, Murphy (1976) found that two cultures that have had minimal

contact with Western civilization (the Eskimo of the Bering Strait and the Yoruba of Nigeria) possess constructs such as schizophrenia, psychopathy, and alcoholism. In the absence of a common set of labels, it is difficult to reconcile these observations with any form of labeling theory maintaining that diagnostic labels are entirely crossculturally relative. A weaker form of labeling theory may be able to accommodate these data.

Second, on a related note, the stigma of mental illness predates formal classifications of psychopathology. Even the first edition of the *DSM* (American Psychiatric Association, 1952) is, historically speaking, a relatively recent human invention. Stigmatization of individuals suffering from mental illness has arguably been an acute social problem for centuries. Indeed, throughout recorded history, treating mental illness in increasingly humane and effective ways has probably done more to reduce the associated stigma than creating, revising, or eliminating diagnostic labels.

Third, clinician-patient confidentiality raises the question of just how a diagnosis itself could be responsible for stigmatization among family, friends, or coworkers. Whereas people may discover that an individual visited a mental health practitioner, his or her formal diagnosis is considerably more private information. Alleging that a diagnosis results in stigma presumes that other people come to know of it. In contrast, Phillips (1963, 1964) showed that societal rejection stems from nondiagnostic information such as the visibility of deviations from normal behavior and the source of help that individuals seek (e.g., visiting a clergyman or physician is less stigmatizing than visiting a psychiatrist or a mental hospital). Though individuals may reveal their diagnoses to others—intentionally or unintentionally—it is doubtful that this occurs with sufficient frequency to account for the observed prevalence of stigma. If labeling theory were formulated to allow for stigmatizing effects of general labels such as "mentally ill," which could be more readily inferred from many kinds of information, this would render moot the question of whether abandoning or revising specific diagnoses would have an influence on stigma.

Fourth, there is the *positive* influence that diagnostic labels may have in the causal attribution process. According to the discounting principle of causal inference (Kelley, 1973), knowledge of one plausible explanation for observed behavior will serve to diminish the perceived influence of other causes. In this way, a valid diagnostic label may prompt observers to discount other dispositional characteristics, such as personality traits, as causes of aberrant behaviors and therefore avoid reac-

tions of blame and social rejection. Consistent with this account, labels have had positive effects in many studies, particularly those with disabled children. Across many dimensions, adults rated mentally retarded children more favorably when they were identified by a diagnostic label than when they were not (Seitz & Geske, 1976). Whereas a label provided an explanation for the inappropriate behavior of a disabled child, nonlabeled children were blamed and punished for the same behaviors (Gibbons & Kassin, 1982). Peers rated essays written by children diagnosed with ADHD more positively than those written by nondiagnosed children (Cornez-Ruiz & Hendricks, 1993). Finally, Wood and Valdez-Menchaca (1996) found positive aspects to labeling children with expressive language disorder and suggested that a diagnostic label "may cause teachers to adopt a more supportive attitude toward the child . . . labeling can provide a more informative context in which to evaluate the relative strengths and weaknesses of a child with disabilities" (p. 587). Of course, caution needs to be exercised to avoid an unwanted side effect that might operate along similar lines: Discounting dispositional attributions should not be allowed to absolve individuals of personal responsibility for negative actions.

Finally, data contradict the original formulation of labeling theory in that secondary deviance may play a negligible role in mental illness. In their review, Link and Cullen (1990) affirm that primary deviance is the major determinant of diagnostic labels and note their many positive effects, particularly those that are treatment related. The evidence for secondary deviance's impact on mental health is much weaker; it may produce negative effects, but they appear to be small in magnitude, primarily social in nature, and contingent on additional factors. For example, Link et al. (1987) demonstrated that the perceived dangerousness of mental patients can moderate judgments of social rejection. The role of perceived dangerousness is consistent with the finding that people with no contact with the mentally ill perceived them as more dangerous and chose to maintain greater social distance than those who had such contact (Penn et al., 1994). Another potential moderator of stigmatization is the sex of target individuals and those who evaluate them. Phillips (1964) found that, for identical behaviors, men were rejected more than women. In a series of five studies by Farina and colleagues (Farina, Felner, & Boudreau, 1973; Farina & Hagelauer, 1975; Farina, Murry, & Groh, 1978) that controlled behaviors, no labeling effect was observed when women made ratings, whereas men tended to reject their fellow men more than they rejected women. Also, even though specific diagnostic labels do not appear to

carry significant costs, individuals who seek treatment may be initiating a self-fulfilling prophecy that can account for the social difficulties that are sometimes experienced. The patient role gives personal relevance to beliefs about the stigma of mental illness (Link, 1987), especially the fear of social rejection. One response to this fear is to withdraw from potentially rejecting social contacts (Link, Cullen, Struening, Shrout, & Dohrenwend, 1989), and the resulting impairment to social support networks may lead to a confirmation of the initially feared rejection. Other researchers have pointed to similar pathways. Claussen (1981) suggested that stigmatization is not a result of labeling, but of self-doubts that set in motion counterproductive processes. Farina, Gliha, Boudreau, Allen, and Sherman (1971) showed that the belief that others are aware of one's status as a mental patient can produce this type of self-fulfilling prophecy. Modifications to diagnostic practices are unlikely to prevent such patient-driven, label-independent self-fulfilling prophecies.

Particularly among methodologically rigorous investigations, there is no compelling evidence for the alleged stigmatizing effect of diagnostic labels. The clear empirical consensus was well summarized a quarter century ago: "It seems likely that any rejection directed towards psychiatric patients comes from their aberrant behavior rather than from the label that has been applied to them" (Lehmann, Joy, Kreisman, & Simmens, 1976, p. 332). Despite this showing, it is not at all uncommon to encounter the claim that diagnostic labels are stigmatizing. Indeed, the studies of Rosenhan (1973a), Langer and Abelson (1974), and Temerlin (1968) continue to enjoy an unusual popularity—having been cited, more than a decade after their original publication, nearly 1,000 times—whereas their critiques are virtually absent from published discussions. This popularity raises two related questions: Why do people persist in believing stigma to be caused by diagnostic labels, and what viable alternatives have been proposed?

BELIEF PERSEVERANCE

In discussing this subject with students I have discovered that many, including even those still in their first year of college, have been explicitly taught that diagnoses are stigmatizing. If the scientific support is underwhelming, why do some educators and professionals in psychology and related professions continue to hold this unwarranted belief? I would like to propose three reasons, though surely there are others.

The first reason is that diagnoses pose an easy target for those who are justifiably concerned about the stigma of mental illness. The positive effects of diagnoses (e.g., connecting patients with treatments, facilitating professional communication and research) are readily taken for granted. Some of the most visible attacks on diagnostic labels (e.g., Rosenhan, 1973a; Szasz, 1961) appear to presume that the efficacy and effectiveness of psychotherapy would survive intact without classifying patients in any way. As Carson (1996) observed, "It is much easier to be a critic in this area [challenging diagnostic conventions] than it is to suggest compelling and pragmatically realistic solutions" (p. 1137). Constructive criticism is worthwhile, but one should keep in mind both the positive and negative consequences of diagnostic labels. Indeed, conspicuously absent in most discussions of stigma, even those that render serious indictments of diagnoses, are proposals for viable alternatives to current diagnostic practices.

Though he did not develop them sufficiently to analyze their implications in detail, Rosenhan (1973a) hinted at two possibilities that others have endorsed: eliminating diagnoses altogether or restricting them to behavioral descriptions. In my experience, students who have been taught that diagnoses cause stigma almost uniformly support the former solution. However, the simple step of eliminating diagnoses is imprudent for several reasons. In addition to their descriptive utility, valid diagnoses carry valuable surplus meaning regarding the etiology, treatment, course, and/or outcome of psychological disorders (Kendell, 1975; Meehl, 1973a; Millon, 1991; Morey, 1991). To eliminate diagnoses would pose substantial problems without offering remedies. For example, it would greatly impede communication within and across treatment centers and research sites as well as making it impossible—under the present health care system—to coordinate third-party payments for therapeutic services. Moreover, eliminating diagnoses would significantly impair the ability to connect patients to empirically supported treatments (Chambless et al., 1996; Dobson & Craig, 1998). As clinicians develop and learn more about effective treatments, they will also need to learn how best to choose them based on client characteristics, and some system of diagnosis is required to do this well.

Restricting practitioners to purely behavioral descriptions² of their clients is also of dubious merit. Even supposing that stigmatization could be attributed to

diagnostic labels, it is unclear how behavioral descriptions would curtail negative reactions. For example, Rosenhan (1973b, p. 1647) supposed that the question "How might you feel if your colleagues believed you were a paranoid schizophrenic?" rhetorically demonstrated that the stigmatizing effects of labels are experientially obvious, that they cannot be denied. With Spitzer (1976), one might fairly question whether "the answer to his hypothetical question would be any different if put solely in behavioral terms without a diagnostic label—'how might you feel if your colleagues believed that you had an unshakable but utterly false conviction that everybody was out to harm you?'" (p. 465). Diagnostic labels and the behaviors that they denote are likely to prompt similar reactions, foiling a simple substitution of one for the other. In addition, mere behavioral description is, scientifically, a step backward. Only in the most primitive stages of basic clinical science does one simply identify problematic behaviors, for it soon becomes necessary to search for patterns of signs and symptoms that frequently co-occur and functional commonalities across individuals who exhibit superficially different behaviors. Thus, diagnosis at the level of clinical syndromes—though not the ultimate goal of a scientific classification scheme (Kendell, 1975; Millon, 1991)—aspires to an intermediate stage of far greater theoretical and practical utility than mere behavioral descriptions.

Whatever one's opinion of any particular system of reaching diagnoses, the classification of patients is unavoidable in clinical work. Kendell (1975) illustrates this by considering three aspects of human behavior: (1) ways in which all people are the same; (2) ways in which some people are the same, but different from others; and (3) ways in which people are totally unique. To the extent that people are all the same—in which case classification is unnecessary—one cannot rationally choose among available treatments or, for that matter, distinguish mental illness from mental health. In the extreme, this view would deny any traces of human uniqueness, thus providing an unacceptable foundation for clinical science and practice. To the extent that people are totally unique—in which case classification is impossible—all learning from personal experience and communication is for naught, all understanding and skills gained in the treatment of one patient will be useless in the treatment of the next. In the extreme, this view would render the scientific study of psychology impossible, for all thought and behavior would be utterly unpredictable. This, too, is clearly not the basis for clinical practice.

Therefore, there *must* be important characteristics that are shared by some patients, but not all. Although

2. Behavioral descriptions are not to be confused with behaviorists' functional analyses of behavior. For a discussion of the latter, see the special section of the *Journal of Consulting and Clinical Psychology* introduced by Follette (1996).

universal and unique patient characteristics are necessary elements in case formulation and treatment planning, they clearly must be complemented by a consideration of shared patient characteristics. Diagnostic systems are built upon shared signs and symptoms: "As soon as one begins to recognize features that are common to some patients but not to all, and to distinguish those which are important from those, like eye color, which are not, one is classifying them, whether one recognizes it or not. The only point at issue is what sort of classification one is going to have" (Kendell, 1975, p. 6).

Thus, clinical work inescapably involves classification. Though there are many valid criticisms of conventional diagnostic systems and practices (e.g., reaching a diagnosis may prompt confirmation bias, premature closure, or diagnostic overshadowing in eliciting and processing additional information), there is little evidence that diagnoses stigmatize patients and none suggesting that stigma reduction is attainable by changing to an alternative method of diagnosis. Until compelling evidence is produced, it seems wisest to strive to improve the reliability, validity, and utility of diagnostic procedures rather than attempting to evade their necessity or replacing them with behavioral descriptions.

A second reason why people may continue to hold diagnostic labels accountable for the stigma of mental illness is that studies such as Rosenhan (1973a), Langer and Abelson (1974), and to a lesser extent Temerlin (1968) are well known, cited regularly in the literature, and summarized in an easily accessible manner in a great number of sources. Critiques of these studies, on the other hand, are much less well known, cited only infrequently in the literature, and virtually never summarized. Thus, there is an availability bias (Tversky & Kahneman, 1973) such that one is far more likely to encounter readable descriptions—and implicitly or explicitly favorable evaluations—of the original studies than of their critiques. Moreover, although the critiques are well reasoned and clearly written, the original studies are nonetheless more vivid, conceptually easier to understand, and more emotionally involving, all of which serves to increase memorability and foster positive reactions (Nisbett & Ross, 1980).

A third potential reason for belief perseverance in the face of negative evidence is that each of us brings to the table prior beliefs stemming in part from professional affiliations and theoretical identifications. From an early point in the educational system, many students encounter strong positions on social issues such as the stigma of mental illness. Parents, peers, teachers, mentors, and colleagues can exert profound influences well before one

appreciates the need for dispassionate collection, analysis, and interpretation of scientific data. Thus, individuals may arrive at firmly entrenched beliefs about diagnostic labels before they ever seriously consider the scientific foundations of these beliefs. Few of us approach this (or any other) important topic free from bias.

SUGGESTIONS FOR FUTURE LABELING RESEARCH

Tentative conclusions on the scientific status of alleged labeling effects have been offered, but there are a number of ways in which future research on labeling could make valuable contributions to the literature. First, regardless of what methods are chosen, researchers should be more explicit in stating the theoretical formulation from which they are drawing predictions. At the outset, a reasoned argument should be provided to establish how differing results would support or refute the hypotheses under investigation.

Second, investigations should test for behavioral discrimination, rather than abstract, impersonal indications of negative attitudes. Although only a minority of studies that use questionnaire or interview methods have obtained negative labeling effects, it would be worthwhile to learn whether these findings translate into overt acts of discrimination. To address the question of whether laboratory-based findings generalize to real-world behaviors, it may be especially beneficial to conduct field research to improve external validity. For example, researchers could send trained individuals to apply for rental housing throughout a city, randomly determining (1) whether to exhibit normal or abnormal behaviors and (2) whether to note a diagnosed mental disorder in each instance. Success rates in obtaining the requested housing, not to mention the reasons given (if any) for granting or denying applications, would provide interesting and informative comparisons across the 4 experimental conditions.

Third, future research should also more carefully ensure that judgments are evaluated against an appropriate criterion measure. One can achieve this goal by either manipulating labels that are empirically unrelated to the dependent variable or developing a criterion measure that explicitly takes into account the validity of all sources of information. For example, Socall and Holtgraves (1992) went to considerable lengths to equate conditions on extralabel factors that might be related to their dependent measures (social distance and beliefs about predictability and outcome). They employed 3 pairs of conditions, 1 psychological and 1 medical within each pair, that plausi-

bly accounted for the same sets of symptoms described in their vignettes: generalized anxiety disorder (GAD) versus allergic food reaction, major depressive disorder (MDD) versus drug reaction to antihypertensive medication, chronic schizophrenia versus brain tumor. Moreover, to further reduce unwanted differences between stimulus conditions, all patients were described as successfully treated. Despite these efforts, meaningful differences between the psychological and medical conditions may still explain why participants' responses were more negative toward those with psychological disorders than those with medical conditions. Allergic food reactions, drug reactions, and operable brain tumors are little or no more likely to recur among those successfully treated than among those who have never experienced them, whereas the prognoses of patients treated for GAD, MDD, and chronic schizophrenia are considerably worse. Thus, particularly post-treatment, the implications of previously suffering from the psychological disorders are different than those of previously suffering from the medical conditions. Likewise, seemingly sound strategies such as counterbalancing labels by randomly targeting different individuals among those appearing in a videotape from participant to participant may create demand characteristics. Singling out 1 of 3 children with the label "developmentally delayed" and then asking related questions (e.g., Vogel & Karraker, 1991) may communicate the experimenter's expectations to participants.

An alternative approach to controlling all extralabel information is to construct appropriate criterion measures. One would need accurate data on the true difference between populations corresponding to one's experimental conditions to develop a defensible criterion against which to evaluate participants' judgments. For example, one could describe several target individuals who vary in the degree to which they suffer from sleep disturbances, reduced appetite, depressed mood, and so forth, such that the behavioral descriptions clearly communicate varying levels of depressive severity. The inclusion versus exclusion of an appropriate diagnostic label (e.g., dysthymia, major depressive disorder) could then be manipulated across conditions, with participants asked to rate target individuals on a number of variables for which valid norms³ exist in the

research literature (e.g., attendance or productivity at work, amount or quality of social interaction). Within each experimental condition, participants' ratings would be compared to the relevant normative data to determine whether judgments were reasonably accurate or biased in a positive or negative direction. Only after comparisons with the normative criterion had been factored into the analysis (e.g., by computing difference scores as judgment minus criterion) could the most meaningful comparisons be carried out across experimental conditions. Thus, it would be possible to find that individuals' expectations are accurate or biased (in either direction) for various behavioral descriptions, each in the presence or absence of a label. One could also incorporate subject variables (e.g., educational background, theoretical orientation, previous contact with mentally ill individuals) into the design to test additional hypotheses. I am not aware of any labeling studies that have taken this approach by evaluating participants' judgments against carefully selected criteria.

Fourth, contextual factors should be better studied to understand the role that diagnostic labels play in reaching judgments about individuals with mental illnesses. For example, in a recent experiment (Ruscio, 2002), behaviors clearly indicative of a particular mental disorder were held constant while the type of doctor (physician vs. psychiatrist), diagnosis (present vs. absent), and treatment (none, medication, psychotherapy) varied factorially across between-subjects experimental conditions. Four hundred eight undergraduate students were each asked to pretend to be a manager at a marketing firm considering the work of 2 employees with good track records within the organization who have been having problems at work. The first scenario read as follows:

Recently, Tom's work performance has suffered as a result of a few brief episodes with the sudden onset of intense apprehension, fearfulness, or terror that are often associated with feelings of impending doom. During these attacks, he experiences shortness of breath, palpitations, chest pain or discomfort, choking or smothering sensations, and fear of "going crazy" or losing control. You refer Tom to the consulting [physician/psychiatrist] who performs evaluations for your company. This [physician/psychiatrist] performs a thorough assessment, [prefers not to render a diagnosis/reaches a diagnosis of panic attack], and recommends that [no treatment is necessary/an anti-anxiety medication be administered to rectify a biochemical imbalance/cognitive psychotherapy be administered to teach how to properly interpret bodily signs].

3. Norms would be judged relevant to the extent that the population of individuals that was sampled includes the target individual—as described behaviorally—as a highly representative, if hypothetical, case. Ideally, the behavioral descriptions would be based heavily on the available norms to ensure a defensible matching. Likewise, the dependent measures should assess variables in units of measurement suitable for comparison with the normative criteria.

Participants then made ratings of how well they would expect Tom to complete an important project and how comfortable they would feel working closely with Tom; both ratings were made for a 2-week and a 6-month period. A second scenario was similar in all essential regards except that Bob, the second employee, suffered from a clear-cut case of paranoid schizophrenia. The order of scenarios was counterbalanced.

Statistical analysis of participants' ratings revealed 3 main effects and 3 interactions. Ratings were more favorable for the 6-month period than for the 2-week period, for Tom than for Bob, and when some treatment was recommended (either medication or psychotherapy) than when no treatment was recommended. Performance expectations were slightly higher for Tom than for Bob, whereas comfort ratings were much higher for Tom than for Bob. For the 2-week period, ratings were less favorable when the doctor recommended no treatment than when the doctor recommended either medication or psychotherapy, whereas for the 6-month period the same pattern was even more pronounced. Interestingly, when the doctor was a physician, ratings were more favorable when a diagnosis was made than when no diagnosis was made, whereas when the doctor was a psychiatrist, ratings differed little between when a diagnosis was made and when no diagnosis was made. Bearing in mind that in all conditions Tom obviously suffered from panic attacks and Bob obviously suffered from paranoid schizophrenia, it is intriguing that the influence of a diagnosis depended on the type of doctor who reached it. Experiments such as this shed light on the role that diagnostic labels play within the larger context of information potentially relevant to stigmatization. Additional research of this type may elaborate and clarify the multifactorial influences on the judgments of laypersons and professionals.

Fifth and finally, future research should test for positive effects of diagnostic labels and the mechanisms by which they occur. A handful of experiments suggest, in keeping with the discounting principle of attribution theory, that a valid diagnostic label can prevent blaming or rejecting responses to others' behavior. There may be additional means by which labels impart benefits. Receiving a diagnosis can psychologically validate an individual's subjective distress. Indeed, those who suffer from an unrecognized condition may propose a new diagnostic category with the hope of gaining public recognition, sympathy, and support. Knowing that others endure similar suffering can provide some measure of comfort, as can the possibility that the condition is well understood and potentially treatable. Gerald Senf

(quoted in Corning, 1986) captures the psychological desire to be diagnosed: "It would be unsettling to walk into a doctor's office with an ailment, receive an examination, and then have the doctor say to you, 'I've never seen anything like this before'" (p. 287). Future investigations should test for positive labeling effects on perceptions of oneself and others as well as evaluating the mechanisms by which positive effects are produced (e.g., clear and compassionate communication) and their generality across diagnostic categories, patient populations, and other potentially relevant factors. Because valid information, such as feedback from psychological tests, can itself be therapeutic (Finn & Tonsager, 1992), research exploring how to capitalize on the beneficial aspects of labels is urgently needed so that diagnoses can be utilized to the best possible advantage.

CONCLUSIONS

Until scientific data demonstrate a stigmatizing impact of diagnostic labels—relative to a viable alternative system—it would be unwise to maintain belief in such a negative effect. In addition to ignoring their many beneficial uses and effects, there are a number of important reasons not to disparage diagnostic labels on the unsubstantiated grounds that they are stigmatizing.

Unfounded criticisms of diagnosis can set in motion a counterproductive self-fulfilling prophecy described by Meehl (1973b). Instructors, mentors, or supervisors who do not value diagnostic judgments—regarding them as intrinsically unreliable, invalid, or of little or no utility—will teach future clinicians that diagnoses are not important and provide insufficient corrective feedback on diagnostic knowledge and skills. Trainees, internalizing the message that diagnosis is unimportant and receiving little help or encouragement to learn, will devote inadequate time and effort to developing diagnostic competence. Upon becoming clinicians, their diagnoses will be of poor quality, spuriously confirming their elders' initial beliefs. Research on labeling does not justify setting the stage for lackadaisical training experiences.

Rather than simply perpetuating belief in labeling theory, instructors and clinical supervisors can take advantage of several opportunities when discussing the stigma of mental illness. Presenting both sides of the issue is both intellectually honest and a vehicle for facilitating critical thinking and independent judgment. For example, having students or trainees read the Rosenhan-Spitzer debate can spark their interest and promote a

lively, constructive discussion. Moreover, to truly comprehend the literature on labeling requires a conceptual grasp of Bayesian principles, information integration, and probabilistic reasoning. In their psychological training, students should be exposed to research on clinical judgment and alternative methods for combining data of differential validity, along with the imperfect predictive validity that can be expected (Faust, 1986; Lilienfeld, Wood, & Garb, 2000; Swets, Dawes, & Monahan, 2000). Discussing issues of stigma and labeling adds concreteness and interest to subject matter that can be comparably dull and difficult in the abstract.

Perhaps most important of all, focusing on other means of alleviating the stigma of mental illness may serve this purpose better than flogging diagnostic labels. Presumably, most discussions of stigmatization occur because people want to do something about the problem. If labels are not the real villain, then researchers, instructors, and practitioners should turn their attention elsewhere. For example, Penn et al. (1994) showed that providing community members with information about the post-treatment living conditions of schizophrenics (i.e., supervised care), reduced negative judgments. Likewise, as research on labeling children with disabilities suggests, the thoughtful and appropriate use of diagnostic labels may help to reduce stigmatization. Noting that labeling theory is vague in its specification of the mechanisms by which labels influence people, Gill and Maynard (1995) argue that clinicians and clients are not naive, that labels are not necessarily the blunt instruments they are often depicted to be. By studying interactions between professionals and the parents of disabled children, Gill and Maynard showed that diagnostic information can be elicited, and eventual diagnoses conveyed, in compassionate and meaningful ways. Given the therapeutic importance of empathy and related features of doctor-patient relationships (Orlinsky & Howard, 1986), perhaps better attention to the psychotherapist's analogue of a physician's "bedside manner" would prevent the self-fulfilling prophecies that diagnoses can trigger.

More broadly, Corrigan and Penn (1999) discuss several strategies for discrediting stigma: protest, education, and the promotion of contact. Drawing upon social psychological research on stereotyping, they cautioned that efforts to suppress misinformation through protest can backfire and educational efforts may be limited by the resilience of prior beliefs. Contact can be enhanced by many factors, such as equal status, cooperative interaction, and institutional support. In light of the evidence that diagnoses play a minimal role in the stigma of men-

tal illness, psychologists should focus on formulating, debating, and evaluating alternative strategies for combating stigma; recognizing the necessity and value of diagnostic labels; improving the reliability and validity of diagnostic judgments; and compassionately eliciting relevant information, communicating diagnoses, and formulating humane, effective treatment plans.

REFERENCES

- Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*, 888-918.
- American Psychiatric Association (1952). *Diagnostic and statistical manual of mental disorders*. Washington, DC: Author.
- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Angrist, S., Lefton, M., Dinitz, S., & Pasamanick, B. (1968). *Women after treatment: A study of former mental patients and their neighbors*. New York: Appleton-Century-Crofts.
- Boisvert, C. M., & Faust, D. (1999). Effects of the label "schizophrenia" on causal attributions of violence. *Schizophrenia Bulletin*, *25*, 479-491.
- Carson, R. C. (1996). Aristotle, Galileo, and the DSM taxonomy: The case of schizophrenia. *Journal of Consulting and Clinical Psychology*, *64*, 1133-1139.
- Chambless, D. L., Sanderson, W. C., Shoham, V., Johnson, S. B., Pope, K. S., Crits-Christoph, P., et al. (1996). An update on empirically validated therapies. *Clinical Psychologist*, *49*, 5-18.
- Claussen, J. A. (1981). Stigma and mental disorder: Phenomena and terminology. *Psychiatry*, *44*, 287-296.
- Comer, R. J. (1998). *Abnormal psychology* (3rd ed.). New York: W. H. Freeman.
- Cornez-Ruiz, S., & Hendricks, B. (1993). Effects of labeling and ADHD behaviors on peer and teacher judgments. *Journal of Educational Research*, *86*, 349-355.
- Corning, W. C. (1986). Bootstrapping toward a classification system. In T. Millon & G. L. Klerman (Eds.), *Contemporary directions in psychopathology: Toward the DSM-IV* (pp. 279-303). New York: Guilford.
- Corrigan, P. W., & Penn, D. L. (1999). Lessons from social psychology on discrediting psychiatric stigma. *American Psychologist*, *54*, 765-776.
- Crown, S. (1975). "On being sane in insane places": A comment from England. *Journal of Abnormal Psychology*, *84*, 453-455.
- Cullari, S. (1998). Treatment resistance. In S. Cullari (Ed.), *Foundations of clinical psychology* (pp. 249-271). Boston: Allyn & Bacon.

- Davis, D. A. (1979). What's in a name? A Bayesian rethinking of attributional biases in clinical judgment. *Journal of Consulting and Clinical Psychology, 47*, 1109-1114.
- Davison, G., & Neale, J. M. (2001). *Abnormal psychology* (8th ed.). New York: Wiley.
- Dawes, R. M. (1991). Probabilistic versus causal thinking. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology* (Vol. 1, pp. 235-264). Minneapolis: University of Minnesota Press.
- Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York: Free Press.
- Dawes, R. M. (2001). *Everyday irrationality: How pseudo-scientists, lunatics, and the rest of us systematically fail to think rationally*. Boulder, CO: Westview Press.
- Dobson, K. S., & Craig, K. D. (Eds.). (1998). *Empirically supported therapies: Best practice in professional psychology*. Thousand Oaks, CA: Sage.
- Farber, I. E. (1975). Sane and insane: Constructions and misconstructions. *Journal of Abnormal Psychology, 84*, 589-620.
- Farina, A., Felner, R. D., & Boudreau, L. A. (1973). Reactions of workers to male and female mental patient job applicants. *Journal of Consulting and Clinical Psychology, 41*, 363-372.
- Farina, A., Gliha, D., Boudreau, L. A., Allen, J. G., & Sherman, M. (1971). Mental illness and the impact of believing others know about it. *Journal of Abnormal Psychology, 77*, 1-5.
- Farina, A., & Hagelauer, H. D. (1975). Sex and mental illness: The generosity of females. *Journal of Consulting and Clinical Psychology, 43*, 122.
- Farina, A., Murray, P. J., & Groh, T. (1978). Sex and worker acceptance of a former mental patient. *Journal of Consulting and Clinical Psychology, 46*, 887-891.
- Faust, D. (1986). Research on human judgment and its application to clinical practice. *Professional Psychology: Research and Practice, 17*, 420-430.
- Faust, D., & Ziskin, J. (1988). The expert witness in psychology and psychiatry. *Science, 241*, 31-35.
- Finn, S. E., & Tonsager, M. E. (1992). Therapeutic effects of providing MMPI-2 test feedback to college students awaiting therapy. *Psychological Assessment, 4*, 278-287.
- 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.
- Fracchia, J., Canale, D., Cambria, E., Ruest, E., & Sheppard, C. (1976). Public views of ex-mental patients: A note on perceived dangerousness and unpredictability. *Psychological Reports, 38*, 495-498.
- Frances, A., First, M. B., Widiger, T. A., Miele, G. M., Tilly, S. M., Davis, W. W., et al. (1991). An A to Z guide to DSM-IV conundrums. *Journal of Abnormal Psychology, 100*, 407-412.
- Freeman, H. E., & Simmons, O. G. (1963). *The mental patient comes home*. New York: Wiley.
- Fryer, J. H., & Cohen, L. (1988). Effects of labeling patients "psychiatric" or "medical": Favorability of traits ascribed by hospital staff. *Psychological Reports, 62*, 779-793.
- Gibbons, F. X., & Kassin, S. M. (1982). Behavioral expectations of retarded and nonretarded children. *Journal of Applied Developmental Psychology, 3*, 85-104.
- Gill, V. T., & Maynard, D. W. (1995). On "labeling" in actual interaction: Delivering and receiving diagnoses of developmental disabilities. *Social Problems, 42*, 11-37.
- Gilovich, T. (1991). *How we know what isn't so*. New York: Free Press.
- Goodyear, R. K., & Parish, T. S. (1978). Perceived attributes of the terms *client*, *patient*, and *typical person*. *Journal of Counseling Psychology, 25*, 356-358.
- Gove, W. R. (1970). Societal reaction as an explanation of mental illness: An evaluation. *American Sociological Review, 35*, 873-884.
- Gove, W. R. (1982). The current status of the labeling theory of mental illness. In W. R. Gove (Ed.), *Deviance and mental illness* (pp. 273-300). Beverly Hills, CA: Sage.
- Gove, W. R., & Fain, T. (1973). The stigma of mental hospitalization: An attempt to evaluate its consequences. *Archives of General Psychiatry, 28*, 494-500.
- Gove, W. R., & Howell, P. (1974). Individual resources and mental hospitalization: A comparison and evaluation of the societal reaction and psychiatric perspectives. *American Sociological Review, 39*, 86-100.
- Grove, W. M., & Barden, R. C. (1999). Protecting the integrity of the legal system: The admissibility of testimony from mental health experts under Daubert/Kumho analyses. *Psychology, Public Policy, and Law, 5*, 224-242.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237-251.
- Kelley, H. H. (1973). The process of causal attribution. *American Psychologist, 28*, 107-128.
- Kendell, R. E. (1975). *The role of diagnosis in psychiatry*. Oxford, England: Blackwell Scientific.
- Kirk, S. A. (1974). The impact of labeling on rejection of the mentally ill: An experimental study. *Journal of Health and Social Behavior, 15*, 108-117.
- Langer, E. J., & Abelson, R. P. (1974). A patient by any other name . . . : Clinician group difference in labeling bias. *Journal of Consulting and Clinical Psychology, 42*, 4-9.
- LaPiere, R. T. (1934). Attitudes vs. actions. *Social Forces, 13*, 230-237.
- Lehmann, S., Joy, V., Kreisman, D., & Simmens, S. (1976). Responses to viewing symptomatic behaviors and labeling of prior mental illness. *Journal of Community Psychology, 4*, 327-334.
- Lewinsohn, P. M., & Rosenbaum, M. (1987). Recall of parental behavior by acute depressives, remitted depressives, and nondepressives. *Journal of Personality and Social Psychology, 52*, 611-619.

- Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest*, 1, 27–66.
- Link, B. G. (1987). Understanding labeling effects in the area of mental disorders: An assessment of the effects of expectations of rejection. *American Sociological Review*, 52, 96–112.
- Link, B. G., & Cullen, F. T. (1990). The labeling theory of mental disorder: A review of the evidence. *Research in Community and Mental Health*, 6, 75–105.
- Link, B. G., Cullen, F. T., Frank, J., & Wozniak, J. F. (1987). The social rejection of former mental patients: Understanding why labels matter. *American Journal of Sociology*, 92, 1461–1500.
- Link, B. G., Cullen, F. T., Struening, E., Shrout, P. E., & Dohrenwend, B. P. (1989). A modified labeling theory approach to mental disorders: An empirical assessment. *American Sociological Review*, 54, 400–423.
- Meehl, P. E. (1973a). *Psychodiagnosis: Selected papers*. Minneapolis: University of Minnesota Press.
- Meehl, P. E. (1973b). Why I do not attend case conferences. In P. E. Meehl (Ed.), *Psychodiagnosis: Selected papers* (pp. 225–302). Minneapolis: University of Minnesota Press.
- Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, 66, 195–244.
- Meehl, P. E. (1997). Credentialed persons, credentialed knowledge. *Clinical Psychology: Science and Practice*, 4, 91–98.
- Millon, T. (1975). Reflections on Rosenhan's "On being sane in insane places." *Journal of Abnormal Psychology*, 84, 456–461.
- Millon, T. (1991). Classification in psychopathology: Rationale, alternatives, standards. *Journal of Abnormal Psychology*, 100, 245–261.
- Morey, L. C. (1991). Classification of mental disorder as a collection of hypothetical constructs. *Journal of Abnormal Psychology*, 100, 289–293.
- Murphy, J. (1976). Psychiatric labeling in cross-cultural perspective. *Science*, 191, 1019–1028.
- Myers, D. G. (1998). *Psychology* (5th ed.). New York: Worth.
- Nairne, J. S. (2000). *Psychology: The adaptive mind* (2nd ed.). Belmont, CA: Wadsworth.
- Nisbett, R. E., Borgida, E., Crandall, R., & Reed, H. (1976). Popular induction: Information is not necessarily informative. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Orlinsky, D. E., & Howard, K. I. (1986). Process and outcome in psychotherapy. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (3rd ed., pp. 311–381). New York: Wiley.
- Penn, D. L., Guynan, K., Daily, T., Spaulding, W. D., Garbin, C. P., & Sullivan, M. (1994). Dispelling the stigma of schizophrenia: What sort of information is best? *Schizophrenia Bulletin*, 20, 567–577.
- Phillips, D. L. (1963). Rejection: A possible consequence of seeking help for mental disorders. *American Sociological Review*, 28, 963–972.
- Phillips, D. L. (1964). Rejection of the mentally ill: The influence of behavior and sex. *American Sociological Review*, 29, 679–687.
- Phillips, D. L. (1966). Public identification and acceptance of the mentally ill. *American Journal of Public Health*, 56, 755–763.
- Rosenhan, D. L. (1973a). On being sane in insane places. *Science*, 179, 250–258.
- Rosenhan, D. L. (1973b). Sane: Insane. *Journal of the American Medical Association*, 224, 1646–1647.
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, 86, 638–641.
- Ruscio, J. (2002). *Contextual factors in laypersons' stigmatization of individuals suffering from mental illness*. Manuscript submitted for publication.
- Scheff, T. J. (1966). *Being mentally ill: A sociological theory*. Chicago: Aldine.
- Scheff, T. J. (1974). The labeling theory of mental illness. *American Sociological Review*, 39, 444–452.
- Schwartz, S., Weiss, L., & Lennon, M. C. (2000). Labeling effects of a controversial psychiatric diagnosis: A vignette experiment of luteal phase dysphoric disorder. *Women and Health*, 30, 63–75.
- Seitz, S., & Geske, D. (1976). Mothers' and graduate trainees' judgments of children: Some effects of labeling. *American Journal of Mental Deficiency*, 81, 362–370.
- Skinner, L. J., Berry, K. K., Griffith, S. E., & Byers, B. (1995). Generalizability and specificity associated with the mental illness label: A reconsideration twenty-five years later. *Journal of Community Psychology*, 23, 3–17.
- Socall, D. W., & Holtgraves, T. (1992). Attitudes toward the mentally ill: The effects of label and beliefs. *Sociological Quarterly*, 33, 435–445.
- Spitzer, R. L. (1975). On pseudoscience in science, logic in remission, and psychiatric diagnosis: A critique of Rosenhan's "On being sane in insane places." *Journal of Abnormal Psychology*, 84, 442–452.
- Spitzer, R. L. (1976). More on pseudoscience in science and the case for psychiatric diagnosis. *Archives of General Psychiatry*, 33, 459–470.
- Sushinsky, L. W., & Wener, R. (1975). Distorting judgments of mental health: Generality of the labeling bias effect. *Journal of Nervous and Mental Health*, 161, 82–89.
- Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1, 1–26.
- Szasz, T. S. (1961). *The myth of mental illness: Foundations of a theory of personal conduct*. New York: Hoeber.

- Temerlin, M. K. (1968). Suggestion effects in psychiatric diagnosis. *Journal of Nervous and Mental Disease, 147*, 349-353.
- Todd, J., & Bohart, A. C. (1999). *Foundations of clinical and counseling psychology* (3rd ed.). New York: Longman.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 4*, 207-232.
- Vogel, D. A., & Karraker, K. H. (1991). Effects of developmentally-delayed labeling on adults' perceptions of toddlers. *Child Study Journal, 21*, 251-261.
- Weiner, B. (1975). "On being sane in insane places:" A process (attributional) analysis and critique. *Journal of Abnormal Psychology, 84*, 433-441.
- Wicker, A. W. (1969). Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues, 25*, 41-78.
- Widiger, T. A., & Clark, L. A. (2000). Toward DSM-V and the classification of psychopathology. *Psychological Bulletin, 126*, 946-963.
- Wood, M., & Valdez-Menchaca, M. C. (1996). The effect of a diagnostic label of language delay on adults' perceptions of preschool children. *Journal of Learning Disabilities, 29*, 582-588.