The involvement of psychologists and psychiatrists within the legal arena continues to grow rapidly but remains highly controversial. Extensive research on clinical judgment provides a scientific basis for clarifying the growing disputes about the values of such professional activities. Studies show that professionals often fail to reach reliable or valid conclusions and that the accuracy of their judgments does not necessarily surpass that of laypersons, thus raising substantial doubt that psychologists or psychiatrists meet legal standards for expertise. Factors that underlie the research findings and implications for courtroom testimony are discussed.

The Expert Testimony of Psychologists and Psychiatrists at hearings and trials alters many lives (1). Clinicians participate in up to 1 million legal cases annually (2). Depending on the expert’s opinion, an individual may be confined to a mental institution, receive huge monetary awards, obtain custody of a child, or lose his or her life.

Do psychiatrists and psychologists promote or obstruct courtroom justice? Passionate debate on a matter of such great practical and moral import may be inevitable but cannot resolve the issues. However, there is sufficient scientific evidence on the diagnostic and predictive accuracy of clinicians to permit dispassionate evaluation (3). We review the scientific evidence and its application to legal standards for expertise.

The interpretation of research on the performance of clinicians requires familiarity with legal standards for expert status. The standards are somewhat complicated, broadly interpreted, and continually evolving, but they can be reduced to two essentials (4). First, in branches of medicine, an expert must be able to state opinions with “reasonable medical certainty.” This phrase itself is somewhat ambiguous but “pretty likely accurate” is as close a translation as any. An expert should be reasonably certain about the issues or questions pertinent to the case at hand, not merely about his field in general or other specific issues. The physician who is reasonably certain about diagnosis but not etiology may be allowed to testify on the former but not the latter. Second, an expert should be able to help the judge or jury reach a more valid conclusion than would be possible without the expert’s testimony. If laypersons are as accurate as the expert or equally accurate with or without the expert’s help, the expert is not needed.

These standards are easily framed as empirical questions that we will examine in order. (i) Can expert witnesses in psychology and psychiatry answer forensic questions with reasonable accuracy? (ii) Can experts help the judge and jury reach more accurate conclusions than would otherwise be possible?

Reasonable Certainty

Reliability and validity of clinical judgment. The expert witness may attempt to determine the current, prior, or future state of the person under examination. Can the person comprehend the consequences of his actions at the time of the crime? Will an injury preclude future employability? These questions vary in difficulty. We will first review clinicians’ successes in answering the simpler and more practiced questions common to everyday clinical work—those that form the basis for addressing the more complex and less familiar questions common to the courtroom.

There is perhaps nothing more fundamental or basic to the science of psychiatry than the classification or diagnosis of patients, and no more fundamental a hurdle than reliability, or cross-clinician consistency in the diagnoses rendered. If clinicians assign widely varying diagnoses, classification will be as much a product of extraneous factors or the diagnostician’s idiosyncrasies as it is the examinee’s actual status.

Psychiatry has been continuously plagued by difficulties in achieving reliable classification. The American Psychiatric Association has revised the official diagnostic manual at a quickening pace: The first Diagnostic and Statistical Manual of Mental Disorders (DSM-I) was published in 1952, DSM-II in 1968, DSM-III in 1980, and DSM-III-Revised in 1987. The next revision, DSM-IV, is slated for publication in the early 1990s. This process of revision little resembles the refinement of categories or cumulative gains common to advanced scientific fields. DSM-I and II often produced poor inter-rater agreement (5), and the diagnostic system was radically altered with the publication of DSM-III. DSM-III introduced more specific classification procedures, changed hundreds of diagnostic criteria, and added or eliminated numerous categories of disorder. DSM-III-R introduced about 200 additional changes in diagnostic guidelines and criteria.

The initial DSM-III field trials appeared to demonstrate improved diagnostic reliability, but serious methodological shortcomings raised doubts about the results (6). A number of subsequent studies showed that rate of disagreement for specific diagnostic categories often equals or exceeds rate of agreement (7-9). The reliability of DSM-III-R awaits testing because many of the changes were
intended to improve agreement but were made in the absence of formal checks on reliability.

Problems with diagnostic reliability illustrate more general difficulties in achieving interclinician agreement on descriptions of current status. For example, Stoller and Geertsma (10) found that highly experienced psychiatrists who viewed the same psychiatric interview could not agree on the patient's diagnosis, intrapsychic motivations and conflicts, or conscious and unconscious feelings. Problems achieving reliable, much less valid, descriptions of current status help anticipate the results of studies examining the more difficult judgments involved in the determination of prior and future states.

Studies that compare clinicians' predictions against objectively determinable, hard data commonly show that error rate exceeds accuracy rate (3). In one study, for example, a series of military recruits was retained in service despite psychiatrists' recommendations that they be discharged for severe psychiatric liabilities (11). After 2 years, most of these individuals had remained on active duty and their overall rate of success and adjustment was not substantially different from that of matched controls initially judged to be free of pathology.

**Reliability and validity of forensic judgments.** The research on reliability and validity cited above mainly examines clinical questions, not forensic questions, and the two can differ substantially. A clinical diagnosis, for example, may relate minimally to the issues of forensic interest. The clinical criteria for "insanity" or psychosis do not include such tests of legal insanity as the capacity to appreciate the consequences of one's action or to resist an impulse. A determination that the clinical criteria have been met does not establish diminished mental capacity, a diagnosis such as "post-traumatic stress disorder" offers little guidance. According to the clinical criteria, for example, may relate minimally to the issues of forensic defense of diminished mental capacity, a diagnosis such as "post-traumatic stress disorder" among individuals who fall within the same diagnostic category might be used to argue that the person is not capable of understanding the nature of the charges or ability to conform to the requirements of the law. A clinical determination of law is different from that of matched controls initially judged to be free of pathology.

- **Assistance to the Judge and Jury**

Studies show that professional clinicians do not in fact make more accurate clinical judgments than laypersons (3). Some studies show a slight professional advantage and some a slight lay advantage, but most often the groups perform similarly. An early study examined success in distinguishing the visual-motor productions of normal versus brain-damaged individuals on a commonly employed screening test (20). Professional psychologists performed no better than office secretaries. In another study, lay interviewers using standard national sample of U.S. neuropsychologists, practitioners who specialize in the assessment of brain-behavior relations performed better than other professionals. Virtually every available professional who entered the forensic arena also shift from their more familiar role as the patient's helping agent and instead seek to uncover truth, whatever its implications for the person under examination (3, 12). The clinician thus becomes a potential adversary. The forensic role is often less familiar or practiced, the clinician's engrained tendency to support or empathize may cloud objectivity, and the person being examined may be less inclined to disclose information openly and honestly. Clinicians, who usually focus primarily on the patient's subjective reality, must now attempt to determine objective reality, a task for which they may be minimally trained.

The expert witness thus becomes engaged in less familiar questions and activities, often with minimal research backing. Not surprisingly, studies examining the accuracy of judgments directly pertinent to forensic assessment, such as the ability to detect the simulation of disorder (that is, malingering) or to predict violence, have shown particularly high rates of error among clinicians.

A determination of a subject's credibility is often essential in forensic assessment. The potential benefits of a favorable courtroom decision, such as relief from serious criminal charges or large financial gains, can lead individuals to feign disorder. Studies show, however, that clinicians often cannot distinguish the psychological test results of normal subjects asked to feign psychosis (or to simulate brain damage) and actual diagnosed cases (13–15). Faust et al. (16) asked children to lower their performance on tests used to assess brain dysfunction but provided no specific instructions to the children on how to achieve this end. Most practitioners who subsequently reviewed the test results abnormal and identified brain damage as the underlying cause. Although the researchers had listed malingering as one of three possible explanations for the test findings, not one practitioner made the correct identification.

Forensic experts frequently appraise the potential for violent behavior. Their opinions may influence decisions involving criminal sentencing or involuntary commitment. Studies on the prediction of violence are consistent: clinicians are wrong at least twice as often as they are correct (17). Steadman (18) followed 967 individuals who were originally placed in maximum security hospitals on the basis of a psychiatric determination of dangerousness but later released by court order into ordinary mental hospitals. Four years later about half the sample was still in ordinary mental hospitals where violence should have been easily detected, but only 26 subjects in the sample were known to have committed violent acts. These results may overestimate clinical error, for one cannot determine how many individuals discharged into the community committed undetected violent acts. Studies on short-term prediction that are limited to patients within controlled settings provide more reliable measurement of violent episodes. Clinicians' accuracy may not be as low as some of the long-term studies suggest, or short-term prediction may not be as difficult, but error still predominates (19).
ly equal or outperform professionals and laypersons (26). If expertise is defined solely by accuracy, the actuarial method is the "expert." Nevertheless, actuarial procedures typically yield modest levels of accuracy and few procedures, as yet, directly address forensic questions. A notable exception is the set of indices on the Minnesota Multiphasic Personality Inventory (MMPI), which are sensitive to the exaggeration or simulation of disorder (27).

Although experts develop actuarial procedures, actuarial output is often readily understood by laypersons. An output statement may read, for example, "Individuals who obtain similar test results engage in violent behavior in about 20% of cases." Although more than 100 studies demonstrate the superiority of actuarial data combination over clinical judgment, few experts rely strictly on actuarial procedures; indeed, many do not even know that such methods exist. Other experts modify actuarial conclusions at their discretion, although research suggests that this decision strategy results in fewer corrected errors than correct conclusions overturned (26). When actuarial procedures are applicable and intelligible to laypersons, the expert's involvement in the interpretive process is unnecessary. In fact, the expert will most likely move the jury farther from the truth, not closer to it, given the common tendency to counteract actuarial conclusions and thereby decrease overall judgmental accuracy.

Factors Limiting Clinical Judgment

An understanding of the factors that underlie research findings and that foster clinicians' misappraisal of their judgmental accuracy may help assuage unfounded inferences about experts' mental power or honesty. A more productive social science–law relation also ultimately depends on a better understanding of the factors underlying judgment error and the development of corrective procedures.

Limits in scientific knowledge. Practitioners are limited by the state of their science. The inadequacies of classification have been described. In addition, psychology lacks a formalized, general theory of human behavior that permits accurate prediction. Most personality theories are verbal sumations of loosely bound conjectures. The subject matter of the field itself—human thought and behavior—resists objective, direct, or reliable observation and measurement.

One manifestation of the fledgling state of scientific psychology is the tremendous diversity within the field, a situation that is incongruous with the law's preference for standard procedures and authoritativeness. There are dozens of personality theories and hundreds of approaches to psychotherapy (28). Two neuropsychologists may administer entirely different test batteries to the same examinee. This diversity in theory and practice breeds the divergence in opinion that makes the "battle of the experts" a regular courtroom occurrence.

Furthermore, the instability of theory and method hinders the accumulation of scientific knowledge. What is new may not be better, but only a fresh attempt to solve a recalcitrant problem. Each time the official diagnostic manual changes one must discard hundreds of investigations relating scores on psychological tests to what are now obsolete categories of disorder. The MMPI, perhaps the best researched psychological test, is itself undergoing revision; thus, investigators must reevaluate the relation of MMPI scores to a recently revised diagnostic manual that will be re-revised within a few years.

Limits in clinical judgment. The clinician, who is limited by the state of his scientific field and likely disregards or undervalues actuarial data combination, depends mainly on subjective methods of data interpretation. Without the safeguards of the scientific method, clinicians are highly vulnerable to the problematic judgment practices and cognitive limitations common to human beings (29, 30).

For example, clinicians disregard or underuse information about the frequency of occurrence, or base rates (31). Many diagnostic signs within psychology show associations of modest strength, at best, with the condition or event of interest. For example, a test indicator of suicidal intent may occur in 80% of true cases but also in 10% of negative cases. As such, the value of this and other diagnostic indicators is never constant but relative to the frequency of events. If suicidal intent is present in one per 1000 patients, this one patient will likely be identified correctly. However, 10% of the remaining 999 patients, or about 99, will be misidentified as suicidal, resulting in almost 100 times more errors than correct identifications. If the frequencies shift, the sign's value shifts also. Given typical limitations in the strength of signs and the low frequency of most psychiatric disorders, numerous diagnostic signs produce more errors than correct identifications. Many faulty signs remain popular because disregard of base rates and associated principles of probability preclude an accurate determination of their worth.

Clinicians also overvalue supportive evidence and undervalue counterevidence (32). In psychology, the selective pursuit of supportive evidence is especially pernicious. Individual behavior is highly variable across time and situation, and tremendous overlap exists across criteria for various psychiatric disturbances and between the characteristics of aberrant and normal individuals. The lives of normal individuals commonly contain the full range of trauma, stress, and turmoil found among the disordered (33). Clinicians typically expect to find abnormality, and a search for supportive evidence will almost always "succeed" regardless of the examinee's mental health. In one study that enhanced the expectancy to find abnormality, every psychiatrist who heard a script portraying a well-adjusted individual nevertheless diagnosed mental disorder (34). This tendency to assume the presence of abnormality and then seek supportive evidence fosters "overpathologizing," that is, the frequent misidentification of individuals as abnormal.

Selective attention to supportive evidence also fosters "illusory correlations," or the belief in relations that appear to be, but are not, valid (23). Suppose that a diagnostic "sign" and a disorder are actually unrelated but sometimes co-occur by chance alone. The clinician who neglects instances in which the sign or disorder appears independently and rather focuses on co-occurrences, comes to believe that the two are related. For example, some clinicians believe that individuals who produce human figure drawings with accentuated eyes have "paranoid" traits. The repeated "discovery" of "confirming" instances, embedded in the context of salient personal experience, creates a compelling illusion that overpowers any awareness of contrary instances or scientific research. Clinicians continue to use human figure drawings despite scientific evidence that disconfirms the perceived association between accentuated eyes and paranoia, and other assumed relations between drawing characteristics and personality traits.

Studies on experience and accuracy show that the conditions under which clinicians practice do not promote experiential learning, a finding that confirmatory bias and illusory correlation help to explain. Clinicians often receive little or no outcome information or feedback about their judgments, which precludes self-correction. The feedback clinicians do receive is often garbled and prone to the same problematic judgment practices that hinder original case appraisals.

Most clinical feedback occurs in the context of therapy. This feedback is skewed and confounded with outcome. To illustrate—clientele particularly pleased with services may be most likely to make follow-up contacts with the therapist, in which they further express praise and thanks. The therapist obtains a select, rather than
a representative sample of the varying pieces that comprise outcome as a whole. Further, therapists' initial appraisals produce actions that can lead to self-fulfilling prophecies. The therapist who decides he would not work well with a patient and transfers the case will never find his judgment disconfirmed. Additionally, clients may purposely or inadvertently provide misleading feedback. Clinicians often evaluate their own judgmental accuracy by observing patients' agreement with their interpretations or descriptions. However, research shows that individuals believe in overly general personality descriptors of dubious validity, a form of suggestibility that provides a livelihood for astrologers and palm readers and misguides clinicians (35).

Selective attention to supportive evidence similarly affects clinicians' appraisals of their own judgmental accuracy (36). A clinician will inevitably receive some outcome information that appears to support his conclusions. The clinician who tells patients that they appear depressed will often obtain affirmation regardless of accuracy, either because patients mistakenly accept the clinician's opinion or are hesitant to disagree with a person upon whom they depend. The number of instances that appear to provide confirmation exceeds its actual frequency, a problem compounded by the under-weighting of conflicting evidence. Given the ambiguity of feedback and the clinician's reliance on theories that allow contradictory interpretations of identical outcomes, counterevidence is easily incorporated into prior beliefs. The patient who challenges a conclusion is viewed as "resisting" the truth or "repressing" it from conscious awareness. The result of these clinical practices and mental habits is overconfidence in judgmental abilities (37). In a study on the detection of malingering, most clinicians expressed extreme confidence on a diagnostic task in which error rate ranged from 90 to 100% (15).

Selective attention to supportive evidence similarly affects clinicians' appraisals of their own judgmental accuracy (36). A clinician will inevitably receive some outcome information that appears to support his conclusions. The clinician who tells patients that they appear depressed will often obtain affirmation regardless of accuracy, either because patients mistakenly accept the clinician's opinion or are hesitant to disagree with a person upon whom they depend. The number of instances that appear to provide confirmation exceeds its actual frequency, a problem compounded by the under-weighting of conflicting evidence. Given the ambiguity of feedback and the clinician's reliance on theories that allow contradictory interpretations of identical outcomes, counterevidence is easily incorporated into prior beliefs. The patient who challenges a conclusion is viewed as "resisting" the truth or "repressing" it from conscious awareness. The result of these clinical practices and mental habits is overconfidence in judgmental abilities (37). In a study on the detection of malingering, most clinicians expressed extreme confidence on a diagnostic task in which error rate ranged from 90 to 100% (15).

Self-appraisal of clinical judgment. Overconfidence is one facet of a more general problem appraising one's own judgmental success and decision processes. Research methods that compare subjective impressions to objective measures of data utilization have revealed substantial discrepancies (30). Clinicians may believe that certain variables that actually exerted minimal influence on their conclusions played a key role, and vice versa (38). For example, a clinician's conclusion may be largely determined by potentially biasing information (for example, a prior opinion) which is sincerely thought to have had no influence.

Clinicians commonly propose that their conclusions rest on a careful weighting of many variables, whereas objective analysis typically shows that only a few variables, perhaps two or three, exert a significant impact (39-40). Clinicians also assert that complex confounding analysis or data integration is necessary to reach accurate conclusions—that one never considers datum in isolation but rather the "whole" or overall pattern of results. However, numerous studies suggest that no clinician, or human being for that matter, can begin to manage such complex cognitive operations (41-43). The attempt to grasp interactions among even two or three variables can outstrip human cognitive capacities. Further, clinicians' judgments can usually be reproduced or duplicated by mathematical formulas that simply add variables together and disregard interactions (24, 44).

The expert's misappraisals of his judgmental accuracy and processes create special complications in the courtroom. Courtroom opinions often defy the type of objective verification possible in the sciences. How does one verify a statement like the following: "I knew just what she was thinking when she committed suicide." The judge or jury, lacking both objective data on the particular expert's judgmental success and familiarity with the relevant research, often must rely on indirect, intuitively plausible markers of accuracy: the expert's stated confidence and description of his judgmental processes and powers, and his background training, experience, and credentials. These supposed markers of accuracy are potentially prejudicial. Clinicians misclassify confidence and misappraise their own judgmental processes and success. Training and experience are unrelated to accuracy. The expert, misled by subjective self-appraisal and illusory beliefs, and unshaken by massive negative scientific evidence, attempts to persuade jurors to share the same misplaced faith in false markers. The expert's persuasive effort may well succeed because it aligns so closely with common belief.

Conclusions and Implications

We began by asking whether expert witnesses achieve reasonable certainty and aid the trier of fact. The scientific evidence clearly suggests that clinicians fail to satisfy either legal standard for expertise. Clinicians frequently cannot agree on psychiatric diagnoses of current states, much less provide trustworthy answers to less familiar and more difficult forensic questions, which often demand projections backward or forward in time. Considerable research also shows that clinicians' judgmental accuracy does not surpass that of laypersons. However, actuarial methods may satisfy one of the standards. Although actuarial procedures rarely address questions of direct forensic interest and usually achieve modest results, rather than reasonable certainty, their accuracy does surpass both professionals and laypersons. It is for the courts to decide whether clinicians' failure to meet both standards should exclude them as expert witnesses, and whether satisfaction of the second standard alone is sufficient to admit actuarial conclusions as courtroom evidence.

Should the courts admit actuarial outcomes, research suggests a limited role for experts. A knowledgeable expert can inform the court whether an actuarial procedure is applicable to the particular examinee and question of interest. For example, MMPI indices for malingering may sometimes aid the court, but the MMPI should not be used with individuals of limited intellectual endowment. The expert may also help as needed to explain output statements, which may contain psychological jargon, and can review relevant research on the accuracy of the particular actuarial technique. However, according to available research, the expert's involvement should end in the explanation of the actuarial procedure. The expert's involvement in the interpretation of the clinical data, or attempts to "refine" or modify actuarial conclusions, produce inferior overall results.

Experts who are aware of the negative scientific evidence may assert that the research does not apply to them. Many of the psychologists and psychiatrists who participated in judgment studies probably held the same prior belief, although the research showed otherwise. Clinicians who claim exemption almost always lack objective data on their judgmental accuracy. Given the many studies that raise serious doubt about clinical judgment and the obstacles to valid self-appraisal of judgmental success, the clinician who makes a counterclaim should bear the burden of proof. The validity of counterclaims could be appraised directly. Certifying bodies could conduct objective evaluation of the clinician's performance on a representative sample of cases that can be verified against objective data. There are no definitive means for verifying certain types of clinical judgments, including most diagnoses, but research methods permit objective evaluation of performance on many judgment tasks. For example, clinicians can be asked to predict occupational success and their judgments compared to known outcomes in actual cases.

What of the possible conclusion that the involvement of expert witnesses is not helpful but does no harm? As discussed, expert testimony may exert a prejudicial affect on juries. Confidence and accuracy can be inversely related, and yet the jury may well accept
the opinion of an expert who exudes confidence over that of an opposing expert who expresses appropriate caution. Expert evidence is readily subject to abuse due to its highly subjective nature and vulnerability to biases. The involvement of experts wastes many hours of already too scarce court time and costs taxpayers millions of dollars. Experts also create malpractice risks for colleagues. Each time an expert witness claims he can predict violent behavior with reasonable certainty, he endorses a falsehood. A competent clinician who could not have anticipated his patient’s violent episode may thus be held legally accountable.

As the courts and the public come to realize the immense gap between experts’ claims about their judgmental powers and the scientific findings, the credibility of psychology and psychiatry will suffer accordingly. Psychological research should eventually yield more certain knowledge and methods that provide meaningful assistance to the trier of fact. Ironically, unlike the current situation in which expert testimony is often admitted despite the negative research on its value, the erosion of credibility may reverse this trend. The courts, having learned to distrust clinicians’ claims, may refuse to admit testimony based on truly useful knowledge and methods despite more adequate supportive studies.

REFERENCES AND NOTES

1. This article addresses clinical evaluation and does not necessarily pertain to courtroom experts who do not assess individuals but rather limit their testimony to research findings, such as studies on the reliability of eyewitnesses. In the subsequent text, “clinician” is used to refer to both psychologists and psychiatrists.


38. D. Faust, Prof. Psychol. 17, 420 (1986).


42. D. Faust, Prof. Psychol. 17, 420 (1986).


44. R. M. Dawes, Am. Psychol. 34, 571 (1979).

45. We would like to thank R. Dawes, L. Goldberg, B. Nurcombe, J. Pine, and D. Wedding for their most helpful comments and suggestions.

1988 AAAS Philip Hauge Abelson Prize

To Be Awarded to a Public Servant or Scientist

The AAAS Philip Hauge Abelson Prize of $2,500 and a commemorative plaque, established by the AAAS Board of Directors in 1985, is awarded annually either to:

(a) a public servant, in recognition of sustained exceptional contributions to advanced science, or

(b) a scientist whose career has been distinguished both for scientific achievement and for other notable services to the scientific community.

AAAS members are invited to submit nominations now for the 1988 prize, to be awarded at the 1989 Annual Meeting in San Francisco. Each nomination must be seconded by at least two other AAAS members. The prize recipient will be selected by a seven-member panel appointed by the Board. The recipient’s travel and hotel expenses incurred in attending the award presentation will be reimbursed.

Nominations should be typed and should include the following information: nominee’s name, institutional affiliation and title, address, and brief biographical resume; statement of justification for nomination; and names, identification, and signatures of the three or more AAAS member sponsors.

Eight copies of the complete nomination should be submitted to the AAAS Executive Office, 1333 H Street, N.W., Washington, D.C. 20005, for receipt on or before 1 August 1988.

1 JULY 1988