

# CLINICAL JUDGMENT AND DECISION MAKING\*

Howard N. Garb

*Wilford Hall Medical Center, Lackland Air Force Base, San Antonio,  
Texas 78236-5300; email: howard.garb@lackland.af.mil*

**Key Words** personality assessment, psychological assessment, psychodiagnosis, behavioral prediction, statistical prediction

■ **Abstract** When clinical psychologists make judgments, are they likely to be correct or incorrect? The following topics are reviewed: (a) methodological advances in evaluating the validity of descriptions of personality and psychopathology, (b) recent findings on the cognitive processes of clinicians, and (c) the validity of judgments and utility of decisions made by mental health professionals. Results from research on clinical judgment and decision making and their relationship to conflicts within the field of clinical psychology are discussed.

## CONTENTS

CLINICAL JUDGMENT AND DECISION MAKING .....	67
METHODOLOGICAL ADVANCES .....	68
COGNITIVE PROCESSES .....	69
Cognitive Heuristics and Biases .....	70
Diagnosis and Psychology of Categorization .....	70
VALIDITY OF JUDGMENTS AND UTILITY OF DECISIONS .....	71
Description of Personality and Psychopathology .....	72
Diagnosis .....	74
Case Formulation .....	76
Behavioral Prediction .....	78
Treatment Decisions .....	80
THE ROMANTIC AND EMPIRICIST TRADITIONS IN CLINICAL PSYCHOLOGY .....	82

## CLINICAL JUDGMENT AND DECISION MAKING

Upon receiving the Bruno Klopfer Distinguished Contribution Award from the Society for Personality Assessment, Caldwell (2004) gave an example of his success interpreting the Minnesota Multiphasic Personality Inventory (Hathaway & McKinley 1943):

\*The U.S. Government has the right to retain a nonexclusive, royalty-free license in and to any copyright covering this paper.

We got a severe 4-6-8 profile on a young woman. I looked at the tortured implications of the pattern and somehow said, "She will have something like cigarette burn scars on her hands, where her father prepared her to steel herself to the suffering of life." The round burn marks were on her hands and extended a little way up her arms (Caldwell 2004, p. 9).

This is a remarkable interpretation. However, it is anecdotal in nature. Caldwell did not present scientific data to support his style of test interpretation. Instead, he selectively related his experiences.

Clinical psychologists are likely to have one of two reactions to the claims made by Caldwell, depending on whether they are influenced by romantic or empiricist ideas (Wood et al. 2003, pp. 92–94). Romantics are likely to be greatly impressed and inspired by Caldwell's accomplishments. Those from the empiricist tradition are likely to have a more skeptical attitude. Empiricists insist that grand claims be scientifically tested. Their point of view is exemplified by a new journal, *The Scientific Review of Mental Health Practice*, which is devoted to the objective investigation of controversial and unorthodox claims in clinical psychology, psychiatry, and social work. The distinction between romantics and empiricists seems especially apt, given the title of Caldwell's address: "My Love Affair with an Instrument."

In this article, research on clinical judgment and its relationship to the romantic and empiricist traditions in clinical psychology is examined. First, the following topics are covered: (a) methodological advances in studying the validity of descriptions of personality and psychopathology, (b) the cognitive processes of clinicians, and (c) the validity of clinical judgments and the utility of treatment decisions.

## METHODOLOGICAL ADVANCES

Important research has been conducted on clinical judgment and decision making, but it has focused on issues that are relatively easy to address. For example, some studies have described interrater reliability. Other studies have been conducted to determine if judgments based on a small amount of information (e.g., results from a single test) agree with results from a large set of information (e.g., judgments based on an interview and history information).

Some important questions seldom have been addressed. For example, when a clinical psychologist describes a client's personality, it is difficult to know how to determine if the psychologist is correct. The description can be compared to other fallible indicators (e.g., self-report ratings, peer ratings), but it is also difficult to determine the validity of these indicators. Similarly, when assessing a child, a clinician can systematically collect information from the child, parents, teachers, and peers. However, what if the information from these different sources is contradictory, as is often the case?

Two new approaches to evaluating the validity of descriptions of personality and psychopathology can be described. One approach is to have clients make ratings every day. For example, in a superb study, Wu & Clark (2003) constructed a behavior record form to measure daily activities. Participants were instructed to indicate whether they had performed a particular behavior on a given day (e.g.,

got into an argument, lost their temper). This particular measure was developed to provide measurements of aggression, impulsivity, and exhibitionism. The advantage of having participants make self-ratings every day is that they do not have to rely on memory. Also, by making ratings for specific behaviors, they do not have to make judgments that require more than a low degree of inference. If we were interested in learning whether psychologists in clinical practice can report whether their clients are aggressive, impulsive, or exhibitionistic, then it would be of value to compare their judgments to results from this behavior record.

A second approach involves improvements in research design and statistical analysis. Kraemer et al. (2003) have proposed an elegant solution for what has long been a recalcitrant problem. Psychologists routinely collect information from more than one source. For example, children can make self-ratings and can be described by parents, teachers, and peers. When these descriptions are not in agreement, it may be unclear which description is valid.

According to Kraemer et al. (2003), an informant's report reflects the influences of (a) the actual characteristics that we want to measure (e.g., traits, symptoms, competencies), (b) the context in which the subject is observed, (c) the perspective of the informant, and (d) the error of measurement. Context refers to the setting/place and circumstance in which the rater knows the person being rated (e.g., home versus outside of home). Perspective refers to characteristics of the informant that influence his or her assessment of the trait (e.g., someone who is making self-ratings will have a different perspective than will other people). Given this framework, our goal is to obtain a measure of a trait that is relatively free from variance attributable to context (or setting) and perspective (biases of the informants). To do this, one must gather ratings from multiple informants who know the client in different contexts and who have differing perspectives. For example, ratings in the context of school could be made by a teacher and a child. Ratings in the context of home could be made by a parent and the child. Perspective would vary as a function of self (child) and other (teacher, parent). All of the trait ratings could then be analyzed using a principal components analysis. If one is successful, the principal components analysis will yield separate factors for trait(s), context, and perspective. If this occurs, then one will have obtained a measure of trait(s) that is closer to the gold standard than the individual ratings made by the informants because one will have removed variance due to context and perspective. This would represent a significant advancement in being able to evaluate the validity of a clinician's ratings of traits and symptoms.

## COGNITIVE PROCESSES

Describing the cognitive processes of clinical psychologists could prove beneficial for improving training and clinical practice. It would be helpful if we could compare the cognitive processes of trainees with those of expert clinicians. This would allow supervisors to give empirically guided feedback to the students and trainees, e.g., on what information to attend to and what questions to ask. Remarkably, this rarely has been done. In fact, little is known about individual differences among clinicians.

Research on cognitive processes is also important because computers and artificial intelligence are likely to play increasingly large roles in our lives (Garb 2000). The more we know about the cognitive processes of clinicians, the more intelligently we will be able to harness the power of artificial intelligence for making judgments and decisions. Although we are far from reaching our goals, advances continue to be made. Recent work on heuristics and biases are mentioned below, followed by a description of work on the psychology of categorization and psychodiagnosis.

## Cognitive Heuristics and Biases

Cognitive heuristics and biases, as formulated by Tversky & Kahneman (1974), have been frequently used to describe how clinical psychologists and other people make judgments (e.g., Arkes 1981, Dawes 1986, Garb 1998, Kayne & Alloy 1988, Turk & Salovey 1988, Wedding & Faust 1989). Thus, progress made in studying heuristics and biases is likely to inform research on clinical judgment. With this in mind, it is important to note that in the opinion of Kahneman (2003, p. 703), the formulation of the affect heuristic is “probably the most important development in the study of judgment heuristics in the past few decades.”

The affect heuristic was introduced and described by Slovic et al. (2002):

*Affective responses* occur rapidly and automatically—note how quickly you sense the feelings associated with the stimulus words *treasure* or *hate*. We argue that reliance on such feelings can be characterized as the *affect heuristic* (p. 397).

One can think of many instances when affect is likely to guide clinical judgment and decision making. On the one hand, the affect heuristic may have positive effects and be related to clinical intuition and the setting of meaningful treatment goals. However, it may also yield negative effects and be related to biases (e.g., race bias and gender bias). It is not yet clear if the research methods that have been employed in studying the affect heuristic in everyday judgment making can be usefully employed in the study of clinical judgment.

Research on the affect heuristic may have important implications for a long-standing issue in clinical psychology: the issue of whether computers or clinical judges should make judgments and decisions (Meehl 1954). To the extent that clinical psychologists can be positively guided by their feelings, they can be expected to do well compared to computers. However, in comparisons of these two methods, clinicians have rarely been more accurate than computers (Grove et al. 2000), which suggests that the affect of the clinician may have limited value for making accurate judgments and decisions. Research is needed to directly address this issue.

## Diagnosis and Psychology of Categorization

Research methods used by cognitive psychologists to study categorization have been usefully employed to study clinical judgment. Results suggest that

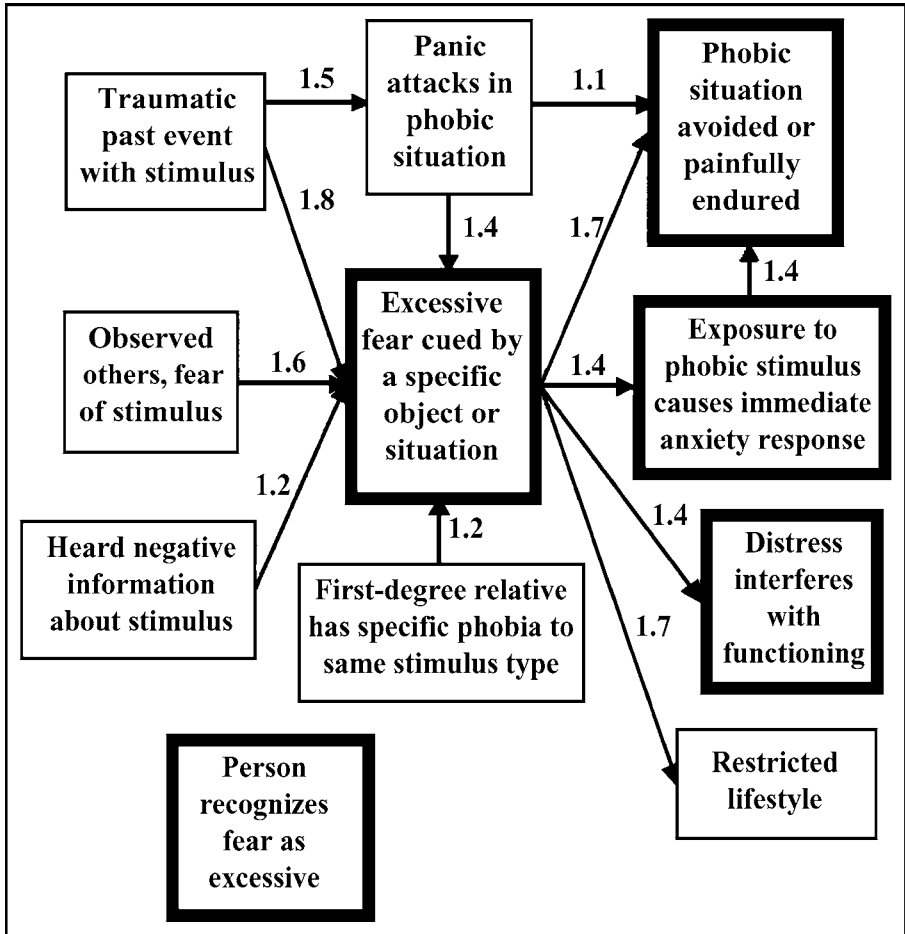
psychologists frequently make diagnoses by comparing clients to prototypes (e.g., Cantor et al. 1980; also see Blashfield et al. 1985, 1989; Evans et al. 2002; Garb 1996). A prototype is a clinician's conception of a hypothetical client who best exemplifies a particular disorder. The results are important because they suggest that interrater reliability will be low when psychologists do not share the same prototypes. The results also suggest that agreement between clinicians' diagnoses and diagnoses based on the *Diagnostic and Statistical Manual of Mental Disorders* (DSM; American Psychiatric Association 1994) criteria will be affected by the dissimilarity of the clinicians' prototypes to the criteria.

Recent results indicate that mental health professionals also make diagnoses by forming causal theories (Kim & Ahn 2002; also see Wakefield et al. 1999). In five outstanding but complex experiments, Kim & Ahn (2002) had psychologists and psychology graduate students make ratings for a series of tasks. Some of the ratings took from 2.5 to 6 hours to complete. For one of the tasks, the DSM criteria and the associated symptoms listed in the DSM manual were presented to judges. Participants were to draw arrows indicating cause-and-effect relations between any of the symptoms that they felt were causally connected. They were also instructed to describe the strength of the postulated causal relations. Psychologists and psychology graduate students provided detailed causal explanations. Most of these causal theories were quite complex. A composite of the drawings made by psychologists and psychology graduate students for a specific phobia is presented in Figure 1. Notice that in this example, the DSM criterion "Person recognizes fear as excessive" was not perceived as being part of a causal relation with other symptoms. Kim & Ahn then gave the judges additional tasks. They were able to determine that the criteria and symptoms that had been described as forming part of a causal relation were weighed more heavily than were other criteria when they made diagnoses. For example, for the diagnosis of specific phobia, the criterion "Person recognizes fear as excessive" was given a mean diagnostic importance rating of only 75.5; ratings for the other DSM criteria ranged from 87.3 to 96.3. This is of interest because when using the DSM, clinicians are supposed to weigh each criterion equally.

Other results reported by Kim & Ahn (2002) are also of interest. For one of the tasks, psychologists and psychology graduate students were instructed to read a set of case studies. When they were later asked about the clients, they recalled causally central symptoms more often than causally peripheral symptoms and isolated symptoms. In fact, the results revealed a bias to falsely recognize symptoms that were causally central to clinicians' theories of different disorders.

## VALIDITY OF JUDGMENTS AND UTILITY OF DECISIONS

When clinicians describe clients, are they likely to be correct or incorrect? Below, results are described for (a) the description of personality and psychopathology, (b) diagnosis, (c) case formulation, (d) prediction, and (e) decision making.



**Figure 1** Composite of participants' drawings of their causal theories of specific phobia. The American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition* (1994) diagnostic criteria are shown in bold. From NS Kim & W Ahn (2002). Copyright American Psychological Association. Reprinted with permission.

## Description of Personality and Psychopathology

Practicing psychologists seem to be less sophisticated at describing personality traits than at making diagnoses. At least for diagnoses, there are explicit rules (DSM criteria), and we can hope to have an idea of what types of judgments clinicians are making. Although we can be certain that clinical psychologists attend to symptoms and consider whether a client has a personality disorder when they evaluate and treat a client, it is frequently not clear if they spend much time thinking about the client's personality traits. For example, research suggests that the evaluation of a

client's normal-range personality traits (e.g., persistence, conscientiousness) can be helpful for treatment planning (Garb 2003), but it is not clear that clinicians consider these traits when thinking about a client. This is a complicated issue because cognitive processing frequently occurs outside of one's awareness (e.g., Kihlstrom 1999). Perhaps a clinician will consciously think about a trait if the client falls toward the end of the dimension. Similarly, a large body of research supports the Big Five Model of personality (Costa & Widiger 2002), but results from this area of research do not guide most psychologists. Instead, psychologists may be guided by implicit models of personality that owe more to everyday experience than to scientific findings. In other words, this area of research is so primitive, we are not even sure what types of personality impressions clinicians form.

In reviewing the literature, it is clear that questions about the validity of descriptions of personality and psychopathology are usually sidestepped or answered indirectly. Remarkably, in these studies, clinicians virtually never make their judgments based on all of the information that is usually available in clinical practice. This is not true of studies on diagnosis, prediction of behavior, and treatment decision making. Important research has been conducted on the description of personality and psychopathology, but it has focused on answering questions that are easily answered.

A popular approach to studying clinical judgment in this area is to examine interrater reliability: One can see if different clinicians make similar ratings when evaluating the same set of clients. Overall, interrater reliability has varied widely for describing personality traits, but has often been good for describing psychiatric symptoms (Garb 1998, pp. 10–14).

Recent work on interrater reliability has focused on the description of defense mechanisms. Overall, results from two studies suggest that psychologists should not be describing defense mechanisms unless they receive extensive training. In both studies, clinicians made ratings using the Defense Functioning Scale, a scale that was introduced for further study in the DSM-IV (American Psychiatric Association 1994). Perry et al. (1998, p. 56) concluded, "Our findings indicate that the defense axis is a feasible, acceptably reliable, and nonredundant addition to DSM-IV," but their results can also be interpreted as indicating that there is little support for adding a defense mechanism axis to the next edition of the DSM. For example, they reported that median kappa reliabilities were 0.42 for describing individual defenses. A kappa value less than 0.40 typically indicates that interrater reliability is poor. A median kappa value of 0.42 suggests that almost half of the kappa reliabilities were below 0.40, so interrater reliability was presumably poor in many cases. In the other study (Hilsenroth et al. 2003), positive results were obtained: The mean intraclass correlation coefficient for ratings of six defense levels was 0.59. However, the study was not conducted under typical clinical conditions: ratings were made after a 2-hour semistructured clinical interview and a 1- to 1.5-hour interpretive/feedback interview.

Studies on validity have also been conducted. Typically, one indicator of a construct has been related to another indicator of the construct (e.g., ratings based

on test results have been compared to ratings based on an interview). This is an appropriate, although limited, strategy. The results illustrate difficulties in making clinical judgments. For example, research generally suggests that it is difficult to draw valid inferences from nonverbal behavior (Ambady & Rosenthal 1992, Garb 1998). Also, psychologists frequently are *not* more accurate than graduate students (Garb 1989, 1998). Difficulty in learning from clinical experience is a topic that is discussed at the end of this article.

Recent work has evaluated the ability of clinicians to detect lying. In most studies, people have not been accurate at this task (DePaulo 1994). However, by coding videotapes using the Facial Action Coding System (Ekman & Friesen 1978), one can detect some individuals who are lying (Frank & Ekman 1997). Furthermore, judges have been able to detect lying after viewing videotapes of interviews under the following condition: when the experimenters select videotapes that differ in the expression of emotion for subjects who lied and subjects who told the truth (Ekman & O'Sullivan 1991, Ekman et al. 1999). Thus, by carefully selecting videotapes to show clinicians, one can increase the likelihood that the clinicians' ratings will be valid. This result is of theoretical interest: When individuals show facial muscular movement that indicates they are lying, clinicians' ratings will have modest validity. However, the result is of limited clinical significance for two reasons. First, individuals who lie do not always show facial muscular movement that is associated with lying. Second, considering that the chance level of accuracy was 50%, a majority of the psychologists obtained only a modest level of accuracy. For example, for "regular" clinical psychologists, 4% achieved 0%–30% levels of accuracy, 52% achieved 40%–60% accuracy, and 44% achieved 70%–100% accuracy.

## Diagnosis

Acceptable levels of interrater reliability have been reported for psychodiagnoses in field trials for both DSM-III (American Psychiatric Association 1980, pp. 470–471) and for the tenth revision of the *International Classification of Diseases* (Sartorius et al. 1993, 1995). However, it is likely that these results were obtained because the clinicians participating in the field trials were familiar with and adhered to diagnostic criteria. In clinical practice, many mental health professionals frequently do not adhere to diagnostic criteria (e.g., Blashfield & Herkov 1996, Davis et al. 1993, Ford & Widiger 1989, Morey & Ochoa 1989); therefore, one might expect that diagnoses made in routine clinical practice will be dissimilar to diagnoses based on semistructured interviews. Semistructured interviews are used to ensure that diagnoses are based on specific criteria and rules. The scientific literature reveals that agreement is generally poor between diagnoses made in routine clinical practice and diagnoses based on semistructured interviews (Shear et al. 2000, Strakowski et al. 1997; also see Garb 1998, pp. 53–54).

Several factors indicate that semistructured interviews are more valid than diagnoses made in clinical practice. First, clinicians do not always ask about important symptoms. In one study (Miller et al. 2001), clinicians evaluated only about 50%

of the key criteria that were assessed using semistructured interviews. Of course, some clinicians will evaluate all of the relevant criteria, but the research suggests that many do not. Second, it can be noted that interrater reliability is better for semistructured interview diagnoses than for clinical diagnoses. Finally, agreement with diagnoses made by expert clinicians has been better for diagnoses based on semistructured interviews than for diagnoses made in routine clinical practice (Basco et al. 2000, Miller et al. 2001). In response to these findings, Widiger & Samuel (2004) recommended that clinicians administer a self-report inventory to alert themselves to the potential presence of maladaptive personality traits followed by a semistructured interview to verify their presence.

At the same time, we must remain aware of the limitations of semistructured interviews. Semistructured interviews clearly do not represent a gold standard. Improvements in methodology need to be applied to further evaluate their validity, e.g., by collecting longitudinal data (Spitzer 1983). In fact, it is possible that in some circumstances, diagnoses based on therapy sessions will be more accurate than diagnoses based on semistructured interviews, because therapists have the opportunity to observe and interact with clients over time. With other clients, it is possible that self-report inventories will be more valid than semistructured interviews (Widiger & Samuel 2004). Perhaps the key to using semistructured interviews is to recognize that they should not be done in rote fashion and that interviewers should be careful in making inferences from the assessment data. In particular, clients should feel that there is a therapeutic element to the interviews and that they can elaborate on their answers.

Interestingly, Westen (1997) has argued against using semistructured interviews for diagnosing personality disorders. He concluded that no personality-disorder semistructured interview or questionnaire has "demonstrated acceptable evidence that it validly assessed the constructs it purported to assess" (p. 895), but he supported his conclusion by citing a review by Perry (1992). The review by Perry (1992) covered results from only nine studies. A more recent review (Widiger 2002) covered results from 35 studies on the diagnosis of personality disorders. Widiger (2002, p. 463) found that "convergent validity generally improves as the degree of structure increases." For example, the worst convergent validity coefficient was obtained in the only study to have used unstructured interviews by practicing clinicians.

Clinicians make fewer diagnoses than do research investigators using semistructured interviews, presumably because they inquire about fewer symptoms (Basco et al. 2000, Zimmerman & Mattia 1999). For example, in one study (Basco et al. 2000), the semistructured interview method identified 96 current comorbid mental disorders in a sample of 200 patients. Clinicians using routine diagnostic methods identified only 35 current comorbid diagnoses. Not only do clinicians make fewer diagnoses than research investigators, research indicates that nearly all comorbid mental disorders are underdiagnosed in routine clinical practice (Garb 1998). For example, the mental disorders of mentally retarded clients are frequently underdiagnosed (Jopp & Keys 2001). Similarly, evidence indicates that mental health

professionals underdiagnose substance abuse disorders among psychiatric inpatients and mental disorders among substance abuse patients (e.g., Hansen et al. 2000, Kirchner et al. 1998). To describe the underdiagnosis of comorbid disorders, the term “diagnostic overshadowing” is frequently used (Reiss et al. 1982).

An advantage of using a semistructured interview is that the questions one asks will not vary as a function of patient characteristics such as race. In fact, to learn if diagnoses made by clinicians are biased, clinicians’ diagnoses are frequently compared with diagnoses made by research investigators. If differences between the two sets of ratings are related to race, gender, or other variables, then one can conclude that one set of diagnoses is biased. Bias has been observed for diagnosis. For example, for diagnoses made in clinical practice and diagnoses based on semistructured interviews, agreement is higher for white patients than for nonwhite patients (Strakowski et al. 1997), and schizophrenia is frequently overdiagnosed in African American patients (e.g., Whaley 2001).

Controversies exist over whether some mental disorders are overdiagnosed. Historically, schizophrenia was overdiagnosed in the United States, at least until the advent of DSM-III in 1980 when the criteria for schizophrenia were made more stringent (Cooper et al. 1972). More recently, questions have been raised about the possible overdiagnosis of dissociative identity disorder (formerly multiple personality disorder; Lilienfeld et al. 1999) and attention deficit/hyperactivity disorder (ADHD; LeFever et al. 2003). Surprisingly, post-traumatic stress disorder (PTSD), after a period of wide acceptance, has become marked by controversy. In fact, in March 2004, Congress held a hearing on the possible overdiagnosis of this disorder. The controversy was instigated in part by the finding that although only 15% of the men who served in Vietnam were assigned to combat units, the National Vietnam Veterans Readjustment Study yielded a lifetime PTSD prevalence rate of 30.9% (Burkett & Whitley 1998, McNally 2003a). In addition, after examining military records via the Freedom of Information Act, Burkett & Whitley (1998) estimated that about 75% of veterans receiving PTSD compensation are pretenders. According to Burkett & Whitley (1998), the high prevalence rate of PTSD reported for the National Vietnam Veterans Readjustment Study may have occurred because military records were not checked to verify statements made by veterans. Congress has recently become involved because they want to know how many soldiers are likely to return from Iraq with PTSD. It is also of interest to note that delayed onset PTSD has not been observed to occur in longitudinal studies of individuals suffering from exposure to noncombat trauma.

## Case Formulation

In addition to describing personality and psychopathology and making diagnoses, psychologists also try to understand the causes of their clients’ behaviors and symptoms. As already mentioned, even when they make other types of judgments (e.g., diagnoses), they frequently consider causal factors (Kim & Ahn 2002). Thus, case formulation is a key task for clinical psychologists—one that is likely to affect how they perform on other tasks.

Unfortunately, making causal judgments can be extremely difficult, and is perhaps more difficult than other tasks facing mental health professionals. It is simply more difficult to explain things than to describe them. In fact, for my book on clinical judgment (Garb 1998), I did not locate a single study on causal judgments in which validity was good or excellent for individual clinicians, regardless of whether the specific task was related to behavioral, cognitive behavioral, or psychodynamic assessment.<sup>1</sup> In reviewing research on psychological assessment, I concluded:

Relatively little research has been conducted to evaluate the reliability and validity of case formulations. The research that has been conducted suggests that the validity of case formulations is often poor (Garb 1998, p. 100).

Recent studies have obtained similar findings. In particular, interrater reliability continues to be poor. For example, psychologists and psychology graduate students in one study (Daleiden et al. 1999) made judgments about the reinforcement functions of children who refuse to attend school. Case descriptions of school-refusing children were obtained from an anxiety disorder clinic. Clinicians were to make ratings for different types of negative reinforcers that could be maintaining the children's behavior. Judgments made by individual clinicians were of questionable interrater reliability (e.g., values of kappa ranged from 0.12 to 0.17). Reliability was acceptable when judgments were aggregated across clinicians (values of kappa were as large as 0.70). However, in clinical practice, case formulations are typically made by a single clinician, not by averaging ratings across clinicians. Similarly, in a second study (Persons & Bertagnolli 1999), mental health professionals made cognitive behavioral formulations after listening to recordings of initial interviews with three depressed women. Interrater reliability for schema ratings (ratings of underlying cognitions) was poor for individual clinicians (mean intraclass correlation coefficient, or ICC, equal to 0.37), but good when schema ratings were averaged over five judges (ICC = 0.72).

Psychologists and other mental health professionals, and even individuals involved in public mental health policy, frequently make causal judgments that are not supported by empirical research. Instead, judgments frequently are based on informal observations or clinical lore. A recent example was given by Baumeister et al. (2003, p. 1):

Teachers, parents, therapists, and others have focused efforts on boosting self-esteem, on the assumption that high self-esteem will cause many positive outcomes and benefits. . . . Our findings do not support continued widespread efforts to boost self-esteem in the hope that it will by itself foster improved outcomes.

At one time, the state of California had funded a task force on self-esteem because of the belief that "raising self-esteem would help solve many of the state's

---

<sup>1</sup>However, interrater reliability was sometimes fair when conditions did not resemble clinical practice, e.g., when ratings were averaged across clinicians.

problems, including crime, teen pregnancy, drug abuse, school underachievement, and pollution” (Baumeister et al. 2003, p. 3). Research has not supported this belief.

Case formulations based on informal observations and clinical lore are often based on the reports of clients and their families. However, research indicates that memories are biased in systematic ways (for numerous examples, see McNally 2003b). In general, clients tend to remember their past behavior and mental status in ways that conform with their own implicit theories of psychopathology. For example, in one study (Rueter et al. 2000), adolescents completed forms indicating whether they had engaged in behaviors related to conduct disorder during the previous year. They completed the forms annually, starting in some instances at age 12 and in other instances at age 14, and ending at age 15. Retrospective ratings were made at age 19, at which time they were asked what criterion behaviors they had engaged in prior to age 15. Participants who met antisocial criteria at age 19 overreported having had behavior problems related to conduct disorder, whereas participants who experienced a decline in antisocial behavior by age 19 underreported having had behavior problems related to conduct disorder in the past. Thus, memories were distorted to conform to present behavior.

To increase the likelihood of making valid causal judgments, psychologists should be more systematic in collecting history data. A well-replicated finding is that many clinicians do not regularly ask clients about important events, e.g., if they have ever been sexually or physically abused (Garb 1998, pp. 88–89). Similarly, Widiger & Clark (2000) implied that clinicians do not rigorously inquire about the course of a client’s psychopathology. They recommended that DSM-V require that clinicians systematically learn about the “life-span history of a patient’s symptomatology . . . by recording, for example, age of onset, lifetime history of disorders, and their longitudinal course” (p. 956). Finally, to decrease reliance on retrospective reports of specific behaviors, clients can carry small handheld electronic diaries that will allow for describing those behaviors in real-time and in real-world contexts (Shiffman et al. 2002).

In summary, psychologists should be cautious when explaining behavior. Even case formulations that seem reasonable may be incorrect. Psychologists should also become more systematic and rigorous in collecting history data.

## Behavioral Prediction

Behavioral prediction is one of the most exciting areas of clinical judgment. Although progress has been slow for improving the validity of predictions for some tasks (e.g., prediction of course of mental disorder, prediction of suicide), exciting advances have been made for the prediction of violence and the prediction of recidivism of sexual offenses. For these latter two tasks, new assessment instruments are being adopted, and statistical prediction rules are having their greatest impact.

**PREDICTION OF VIOLENCE** Although early studies on the long-term prediction of violence indicated that clinicians are wrong twice as often as they are correct

(Monahan 1981), results from subsequent studies indicate that predictions of violence made by clinicians are often valid (Mossman 1994). For example, in a well-known study on the long-term (six-month) prediction of violence (Lidz et al. 1993), clinicians made correct predictions for 58% of the patients who did not become violent and 60% of those who did.

Although mental health professionals are able to predict violence with moderate validity, race bias and gender bias have been documented. Black psychiatric inpatients and black prison inmates often are predicted to be more violent than are white psychiatric inpatients and white prison inmates, even when race is not significantly related to the occurrence of violence (Garb 1998, pp. 113–114). This finding continues to be replicated. In a recent study (Hoptman et al. 1999), psychiatrists at a forensic psychiatric hospital were to predict which patients would become assaultive during a three-month period. Fifty-seven percent of the patients were African American. They committed 55% of the assaults. The psychiatrists had predicted that they would commit 67% of the assaults.

With regard to gender bias, violence tends to be overpredicted for male patients and underpredicted for females (Garb 1998, p. 115). A recent study (Elbogen et al. 2001) suggests that this is related to the gender of the mental health professionals. In this study, male and female clinicians agreed upon general levels of dangerousness for female patients, but female clinicians viewed male psychiatric patients as more dangerous than did male clinicians.

Advances have occurred for the prediction of violence. The importance of evaluating psychopathy when predicting violence has been established (e.g., Monahan et al. 2001). Forensic psychologists, but not other licensed psychologists, are generally careful to evaluate this construct when predicting violence (Tolman & Mullenore 2003). In addition, a number of assessment instruments and aids have gained acceptance among forensic psychologists, including measures of psychopathy (e.g., the Hare Psychopathy Checklist-Revised, PCL-R; Hare 1991), statistical prediction rules (e.g., the Violence Risk Appraisal Guide, VRAG; Quinsey et al. 1998), and rules and guidelines designed to structure the process for evaluating risk for violence (e.g., the HCR-20, a 20-item checklist to assess the risk for future violent behavior in criminal and psychiatric populations; Webster et al. 1998).

These advances have had a large impact on forensic psychology but apparently not on clinical and counseling psychology. In a recent survey (Tolman & Mullenore 2003), forensic psychologists, but not other licensed psychologists, reported being aware of the scientific literature on predicting violence. Both groups of psychologists also described the assessment instruments that they use. The PCL-R and the VRAG were among the top five instruments used by forensic psychologists to evaluate psychopathy. This was not true for the other licensed psychologists. Given the controversy that has surrounded the use of the Rorschach inkblot test (Hunsley & Bailey 1999, Lilienfeld et al. 2000), it is of interest to note that the Rorschach was among the top five instruments used by nonforensic psychologists for conducting general evaluations and for evaluating psychopathy. For forensic psychologists, the Rorschach was not among the top five tests for either task. Similarly, in another

survey (Lally 2003), a majority of diplomates in forensic psychology deemed the Rorschach to be unacceptable for a range of tasks including the evaluation of (a) risk for violence, (b) risk for sexual violence, (c) competency to stand trial, (d) competency to waive Miranda Rights, and (e) malingering.

**PREDICTION OF SEXUAL CRIME** Advances also have occurred for the prediction of sexual recidivism. Recidivism rates for criminals committing sex offenses range from about 35% to 55% (Hanson et al. 2003). In recent years, research on the actuarial prediction of sexual recidivism has increased dramatically: Only one finding was identified in a review of the literature published before 1996 (Hanson & Bussiere 1998), whereas more than 50 findings have been subsequently described (Hanson et al. 2003). Actuarial measures that have been used to predict sexual recidivism include the Violence Risk Appraisal Guide (VRAG; Harris et al. 1993), the Sex Offender Risk Appraisal Guide (SORAG; Quinsey et al. 1998), the Rapid Risk Assessment for Sex Offender Recidivism (RRASOR; Hanson 1997), and the Static-99 (Hanson & Thornton 1999). Positive results have been obtained (Hanson et al. 2003). When results from different studies have been pooled, the accuracy of actuarial predictions has been statistically significant, medium in effect size ( $d = 0.68$ , 95% confidence interval of 0.62 to 0.42), and significantly more accurate than unstructured clinical predictions ( $d = 0.28$ , 95% confidence interval of 0.14 to 0.42). Thus, compared to clinical prediction, the use of actuarial scales is a major improvement for predicting sexual recidivism.

## Treatment Decisions

Treatment decisions often are based on interviews, medical records, and, in some cases, test results. Research topics include (a) the utility of assessment instruments, (b) the utility of standardized versus tailored treatment plans, (c) the appropriateness of evaluations of competency to consent to treatment, and (d) the use of evidence-based recommendations for treatment.

Nelson-Gray (2003) described a method for studying the utility of assessment instruments. In the following example, she related using the method to evaluate a semistructured interview:

If treatment was more successful for individuals whose diagnosis had been established by a set of assessment devices that included the semistructured interview than by a set of assessment devices that excluded the semistructured interview, then . . . the treatment utility of the semistructured interview would be demonstrated (p. 524).

This method is rarely used, so little is directly known about the utility of assessment instruments such as semistructured interviews and psychological tests (Hayes et al. 1987, Meehl 1959, Nelson-Gray 2003).

A treatment plan that is formulated on an individual basis is not necessarily better than a standardized one. In one study (Schulte et al. 1992), 120 phobic

clients were randomly assigned to one of three groups: (a) a control group, (b) a standardized therapy group (exposure in vivo), and (c) a group receiving individually planned treatment. For the clients receiving individually planned treatment, therapists were allowed to use all of the therapeutic methods commonly employed in behavior therapy and cognitive therapy. Standardized treatment proved to be the most successful, which suggests that in some instances the utility of standardized treatment is greater than the utility of tailored treatment plans.

The most vigorous area of research on clinical decision making in the past six to seven years has been on the use of evidence-based treatment recommendations. But before this body of research is discussed, one other study is first described. Although we have known for years that many psychiatrists frequently do not attend to all of the necessary legal criteria when seeking civil commitments for psychiatric hospitalizations (e.g., Bagby et al. 1991, Garb 1998), a recent study raises questions about how they decide if a patient is competent to agree to electroconvulsive therapy (Kitamura & Kitamura 2000). In this study, 176 members of the Japanese Society of Psychiatry and Neurology read transcripts of competency interviews for five patients who were recommended to undergo electroconvulsive therapy because drugs had failed to improve their mental status. When indicating if a patient was competent to make this treatment decision, interrater reliability among clinicians was poor. The intraclass correlation coefficient was only 0.31.

The treatment of mental disorders may be transformed by the adoption of evidence-based practice recommendations. For example, the Agency for Health Care Policy and Research and the National Institute of Mental Health established the Schizophrenia Patient Outcomes Research Team (PORT) to develop and disseminate recommendations for treating schizophrenia based on scientific evidence. Their recommendations addressed the use of antipsychotic agents as well as the use of psychological, family, and other interventions (Lehman et al. 1998b). With the publication of treatment recommendations, investigators have been able to evaluate treatment decisions made by mental health professionals to learn if they are consistent with evidence-based practice. In general, results indicate that implementation of the treatment recommendations would transform clinical practice.

For the treatment of schizophrenia, mental health professionals frequently deviate from recommendations for evidence-based practice. For example, the PORT (Lehman et al. 1998a) surveyed a stratified random sample of 719 individuals diagnosed with schizophrenia in inpatient and outpatient settings. Treatment was in conformance with the treatment recommendations less than half of the time, although it was better for pharmacological than for psychosocial treatments. Treatment recommendations were followed less often for minority patients. For example, dosage of antipsychotic medication for an acute episode should be in the range of 300–1000 chlorpromazine equivalents per day for a minimum of six weeks. Lehman et al. (1998a) found that only 62.4% of inpatients received the recommended dosage at discharge. Minority patients were more likely to be on a high dose (>1000 chlorpromazine equivalents) than were Caucasian patients (27.4% and 15.9%, respectively).

Since the publication of the PORT findings and recommendations, a large number of studies have been conducted to determine how well mental health professionals follow evidence-based practice when treating schizophrenia. In general, half of the patients with schizophrenia frequently do not receive antipsychotic medications within recommended dosages; black patients frequently receive excessively high dosages of antipsychotic medicine, and they are less likely to be tried on atypical psychotics (Chen et al. 2000, Covell et al. 2002, Daumit et al. 2003, Dickey et al. 2003, DosReis et al. 2002, Herbeck et al. 2004, Kreyenbuhl et al. 2003, Owen et al. 2001, Sohler et al. 2003, Valenstein et al. 2001, Valenti et al. 2003, Walkup et al. 2000; also see Botts et al. 2003, Buchanan et al. 2002, Remington et al. 2001, Weissman 2002, Woods et al. 2003). Conformance with psychosocial treatment recommendations was studied less frequently, although, to give just one example, Lehman et al. (1998) reported that patients were referred for family therapy only 13.9% to 26.8% of the time, even though research supports the use of specified forms of family therapy in conjunction with psychotropic medicine for the treatment of schizophrenia.

## THE ROMANTIC AND EMPIRICIST TRADITIONS IN CLINICAL PSYCHOLOGY

As noted above, two traditions in clinical psychology can be described. Both empiricists and romanticists base their judgments on a combination of scientific findings, informal observations, and clinical lore, although empiricists place a greater emphasis on scientific findings. A key distinction is that those in the romantic tradition are likely to accept findings based on clinical validation, whereas those in the empiricist tradition are likely to maintain a skeptical attitude. Thus, when Caldwell (2004) interpreted the Minnesota Multiphasic Personality Inventory and inferred that his client would have scars from being burned with cigarettes, he was offering a clinical validation of his method for interpreting the inventory. Empiricists insist upon empirical validation: They would insist that a study be conducted with appropriate controls. That is, they would insist that an expert make judgments for a series of clients and that a research investigator record the validity of the judgments. It may seem harsh to question an anecdote told by an expert clinician, but when expert psychologists have been the objects of study, their claims frequently have not been supported (Garb 1989; Wood et al. 2003, pp. 136–142).

Silver (2001), upon receiving the American Psychological Association's Award for Distinguished Contributions to Applied Psychology, made an argument that exemplifies the romanticist tradition. He minimized the importance of scientific research and praised the process of clinical validation. Silver related that he first forms an impression of a client using biographical and interview data. He then predicts how a client will perform on a test. If his predictions turn out to be accurate, he concludes that his impression of the client and his use of the test are likely to be accurate. If his predictions are inaccurate, he collects additional information. Silver gave the following example:

I even use psychological tests whose validity is in question. For example, I will administer the Draw-A-Person Test (Machover 1949). Again, I make specific predictions. Without so doing, I would not use this test because of the weak support for its validity (Silver 2001, p. 1009).

Just as Caldwell (2004) referred to clinical experience to validate his use of a test, Silver does the same. The problem with Silver's approach is that a psychologist's prediction of test scores will sometimes be correct by chance, not because the psychologist's overall impression of the client is accurate. When this occurs, the psychologist will be misled into believing that his impression of the client and his use of the test have been validated.

Another example illustrates the romantic and empiricist traditions in clinical psychology (additional examples are given in Garb et al. 2004). As noted above, psychologists frequently make diagnoses by comparing clients to prototypes. That is, they frequently do not adhere to the DSM criteria, but instead form an impression of how similar a client is to their conception of the prototypical client with a particular disorder. Shedler & Westen (2004) have recommended that the DSM diagnostic criteria for personality disorders be revised so that mental health professionals simply rate how similar a client is to a prototype. To generate prototypes, Shedler & Westen asked experienced psychiatrists and psychologists to describe their conceptions (prototypes) of personality disorders. Shedler & Westen reported that the prototypes "were clinically richer than the DSM descriptions" (p. 1350). The approach recommended by Shedler & Westen (2004) falls into the romantic tradition because it relies on the "combined experience of seasoned clinical practitioners" while using statistical methods to aggregate the ratings made by those clinicians (p. 1364). Using this approach, the DSM criteria would not be revised on the basis of research studies on the etiology, nature, and course of a mental disorder, but instead on the basis of clinicians' observations.

Research on clinical judgment and decision making supports the empiricist tradition. A large body of research indicates that it can be surprisingly difficult to learn from informal observations, both because clinicians' cognitive processes are fallible and because accurate feedback on the validity of judgments is frequently not available in clinical practice (Chapman & Chapman 1969; Garb 1989, 1998). Furthermore, when clinical lore is studied, it often is found to be invalid. For example, according to clinical lore, the comprehension and picture-arrangement subtests of the Wechsler intelligence tests are sensitive measures of social judgment. However, when this was examined empirically (Lipsitz et al. 1993), scores on these Wechsler subtests were not significantly related to a clinician-rated measure of overall social competence. Furthermore, there is reason to believe that most inferences based on Wechsler subtest variation "are either untested by science or unsupported by scientific findings" (Kamphaus 1998, p. 46; also see Watkins 2003). In conclusion, psychologists should reduce their reliance on informal observation and clinical validation when (a) choosing an assessment instrument or treatment intervention, (b) revising diagnostic criteria, and (c) making clinical judgments or test interpretations.

**The Annual Review of Clinical Psychology is online at  
<http://clinpsy.annualreviews.org>**

**LITERATURE CITED**

- Ambady N, Rosenthal R. 1992. Thin slices of expressive behavior as predictors of interpersonal consequences: a meta-analysis. *Psychol. Bull.* 111:256–74
- American Psychiatric Association. 1980. *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: Am. Psychiatr. Assoc. 3rd ed.
- American Psychiatric Association. 1994. *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: Am. Psychiatr. Assoc. 4th ed.
- Arkes HR. 1981. Impediments to accurate clinical judgment and possible ways to minimize their impact. *J. Consult. Clin. Psychol.* 49:323–30
- Bagby RM, Thompson JS, Dickens SE, Nohara M. 1991. Decision making in psychiatric civil commitment: an experimental analysis. *Am. J. Psychiatry* 148:28–33
- Basco MR, Bostic JQ, Davies D, Rush AJ, Witte B, et al. 2000. Methods to improve diagnostic accuracy in a community mental health setting. *Am. J. Psychiatry* 157:1599–605
- Baumeister RF, Campbell JD, Krueger JI, Vohs KD. 2003. Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychol. Sci. Public Interest* 4:1–44
- Blashfield RK, Herkov MJ. 1996. Investigating clinician adherence to diagnosis by criteria: a replication of Morey and Ochoa 1989. *J. Personal. Disord.* 10:219–28
- Blashfield RK, Sprock J, Haymaker D, Hodgins J. 1989. The family resemblance hypothesis applied to psychiatric classification. *J. Nerv. Ment. Dis.* 177:492–97
- Blashfield RK, Sprock J, Pinkson K, Hodgins J. 1985. Exemplar prototypes of personality disorder diagnoses. *Compr. Psychiatry* 26:11–21
- Botts S, Hines H, Littrell R. 2003. Antipsychotic pharmacy in the ambulatory care setting, 1993–2000. *Psychiatr. Serv.* 54:1086
- Buchanan RW, Kreyenbuhl J, Zito JM, Lehman A. 2002. The Schizophrenia PORT Pharmacological Treatment Recommendations: conformance and implications for symptoms and functional outcome. *Schizophr. Bull.* 28:63–73
- Burkett BG, Whitley G. 1998. *Stolen Valor: How the Vietnam Generation Was Robbed of Its Heroes and Its History*. Dallas, TX: Verity
- Caldwell AB. 2004. My love affair with an instrument. *J. Personal. Assess.* 82:4–10
- Cantor N, Smith EE, French R, Mezzich J. 1980. Psychiatric diagnosis as prototype categorization. *J. Abnorm. Psychol.* 89:181–93
- Chapman LJ, Chapman JP. 1969. Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *J. Abnorm. Psychol.* 74:271–80
- Chen RS, Nadkarni PM, Levin FL, Miller PL, Erdos J, et al. 2000. Using a computer database to monitor compliance with pharmacotherapeutic guidelines for schizophrenia. *Psychiatr. Serv.* 51:791–94
- Cooper JE, Kendell RE, Gurland BJ, Sharpe L, Copeland JRM, et al. 1972. *Psychiatric Diagnosis in New York and London*. London: Oxford Univ. Press
- Costa PT, Widiger TA. 2002. *Personality Disorders and the Five-Factor Model of Personality*. Washington, DC: Am. Psychol. Assoc. 2nd ed.
- Covell NH, Jackson CT, Evans AC, Essock SM. 2002. Antipsychotic prescribing practices in Connecticut's public mental health system: rates of changing medications and prescribing styles. *Schizophr. Bull.* 28:17–29
- Daleiden EL, Chorpita BF, Kollins SH, Drabman RS. 1999. Factors affecting the reliability of clinical judgments about the function

- of children's school-refusal behavior. *J. Clin. Child Psychol.* 28:396–406
- Daumit GL, Crum RM, Guallar E, Power NR, Primm AB, et al. 2003. Outpatient prescriptions for atypical antipsychotics for African Americans, Hispanics, and whites in the United States. *Arch. Gen. Psychiatry* 60:121–28
- Davis RT, Blashfield RK, McElroy RA. 1993. Weighting criteria in the diagnosis of personality disorder: a demonstration. *J. Abnorm. Psychol.* 102:319–22
- Dawes RM. 1986. Representative thinking in clinical judgment. *Clin. Psychol. Rev.* 6:425–41
- DePaulo BM. 1994. Spotting lies: Can humans learn to do better? *Curr. Dir. Psychol. Sci.* 3:83–86
- Dickey B, Normand S-LT, Hermann RC, Eisen SV, Cortes DE, et al. 2003. Guideline recommendations for treatment of schizophrenia: the impact of managed care. *Arch. Gen. Psychiatry* 60:340–48
- DosReis S, Zito JM, Buchanan RW, Lehman AF. 2002. Antipsychotic dosing and concurrent psychotropic treatments for Medicaid-insured individuals with schizophrenia. *Schizophr. Bull.* 28:607–17
- Ekman P, Friesen WV. 1978. *The Facial Action Coding System*. Palo Alto, CA: Consult. Psychol. Press
- Ekman P, O'Sullivan M. 1991. Who can catch a liar? *Am. Psychol.* 46:913–20
- Ekman P, O'Sullivan M, Frank MG. 1999. A few can catch a liar. *Psychol. Sci.* 10: 263–66
- Elbogen EB, Williams AL, Kim D, Tomkins AJ, Scalora MJ. 2001. Gender and perceptions of dangerousness in civil psychiatric patients. *Leg. Criminol. Psychol.* 6: 215–28
- Evans DL, Herbert JD, Nelson-Gray RO, Gaudiano BA. 2002. Determinants of diagnostic prototypicality judgments of the personality disorders. *J. Personal. Disord.* 16: 95–106
- Ford MR, Widiger TA. 1989. Sex bias in the diagnosis of histrionic and antisocial personality disorders. *J. Consult. Clin. Psychol.* 57:301–5
- Frank MG, Ekman P. 1997. The ability to detect deceit generalizes across different types of high-stakes lies. *J. Personal. Soc. Psychol.* 72:1429–39
- Garb HN. 1989. Clinical judgment, clinical training, and professional experience. *Psychol. Bull.* 105:387–96
- Garb HN. 1996. The representativeness and past-behavior heuristics in clinical judgment. *Prof. Psychol. Res. Pract.* 27:272–77
- Garb HN. 1998. *Studying the Clinician: Judgment Research and Psychological Assessment*. Washington, DC: Am. Psychol. Assoc.
- Garb HN. 2000. Computers will become increasingly important for psychological assessment: not that there's anything wrong with that? *Psychol. Assess.* 12:31–39
- Garb HN. 2003. Incremental validity and the assessment of psychopathology in adults. *Psychol. Assess.* 15:508–20
- Garb HN, Wood JM, Lilienfeld SO, Nezworski MT. 2005. Roots of the Rorschach controversy. *Clin. Psychol. Rev.* In press
- Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. 2000. Clinical versus mechanical prediction: a meta-analysis. *Psychol. Assess.* 12:19–30
- Hansen SS, Munk-Jorgensen P, Guldbaek B, Solgard T, Lauszus KS, et al. 2000. Psychoactive substance use diagnoses among psychiatric in-patients. *Acta Psychiatr. Scand.* 102:432–38
- Hanson RK. 1997. *The Development of a Brief Actuarial Scale for Sexual Offense Recidivism*. Ottawa, Ont.: Dept. Solicitor Gen.
- Hanson RK, Bussiere MT. 1998. Predicting relapse: a meta-analysis of sexual offender recidivism studies. *J. Consult. Clin. Psychol.* 66:348–62
- Hanson RK, Morton KE, Harris AJR. 2003. Sexual offender recidivism risk: what we know and what we need to know. *Ann. NY Acad. Sci.* 989:154–66
- Hanson RK, Thornton D. 1999. *Static-99: Improving Actuarial Risk Assessments for Sex Offenders*. Ottawa, Ont.: Dept. Solicitor Gen.

- Hare RD. 1991. *The Hare Psychopathy Checklist-Revised (PCL-R)*. Toronto, Ont.: Multi-Health Syst.
- Harris GT, Rice ME, Quinsey VL. 1993. Violent recidivism of mentally disordered offenders: the development of a statistical prediction instrument. *Crim. Justice Behav.* 20:315–35
- Hathaway SR, McKinley JC. 1943. *The Minnesota Multiphasic Personality Inventory*. Minneapolis: Univ. Minn. Press
- Hayes SC, Nelson RO, Jarrett RB. 1987. The treatment utility of assessment: a functional approach to evaluating assessment quality. *Am. Psychol.* 42:963–74
- Herbeck DM, West JC, Ruditis I, Duffy FF, Fitek DJ, et al. 2004. Variations in use of second-generation antipsychotic medication by race among adult psychiatric patients. *Psychiatr. Serv.* 55:677–84
- Hilsenroth MJ, Callahan KL, Eudell EM. 2003. Further reliability, convergent and discriminant validity of overall defensive functioning. *J. Nerv. Ment. Dis.* 191:730–37
- Hoptman MJ, Yates KF, Patalinjug MB, Wack RC, Convit A. 1999. Clinical prediction of assaultive behavior among male psychiatric patients at a maximum-security forensic facility. *Psychiatr. Serv.* 50:1461–66
- Hunsley J, Bailey JM. 1999. The clinical utility of the Rorschach: unfulfilled promises and an uncertain future. *Psychol. Assess.* 11:266–77
- Jopp DA, Keys CB. 2001. Diagnostic overshadowing reviewed and reconsidered. *Am. J. Ment. Retard.* 106:416–33
- Kahneman D. 2003. A perspective on judgment and choice: mapping bounded rationality. *Am. Psychol.* 58:697–720
- Kamphaus RW. 1998. Intelligence test interpretation: acting in the absence of evidence. In *WISC-III Clinical Use and Interpretation: Scientist-Practitioner Perspectives*, ed. A Prifitera, DH Saklofske, pp. 39–57. New York: Academic
- Kayne NY, Alloy LB. 1988. Clinician and patient as aberrant actuaries: expectation-based distortions in assessment of covariation. In *Social Cognition and Clinical Psychology: A Synthesis*, ed. LY Abramson, pp. 295–365. New York: Guilford
- Kihlstrom JF. 1999. The psychological unconscious. In *Handbook of Personality*, ed. LR Pervin, O John, pp. 424–42. New York: Guilford. 2nd ed.
- Kim NS, Ahn W. 2002. Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *J. Exp. Psychol.: Gen.* 131:451–76
- Kirchner JE, Owen RR, Nordquist C, Fischer EP. 1998. Diagnosis and management of substance use disorders among inpatients with schizophrenia. *Psychiatr. Serv.* 49:82–85
- Kitamura T, Kitamura F. 2000. Reliability of clinical judgment of patients' competency to give informed consent: a case vignette study. *Psychiatr. Clin. Neurosci.* 54:245–47
- Kraemer HC, Measelle JR, Ablow JC, Essex MJ, Boyce WT, Kupfer DJ. 2003. A new approach to integrating data from multiple informants in psychiatric assessment and research: mixing and matching contexts and perspectives. *Am. J. Psychiatry* 160:1566–77
- Kreyenbuhl J, Zito JM, Buchanan RW, Soeken KL, Lehman AF. 2003. Racial disparity in the pharmacological management of schizophrenia. *Schizophr. Bull.* 29:183–93
- Lally SJ. 2003. What tests are acceptable for use in forensic evaluations? A survey of experts. *Prof. Psychol. Res. Pract.* 34:491–98
- LeFever GB, Arcona AP, Antonuccio DO. 2003. ADHD among schoolchildren: evidence of overdiagnosis and overuse of medication. *Sci. Rev. Ment. Health Pract.* 2:49–60
- Lehman AF, Steinwachs DM, Co-Investigators of the PORT Project. 1998a. Patterns of usual care for schizophrenia: initial results from the Schizophrenia Patient Outcomes Research Team (PORT) client survey. *Schizophr. Bull.* 24:11–20
- Lehman AF, Steinwachs DM, Co-Investigators of the PORT Project. 1998b. Translating research into practice: the Schizophrenia Patient Outcomes Research Team (PORT)

- treatment recommendations. *Schizophr. Bull.* 24:1-10
- Lidz CW, Mulvey EP, Gardner W. 1993. The accuracy of predictions of violence to others. *JAMA* 269:1007-11
- Lilienfeld SO, Lynn SJ, Kirsch I, Chaves JF, Sarbin TR, et al. 1999. Dissociative identity disorder and the sociocognitive model: recalling the lessons of the past. *Psychol. Bull.* 125:507-23
- Lilienfeld SO, Wood JM, Garb HN. 2000. The scientific status of projective techniques. *Psychol. Sci. Public Interest* 1:27-66
- Lipsitz JD, Dworkin RH, Erlenmeyer-Kimling L. 1993. Wechsler Comprehension and Picture Arrangement subtests and social adjustment. *Psychol. Assess.* 5:430-37
- Machover K. 1949. *Personality Projection in the Drawing of the Human Figure*. Springfield, IL: Thomas
- McNally RJ. 2003a. Progress and controversy in the study of posttraumatic stress disorder. *Annu. Rev. Psychol.* 54:229-52
- McNally RJ. 2003b. *Remembering Trauma*. Cambridge, MA: Belknap Press of Harvard Univ. Press
- Meehl PE. 1954. *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Minneapolis: Univ. Minn. Press
- Meehl PE. 1959. Some ruminations on the validation of clinical procedures. *Can. J. Psychol.* 13:102-28
- Miller PR, Dasher R, Collins R, Griffiths P, Brown F. 2001. Inpatient diagnostic assessments: 1. Accuracy of structured vs. unstructured interviews. *Psychiatry Res.* 105: 255-64
- Monahan J. 1981. *The Clinical Prediction of Violent Behavior*. Washington, DC: US Gov. Print. Off.
- Monahan J, Steadman HJ, Silver E, Appelbaum PS, Robbins PC, et al. 2001. *Rethinking Risk Assessment: The MacArthur Study of Mental Disorder and Violence*. New York: Oxford Univ. Press
- Morey LC, Ochoa ES. 1989. An investigation of adherence to diagnostic criteria: clinical diagnosis of the DSM-III personality disorders. *J. Personal. Disord.* 3:180-92
- Mossman D. 1994. Assessing predictions of violence: being accurate about accuracy. *J. Consult. Clin. Psychol.* 62:783-92
- Nelson-Gray RO. 2003. Treatment utility of psychological assessment. *Psychol. Assess.* 15:521-31
- Owen RR, Feng W, Thrush CR, Hudson TJ, Austen MA. 2001. Variations in prescribing practices for novel antipsychotic medications among Veterans Affairs Hospitals. *Psychiatr. Serv.* 52:1523-25
- Perry JC. 1992. Problems and considerations in the valid assessment of personality disorders. *Am. J. Psychiatry* 149:1645-53
- Perry JC, Hoglend P, Shear K, Vaillant GE, Horowitz M, et al. 1998. Field trial of a diagnostic axis for defense mechanisms for DSM-IV. *J. Personal. Disord.* 12:56-68
- Persons JB, Bertagnolli A. 1999. Interrater reliability of cognitive-behavioral case formulations of depression: a replication. *Cogn. Ther. Res.* 23:271-83
- Quinsey VL, Harris GT, Rice ME, Cormier CA. 1998. *Violent Offenders: Appraising and Managing Risk*. Washington, DC: Am. Psychol. Assoc.
- Reiss S, Levitan G, McNally RJ. 1982. Emotionally disturbed mentally retarded people: an underserved population. *Am. Psychol.* 37: 361-67
- Remington G, Shammi CM, Sethna R, Lawrence R. 2001. Antipsychotic dosing patterns for schizophrenia in three treatment settings. *Psychiatr. Serv.* 52:96-98
- Rueter MA, Chao W, Conger RD. 2000. The effect of systematic variation in retrospective conduct disorder reports on antisocial personality disorder diagnoses. *J. Consult. Clin. Psychol.* 68:307-12
- Sartorius N, Kaelber CT, Cooper JE, Roper MT, Rae DS, et al. 1993. Progress toward achieving a common language in psychiatry: results from the field trial of the clinical guidelines accompanying the WHO classification of mental and behavioral disorders in ICD-10. *Arch. Gen. Psychiatry* 50:115-24

- Sartorius N, Ustun TB, Korten A, Cooper JE, van Drimmelen J. 1995. Progress toward achieving a common language in psychiatry. II: Results from the international field trials of the ICD-10 diagnostic criteria for research for mental and behavioral disorders. *Am. J. Psychiatry* 152:1427-37
- Schulte D, Kunzel R, Pepping G, Schulte-Bahrenberg T. 1992. Tailor-made versus standardized therapy of phobic patients. *Adv. Behav. Res. Ther.* 14:67-92
- Shear MK, Greeno C, Kang J, Ludewig D, Frank E, et al. 2000. Diagnosis of nonpsychotic patients in community clinics. *Am. J. Psychiatry* 157:581-87
- Shedler J, Westen D. 2004. Refining personality disorder diagnosis: integrating science and practice. *Am. J. Psychiatry* 161:1350-65
- Shiffman S, Gwaltney CJ, Balabanis MH, Liu KS, Paty JA, et al. 2002. Immediate antecedents of cigarette smoking: an analysis from ecological momentary assessment. *J. Abnorm. Psychol.* 111:531-45
- Silver RJ. 2001. Practicing professional psychology. *Am. Psychol.* 56:1008-14
- Slovic P, Finucane M, Peters E, MacGregor DG. 2002. The affect heuristic. In *Heuristics and Biases: The Psychology of Intuitive Judgment*, ed. T Gilovich, D Griffin, D Kahneman, pp. 397-420. Cambridge, UK: Cambridge Univ. Press
- Sohler NL, Walkup J, McAlpine D, Boyer C, Olfson M. 2003. Antipsychotic dosage at hospital discharge and outcomes among persons with schizophrenia. *Psychiatr. Serv.* 54:1258-63
- Spitzer RL. 1983. Psychiatric diagnosis: Are clinicians still necessary? *Compr. Psychiatry* 24:399-411
- Strakowski SM, Hawkins JM, Keck PE, McElroy SL, West SA, et al. 1997. The effects of race and information variance on disagreement between psychiatric emergency service and research diagnoses in first-episode psychosis. *J. Clin. Psychiatry* 58:457-63
- Tolman AO, Mullendore KB. 2003. Risk evaluations for the courts: Is service quality a function of specialization? *Prof. Psychol. Res. Pract.* 34:225-32
- Turk DC, Salovey P, eds. 1988. *Reasoning, Inference, and Judgment in Clinical Psychology*. New York: Free Press
- Tversky A, Kahneman D. 1974. Judgment under uncertainty: heuristics and biases. *Science* 185:1124-31
- Valenstein M, Copeland L, Owen R, Blow F, Visnic S. 2001. Delays in adopting evidence-based dosages of conventional antipsychotics. *Psychiatr. Serv.* 52:1242-44
- Valenti AM, Narendran R, Pristach CA. 2003. Who are patients on conventional antipsychotics? *Schizophr. Bull.* 29:195-99
- Wakefield JC, Kirk SA, Pottick KJ, Hsieh D. 1999. Disorder attribution and clinical judgment in the assessment of adolescent antisocial behavior. *Soc. Work Res.* 23:227-38
- Walkup JT, McAlpine DD, Olfson M, Labay LE, Boyer C, et al. 2000. Patients with schizophrenia at risk for excessive antipsychotic dosing. *J. Clin. Psychiatry* 61:344-48
- Watkins MW. 2003. IQ subtest analysis: clinical acumen or clinical illusion? *Sci. Rev. Ment. Health Pract.* 2:118-41
- Webster CD, Douglas KS, Eaves D, Hart SD. 1998. *The HCR-20: Assessing Risk for Violence* (Version 2). Odessa, FL: Psychol. Assess. Resource.
- Wedding D, Faust D. 1989. Clinical judgment and decision making in neuropsychology. *Arch. Clin. Neuropsychol.* 4:233-65
- Weissman EM. 2002. Antipsychotic prescribing practices in the Veterans Healthcare Administration—New York metropolitan region. *Schizophr. Bull.* 28:31-42
- Westen D. 1997. Divergences between clinical and research methods for assessing personality disorders: implications for research and the evolution of Axis II. *Am. J. Psychiatry* 154:895-903
- Whaley AL. 2001. Cultural mistrust and the clinical diagnosis of paranoid schizophrenia in African American patients. *J. Psychopathol. Behav. Assess.* 23:93-100
- Widiger TA. 2002. Personality disorders. In

- Handbook of Assessment and Treatment Planning for Psychological Disorders*, ed. MM Antony, DH Barlow, pp. 453–80. New York: Guilford
- Widiger TA, Clark LA. 2000. Toward DSM-V and the classification of psychopathology. *Psychol. Bull.* 126:946–63
- Widiger TA, Samuel DB. 2005. Evidence based assessment of personality disorders. *Psychol. Assess.* In press
- Wood JM, Nezworski MT, Lilienfeld SO, Garb HN. 2003. *What's Wrong With the Rorschach? Science Confronts the Controversial Inkblot Test*. San Francisco: Jossey-Bass
- Woods SW, Sullivan MC, Neuse EC, Diaz E, Baker CB, et al. 2003. Racial and ethnic effects on antipsychotic prescribing practices in a community mental health center. *Psychiatr. Serv.* 54:177–79
- Wu KD, Clark LA. 2003. Relations between personality traits and self-reports of daily behavior. *J. Res. Personal.* 37:231–56
- Zimmerman M, Mattia JI. 1999. Psychiatric diagnosis in clinical practice: Is comorbidity being missed? *Compr. Psychiatry* 40:182–91

## CONTENTS

---

A HISTORY OF CLINICAL PSYCHOLOGY AS A PROFESSION IN AMERICA (AND A GLIMPSE AT ITS FUTURE), <i>Ludy T. Benjamin, Jr.</i>	1
STRUCTURAL EQUATION MODELING: STRENGTHS, LIMITATIONS, AND MISCONCEPTIONS, <i>Andrew J. Tomarken and Niels G. Waller</i>	31
CLINICAL JUDGMENT AND DECISION MAKING, <i>Howard N. Garb</i>	67
MOTIVATIONAL INTERVIEWING, <i>Jennifer Hettema, Julie Steele, and William R. Miller</i>	91
STATE OF THE SCIENCE ON PSYCHOSOCIAL INTERVENTIONS FOR ETHNIC MINORITIES, <i>Jeanne Miranda, Guillermo Bernal, Anna Lau, Laura Kohn, Wei-Chin Hwang, and Teresa La Fromboise</i>	113
CULTURAL DIFFERENCES IN ACCESS TO CARE, <i>Lonnie R. Snowden and Ann-Marie Yamada</i>	143
COGNITIVE VULNERABILITY TO EMOTIONAL DISORDERS, <i>Andrew Mathews and Colin MacLeod</i>	167
PANIC DISORDER, PHOBIAS, AND GENERALIZED ANXIETY DISORDER, <i>Michelle G. Craske and Allison M. Waters</i>	197
DISSOCIATIVE DISORDERS, <i>John F. Kihlstrom</i>	227
THE PSYCHOBIOLOGY OF DEPRESSION AND RESILIENCE TO STRESS: IMPLICATIONS FOR PREVENTION AND TREATMENT, <i>Steven M. Southwick, Meena Vythilingam, and Dennis S. Charney</i>	255
STRESS AND DEPRESSION, <i>Constance Hammen</i>	293
THE COGNITIVE NEUROSCIENCE OF SCHIZOPHRENIA, <i>Deanna M. Barch</i>	321
CATEGORICAL AND DIMENSIONAL MODELS OF PERSONALITY DISORDER, <i>Timothy J. Trull and Christine A. Durrett</i>	355
THE DEVELOPMENT OF PSYCHOPATHY, <i>Donald R. Lynam and Lauren Gudonis</i>	381
CHILD MALTREATMENT, <i>Dante Cicchetti and Sheree L. Toth</i>	409
PSYCHOLOGICAL TREATMENT OF EATING DISORDERS, <i>G. Terence Wilson</i>	439
GENDER IDENTITY DISORDER IN CHILDREN AND ADOLESCENTS, <i>Kenneth J. Zucker</i>	467

THE DEVELOPMENT OF ALCOHOL USE DISORDERS, <i>Kenneth J. Sher, Emily R. Grekin, and Natalie A. Williams</i>	493
DECISION MAKING IN MEDICINE AND HEALTH CARE, <i>Robert M. Kaplan and Dominick L. Frosch</i>	525
PSYCHOLOGY, PSYCHOLOGISTS, AND PUBLIC POLICY, <i>Katherine M. McKnight, Lee Sechrest, and Patrick E. McKnight</i>	557
COGNITIVE APPROACHES TO SCHIZOPHRENIA: THEORY AND THERAPY, <i>Aaron T. Beck and Neil A. Rector</i>	577
STRESS AND HEALTH: PSYCHOLOGICAL, BEHAVIORAL, AND BIOLOGICAL DETERMINANTS, <i>Neil Schneiderman, Gail Ironson, and Scott D. Siegel</i>	607
POSITIVE PSYCHOLOGY IN CLINICAL PRACTICE, <i>Angela Lee Duckworth, Tracy A. Steen, and Martin E. P. Seligman</i>	629
INDEX	
Subject Index	653