"If you are going to read only one psychology book this year, make it *House of Cards*."

—*Contemporary Psychology*

HOUSE OF CARDS

Psychology and Psychotherapy Built on Myth

Robyn M. Dawes

Emotional suffering is very real, and the vast majority of people in these expanding professions sincerely wish to help those suffering. But are they really the experts they claim to be? Is our society justified in granting them special status and paying them from common funds? Are they better therapists than minimally trained people who may share their knowledge of behavioral techniques or empathetic understanding of others? Does possessing a license imply that they are using scientifically sound methods in treating people or providing an "expert opinion"? Should their opinions be recognized in our courts as having any more validity than the opinions of anyone else? In particular, are their opinions any better than those of judges, who have been selected on the basis of their legal record to make tough social decisions? Can these mental health practitioners, for example, make a better determination of whether a young child has been sexually abused than can be made by a careful consideration of the evidence without considering their opinions?

These questions have been studied quite extensively, often by psychologists themselves. There is by now an impressive body of research evidence indicating that the answer to these questions is no. Those claiming to be mental health experts—including many psychiatrists—often assert that their "experience" allows them to apply principles of psychology in a better manner than others could, but the research evidence is that a minimally trained person applying these principles automatically does at least as well. Moreover, the research evidence indicates that—unlike a surgeon, for example—mental health practitioners don't develop skills in applying these principles through experience. Often, moreover, they don't even attempt a systematic application of principles, instead claiming to base their practice and judgment on "trained intuition," which presumably allows them to transcend or ignore these principles when they shouldn't. There are "scientifically based" practitioners who attempt to base what they do on these principles, but there is no system of assurance that others will do so as well in these rapidly expanding fields, and they don't. A license has become, unfortunately, a license to ignore the valid principles and generalizations that do in fact exist in the mental health areas (though not in impressive numbers). And when the practitioners ignore valid principles, they can even become outright dangerous to our civil liberties, as when they ignore what they presumably should know about the malleability of human memory or the suggestibility of young children. ("There was no really good evidence. It was the therapists' notes that convinced me she was guilty.")[1]

The purpose of Part One is to share with the reader the research basis for these negative conclusions. I will sometimes describe specific studies, sometimes rely on summaries of sets of studies. These results have very strong implications for public policy in the mental health area. We should not be pouring out resources and money to support high-priced people who do not help others better than those with far less training would, and whose judgments and predictions are actually worse than the simplest statistical conclusion based on "obvious" variables. Instead, we should take seriously the findings that the effectiveness of therapy is unrelated to the training or credentials of the therapist. We should take seriously the findings that the best predictors of future behavior are past behavior and performance on carefully standardized tests, not responses to inkblot tests or impressions gained in interviews, even though no prediction is as good as we might wish it to be. The conclusion is that in attempting to alleviate psychological suffering, we should rely much more than we do on scientifically sound, community-based programs and on "paraprofessionals," who can have extensive contact with those suffering at no greater expense than is currently incurred by paying those claiming to be experts. We might also be better off relying more on ourselves in addressing our own problems.

This section of the book is based on a philosophy enunciated by Paul C. Stern. A major policy goal of psychological and social science should be to "separate common sense from common nonsense and make uncommon sense more common."[2] The common sense that assumes trained people must possess unique skills simply because they claim to have them is common nonsense. In addition, the common-sense attitudes and beliefs that lead us to accept mental health practitioners in particular as experts must be understood as common nonsense. The uncommon sense to understand the issues involved in evaluating claims of expertise and to grasp the meaning of the research addressing these issues should become common sense. It is to this goal of separating sense from nonsense that the first seven chapters of this book are addressed.

to be an "intuitive understanding" of their clients' problems, an understanding they have supposedly gained "from experience." But when they practice on this intuitive basis, they perform at best as well as minimally trained people who lack their credentials (the topic of Chapter 2) and at worst as licensed, expensive (if inadvertent) frauds (the topic of Chapter 5).

The reason I reach these conclusions is that the ability of these professionals to alleviate emotional distress has been subjected to empirical scrutiny—for example, their effectiveness as therapists (Chapter 2), their insight about people (Chapter 3), and the relationship between how well they function and the amount of experience they have had in their field (Chapter 4). Virtually all the research—and this book will reference more than three hundred empirical investigations and summaries of investigations—has found that these professionals' claims to superior intuitive insight, understanding, and skill as therapists are simply invalid. What our society has done, sadly, is to license such people to "do their own thing," while simultaneously justifying that license on the basis of scientific knowledge, which those licensed too often ignore. This would not be too bad if "their own thing" had some validity, but it doesn't. What the license often does is to provide a governmental sanction for nonsense such as:

> *"In my mind, I know what she was thinking and feeling at the time of her death"*—a Harvard professor of psychiatry, quoted in the *New York Times*, October 21, 1987, p. A22. Where his "psychological autopsy" was allowed into testimony at the trial of Teresa Jackson for (psychological) child abuse following the suicide of her daughter in Fort Lauderdale, Florida

or, from a professional talking about incest victims,

> *"It's so common that I'll tell you that within 10 minutes, I can spot it as a person walks in the door, often before they even realize it. There's a trust, a lack of trust, that's the most common issue. There's a way that a person presents themselves. There's a certain body language that says I'm afraid to expose myself. I'm afraid to be hurt."*—Good Morning America's on-air psychologist on the CNBC program *Real Personal*, April 27, 1992 (after maintaining that "Probably one in four women, one in eight men, have been incested.")

If the only result were nonsense, it would not be so bad. There is a lot of benign nonsense in the world. Unfortunately, such nonsense like this can have a profound effect on other people's lives, and it is expensive nonsense.

Claims to intuitive understanding, like those in above quotes, leave potential clients incapable of distinguishing between service that has a true scientific base and service based simply on the claims of those providing the service. The professional associations have exacerbated this confusion by monitoring and sanctioning their members only for the consistency of their practice with their presumed power and status, not for whether that practice does any good or has any scientific justification. Thus, in a recent flap concerning a female Harvard psychiatrist whose client committed suicide, the focus of the professional board's inquiry was on whether she had sexual relations with him—not on whether encouraging him to regress to an infantile state so that she could "reparent" him had any known value for him or anyone else. The write-up in *Newsweek* treated the public to what various well-known psychiatrists and psychologists "said," "thought," or thought they "knew" about the case but nowhere was there reference to any *evidence* concerning the psychiatrist's mode of treatment.[4] The impression is created that psychotherapy treatment is all a matter of opinion or conjecture. It isn't, but many practitioners treat it that way, while the professional associations support them in doing virtually anything at all that appeals to their "clinical intuition," as if there were no knowledge. The professionals are immune so long as they keep their hands off their clients and don't do anything else that would offend their colleagues' sense of status or propriety, such as be arrested for homosexual solicitation in a men's room or plead *nolo contendere* to a charge of child sexual abuse in order to avoid being jailed as a sex offender.

Finally, the mental health professionals who claim expertise without a scientific base have apparently had a profound effect on our culture's beliefs about what constitutes a good life, what types of behavior are desirable, and—most important—how people "should" feel about the world (see Part II). The most pernicious of these beliefs is that adult behavior is determined mainly by childhood experiences, even very subtle ones, and particularly those that enhance or diminish self-esteem. Self-esteem, in turn, is believed to be an important causal variable in behavior, even though the California Task Force on the Importance of Self-Esteem could find no evidence of such a causal effect. Especially, low self-esteem is believed to yield, with unerring consistency, personally or socially destructive behaviors, so that people who wish to change their behavior must experience an elevation of self-esteem *first* (as the result of therapy or an esteem-raising self-help group) and attempt serious change in their lives only later. Again, the evidence for these beliefs is negative. What these beliefs do is discourage people from attempting to craft a decent life for themselves and instead encourage them to do whatever is necessary to *feel* good—about themselves. Sometimes such striving after "mentally healthy" feelings and attitudes simply result in ludicrous behavior (like clutching a teddy bear while proudly proclaiming oneself undoubtedly an incest victim, despite an inability to remember any credible instances). In general, however, this strategy is self-defeating, because it ignores the simple principle that much of our feeling results from what we do rather than causing us to do it.

By contrast, other professionals do base their recommendations on what is known, or on what is believed to be true on the basis of research findings. They do not offer grandiose and false advice to the general public about how to live, think, and feel. The simple reason is that their own scientific knowledge about human distress makes them aware of its limitations, and most of them are responsible enough not to pretend that these limitations do not exist.

THE GROWTH OF PSYCHOLOGY

As the problem of mental distress becomes ever more severe in this country, the magnitude and status of the professions claiming to have a solution also grow. Psychiatry, with its requirement of medical training and its emphasis on prescribing drugs, has approximately doubled in size in the past thirty years. In contrast, psychology has become big business. In this chapter, I will concentrate on the growth and practice of professional psychology, because it has had the biggest impact on the mental health field since the early 1970s, when clinical psychologists were first licensed as mental health experts. Clinical social work has also had a growing impact, somewhat later—as clinical social workers became licensed in many states through the 1980s. The practice of social workers is more akin to that of psychologists than of

Association,[8] salaries of professional psychologists averaged $73,300 in 1989. Those with two to five years of professional experience averaged $54,068; those with five to nine years averaged $67,005; and those with ten or more years averaged $78,685. A survey taken by the Oregon Psychological Association in 1985 (which involved gross receipts rather than net salaries), when I was president of it, indicated roughly comparable figures. The subjects of this survey ranged from those who had been recently licensed to those who had already established a clientele and a reputation; Oregon at the time was experiencing one of the worst recessions in the country. (The Rand McNally listing of cities at the time had ranked the Eugene–Springfield area as the very worst urban area for economic opportunities.[9]) At board meetings of the Oregon Psychological Association, I was informed that the most common charge of established clinical psychologists in the Portland area was $125 per hour.

The cost, power, and affluence of professional clinical psychologists arises not solely from a belief in their expertise and efficacy but also from sheer numbers. The profession has expanded dramatically in the last thirty years. When I joined the APA in 1959, it had approximately 18,000 members, of whom only 2,500 listed specialties in clinical or counseling psychology.[10] When I quit in 1988, there were 68,000 members, approximately 40,000 of whom were in clinical or counseling.[11] By comparison, the American Psychiatric Association had 10,000 members in 1959 and grew to 34,000 in 1989[12]; assuming that a constant proportion of its members are engaged in practice, that is an increase by a factor of 3.4, while the proportion of American Psychological Association members in professional practice grew by a factor of 16. When I joined in 1959, there were no state procedures for licensing psychologists. Today there are licensing procedures and boards in every state and territory in this country and in every province of Canada. There were roughly 45,500 professional licensed psychologists in this country by 1985.[13] Since then, clinical psychology has doubled its numbers every ten years.[14] For comparison, the doubling rate of lawyers is twelve years,[15] of social workers fourteen years, and of psychiatrists twenty years.[16]

We are all paying for these services through insurance premiums and taxes. In most contexts in a market economy, payment for a good or service is based on a belief that it will work in a certain manner; for example, automobiles with antilock brakes and air bags are generally more expensive than those without because the purchaser has a belief that these safety features will work. If they do not work as advertised, or if they are not part of a car purporting to have them, the seller is liable to lawsuits and prosecution for misinterpretation and for misleading or fraudulent advertising. There is, in addition, some consumer protection for goods and services that purchasers cannot be expected to evaluate on their own without highly specialized training or that are offered primarily to those who lack general competence to understand. Thus, medical practitioners are licensed in every state, as are nursing homes.

UNJUSTIFIED GROWTH IN PSYCHOLOGY

The basic service that professional psychology claims to offer is the skilled application of a scientific understanding of human behavior and feelings, particularly as they relate to issues of mental health, and illness; psychotherapy offers unique skills as well. But as a group professional psychologists and other mental health professionals making the same claims do not possess a special expertise that allows them to provide this service. They are no better as psychotherapists than are others of comparable intelligence who are minimally trained (see Chapter 2); they do not have any special abilities in diagnosing mental distress and predicting human behavior, or in evaluating what causes particular people to behave and feel as they do (see Chapter 3); and they do not learn anything from clinical experience with distressed people that cannot be learned by reading textbooks (see Chapter 4). In fact, there is substantial evidence that the simplest statistical models do better than credentialed and experienced professional psychologists at predicting human outcomes. Moreover, the expertise of mental health experts is limited by the accuracy of the techniques they use; the accurate ones are easy to understand and master, while the ones purporting to require specialized training (like the Rorschach Ink Blot Test) are usually invalid. It follows that the licensing of psychologists in particular protects not the public but the profession (see Chapter 5). In fact, the social sanctioning of "clinical" techniques of dubious validity or proven invalidity through licensing them harms the public.

If research shows that the services of professional psychologists and other mental health experts are not what they are believed or

Kelly was elected president of the American Psychological Association in 1956. As time passed, however, he became increasingly concerned that his vision had been abandoned, even as, after a period of ten or so years of steady growth, the profession exploded in numbers. Graduate programs proliferated, all appealing to a board of the APA for "accreditation." In 1971 the APA made a momentous decision. As evidence indicated that training in theory and research were unrelated to effectiveness as a psychotherapist, the association recognized a new degree, the doctorate of psychology without research training. It was abbreviated as a Psy.D., to differentiate it from the Ph.D., which is technically a "doctorate of philosophy" and which for years has implied not only relevant research training but the production of a dissertation that contributes new knowledge to the field of study. The recognition of the Psy.D. was provisional, pending an evaluation of the programs and the people graduating from them.

What happened, however, was rapid expansion. The original program at the University of Illinois no longer exists, but Psy.D. programs sprang up all over the country, and some of them—such as the Los Angeles branch of the California School of Professional Psychology—even obtained state and American Psychological Association accreditation to switch from granting a Psy.D. to granting a Ph.D. The finding that research training and competence were unrelated to effectiveness as a therapist received stronger and stronger research support, so that derogating research-based practice—as opposed to the "art" of psychotherapy—appeared appropriate to the profession. The fact that the research indicated that one's effectiveness as a therapist was unrelated to *any* professional training was ignored, especially when the question of whether to allow greater autonomy and status for the allied profession of psychiatric social work arose. People with Psy.D.'s became equal to those with Ph.D.'s within the profession through a phrase in most state licensing laws that required a Ph.D. from a program accredited by the APA "or equivalent training." The "fight" with the American Medical Association and the American Psychiatric Association to allow psychologists to be primary providers of mental health services was largely successful, perhaps in part by dint of sheer numbers. The original view that recipients of Psy.D.'s and now Ph.D.'s from professional schools were to function primarily as therapists was lost. These

> sands and thousands of practitioners who are peripherally acquainted with the discipline of psychology.[24]

As we will see in Chapter 2, such peripheral acquaintance need not make for poor therapists. When, however, mental health practitioners in psychology present themselves as experts in legal proceedings, or in deciding whether child abuse actually occurred, or as in advising people about how to live, such peripheral acquaintance is a severe problem. The public trusts such experts on the assumption that they are applying valid psychological principles. But when the experts aren't even aware of these principles, their pronouncements are unsupported.

I know of no comparable changes in the quality of training for psychiatrists. Requirements of passing calculus, physics, biology, and organic chemistry for entrance to medical school have remained constant, however, and the first two years at most medical schools retain a highly academic ("scientific") curriculum. As one psychiatrist friend argues, biochemistry may not be nearly as important to psychiatric work as statistics would be, and much of the course work psychiatrists are required to take appears to be little more than drudgery. Such drudgery does, however, assure that the psychiatrists will have the intelligence and perseverance to succeed at these tough tasks. The topics they master may not be directly relevant to practice, but the qualities a person needs to master them may well be. Most of the major higher status graduate schools in psychology also require evidence of these qualities for admission, and their programs are intellectually demanding for students who intend to become professionals as well as for those who intend to enter research. Moreover, these programs emphasize an approach to practice based on what is known scientifically.

Unfortunately, the lower status schools—*as a group*—do not emphasize research, and many of these professional schools select and train mainly on the basis of impressions of students' personal qualities. Graduates can emerge with little scientific training beyond a year's perfunctory course in statistics, centered mainly on how to enter data into a "canned" computer program. The APA has checklist requirements for a program's accreditation, but satisfying such a list is a far cry from providing rigorous training. It is possible to argue that the

A third reason is that a commitment to rationality and scientific knowledge constrains the poorly trained and hence unskeptical professional from making extreme claims. Some of these claims are sheer nonsense—for example, a claim to be able to know what someone was thinking when she committed suicide, or to be able to tell within ten minutes whether someone has been sexually abused as a child on the basis of that person's general demeanor. Such claims are often believed. The Harvard professor who stated that he knew in his mind what Teresa Jackson's daughter was thinking in her mind when she committed suicide was allowed to testify as an expert in the Teresa Jackson child abuse trial, and Teresa Jackson was convicted. Many other mental health professionals claiming expertise are able—free of the constraints that an understanding of the evidence should provide—to testify whether a person was or was not "insane" during the commitment of a crime, not "insane" in the ordinary social sense of the term, which courts are as capable of judging as is anyone else, but in some supposedly scientific sense of the term. Moreover, the claims are often believed by the general public—often to the detriment of all involved. For example, the belief that schizophrenia and autism are due to a "schizophrenogenic" ("or iceberg") mother, who was unwilling to or incapable of providing the afflicted child with the affection required for normal development, has caused untold misery among the families of such disturbed children. How did that belief come about? From "clinical judgment," which was accepted because it is consistent with our "everyday intuitions." (Chapter 2 will detail the even more disturbing example of lobotomy as a "cure" for schizophrenia.)

One particularly pernicious result of the deemphasis on research has been a series of fads in the area of mental health. The most prevalent one as this book is written is an epidemic of diagnosing people as suffering from multiple personality disorder. This condition supposedly results from repressed child sexual abuse, or even from being raised by parents who practiced satanism—although belief in the existence of satanic cults (as opposed to belief in the KKK or the mafia) is based purely on "aided memories" of people in therapy or "support groups." (A well-publicized story of a satanic cult practice in Texas that led to ten murders was later retracted. It was a drug ring.)

To be sure, professional psychologists still *claim* that their practice is research based, whether or not it is in fact. In 1988, when the then-president of the APA was facing a revolt of research-oriented psychologists who threatened to form their own, research-based organization,[29] he said: "Our scientific base is what sets us apart from the social workers, the counselors, and the Gypsies."[30] And later: "The scientists are the jewels in our crown, and they will continue to be. So we're not going to give them up."[31]

CURRENT BUT UNJUSTIFIED SUCCESS

Having separated itself so far from its research base, how did professional psychology survive? One answer is through lobbying state and national governments for money and privilege. A great deal of money has been put into lobbying with positive results, which can be assessed by reading even randomly chosen issues of the APA *Executive Newsbulletin* or the *State Issues Forum*. Also, positive feedback arises from growth itself—just as the initial growth of VHS recordings led people to buy more VHS sets than Beta sets, which stimulated the growth of VHS recordings, and so on—even though the Beta technology may have been superior.[32] Moreover, salaries of professional psychologists are high, at least relative to salaries of research and academic psychologists. Finally, there is an intrinsic appeal to college seniors in doing "real clinical work" with "real people" after years of academic "preparation for life."

All these factors alone, however, cannot account for the successful growth of professional psychology. For example, lobbying pure and simple may have an effect when the numbers of people a lobby group represents is large relative to the individual legislator's entire constituency, or are "single-issue" oriented. But a hundred thousand professional psychologists and allied practitioners do not constitute such a group; they are geographically diffuse and hardly single-issue people.

Other lobbyists succeed by framing their issue in ways compatible with legislators' views of the world.[33] That is exactly what may account for the APA's lobbying success. Acceptance of what "authorities" claim about their own expertise is not a pathological syndrome, except at its extremes. Belief in authorities who really understand human life is therefore natural to us all. Haven't I myself cited Kelly and Sechrest as authorities? Moreover, as we will see in Chapter 7, authorities claim to be able to "explain" the individual life course, an ability that we all believe we have; yet research findings show that neither the authorities nor the rest of us can do this well, which may surprise the readers of the *New York Review of Books* as much as readers of *People* magazine. An observer cannot explain "why" people do what they do, and people themselves are often aware only of their after-the-fact rationalizations; few take careful notes based on "think aloud" ruminations at the time they make major decisions in their lives, and even if they did, many important factors influencing their behavior would not be included.

More specifically, professional psychologists and other mental health professionals employing the same procedures make the same "attribution errors" in their explanations that we all tend to make—overemphasizing the role of personality as opposed to situational factors in influencing the behavior of others, while simultaneously overemphasizing situational factors as opposed to personality in influencing our own behavior. We readily believe that when other people behave in ways we don't like, it's because they are "sick," but that when we behave in ways we don't like, it's due to the lousy home environment in which we were raised. That leads to the final impact of the claims of professional psychologists. They end up agreeing with the rest of us! Such agreement, of course, implies that they agree with each other as well, and then they can cite each other as additional authority figures.

FORENSIC PSYCHOLOGY WITH "PORTRAITS"

Let me illustrate these problems by an analysis of the presidential speech to the American Psychological Association at its 1990 convention.[34] The president who made it is a leading forensic psychologist. Although his presidential address was not delivered until 1990, he was the president who preceded the one who referred to the "jewel in the crown." His speech is a defense of psychological assessment in court settings. He makes many good points at the beginning when he discusses the presentation of results from well-standardized and validated tests. For example, the highest subtest score on an overall IQ test of someone subsequently brain-damaged cannot be used as an index of how well that person functioned prior to the damage. (Perfectly normal people as well will have some subtest scores higher than others, so that their highest score cannot be taken as a measure of their overall IQ, which is assessed by their average score.) He also stresses the importance of the reliabilities and validities (predictive accuracies) of the measures used.

Toward the end of the speech, however, the president pushes his own use of what he terms *psychological portraits* in court settings, by which he means detailed descriptions of a person's psychological functioning. First, he points out that the research evidence evaluating categorical judgments of professional clinicians yields negative results: "Research published much earlier showed that the type of one- or two-word differential diagnosis, characterizations, and predictions then extant were judged to be lacking in validity (Meehl, 1954, 1956, 1957). Reviews of more current studies (Dawes, Faust and Meehl, 1989), including an excellent recent update of the use of one's head instead of formulas (Kleinmuntz, 1990), reaffirm that conclusion."[35] He then dismisses such judgments in favor of what he terms "valid psychological assessment (portrait) findings."[36]

His evidence for their validity is two extreme cases. One is of a twenty-one-year-old woman who scored at the ninety-eighth percentile in aptitude tests and was Phi Beta Kappa in college. She subsequently suffered a serious head injury in an automobile accident and thereafter tested in the mentally defective range (third percentile) on intelligence tests. The other case he cites is of a man whose intellectual abilities were totally unchanged after exposure to neurotoxins in a workplace. (Not surprisingly, everyone involved agreed with the president's expert professional testimony that the first person had suffered extensive impairment as the result of brain damage while the second hadn't.) Using these extreme instances, he goes on to conclude that: "when such assessment is done well, it is patently obvious to all involved (i.e., juries, judges, and attorneys for *both* plaintiff and defense) that what such a psychologist-expert-witness concluded was valid (true) within the reasonable degree of certainty required in such litigation"[37] (italics in the original). Actually, however, he has made the reverse argument: Because it is patently obvious, it must be true. In other words, his claim that his judgment is valid is supported by the fact that everyone agrees with it, that is, by the lack of a need for his judgment—because the same judgment can be made without him. He misses the point that valid expert testimony in a court should be about matters that untrained judges and juries *cannot* evaluate without assistance.

there is little reason to believe that an overall judgment will have much validity when the components on which it is based have none, which in this context even the speaker himself admits. Then what? Are he and his colleagues to return to all the courts in which they have testified, apologize, and request new hearings for those involved? This approach to practice embraces a principle that I find unacceptable: Earn now, learn later—if at all. Compare this philosophy with the philosophy of extreme caution expressed by W. French Anderson, in the context of gene therapy about which a lot more is known already.

But the president presents the "portraits" anyway. Professional psychologists and other mental health experts are often willing to testify, and they have a profound impact on others' lives in the absence of any evidence that what they do is valid. Their supportive evidence is simply hypothesized, while negative evidence that has actually been collected is ignored. This form of reasoning can be termed *arguing from a vacuum*, because what is purported to be true is supported not by direct evidence but by attacking an alternative possibility. As we will see throughout this book, mental health professionals repeatedly argue from a vacuum to justify their practices.

The American people certainly deserve to have professional mental health experts in the court system only *after* evidence of their accuracy has been supplied, *not before*. We should demand more convincing arguments than this APA president presents. Lacking such evidence, he and his colleagues should be thrown out of court. But their licensing has allowed them in court, and the justification for their presence is based on exactly the type of arguments he presents. Unfortunately, these arguments are persuasive to many people, even though careful examination shows them to imply evidence that simply doesn't exist—the vacuum. They appeal to our uncritical intuitions. They sell. It is my hope that this book will convince the reader not to buy them.

BUT THE SCIENCE EXISTS, ELSEWHERE

The fact that the president cannot point to studies supporting his position does not imply that research psychologists have not conducted literally thousands of studies that have led to a "science of psychology." First, much progress in psychological knowledge has been in the areas of physiology, perception, thinking, judgment, behavioral control, and social beliefs, attitudes, and interactions.[41] Achievements are not limited to theoretical understanding but have applied uses as well. Many of these uses are so common that we do not even think of them as involving "social science": aptitude tests and public opinion polling are examples. These advancements may not have uniformly good consequences, but neither do those in other sciences, like nuclear bombs and medical devices that promote overpopulation and prolong life in a vegetative state or one of unrelenting and extreme pain. Aptitude tests that predict success in a racist environment may be used for racist selection to that environment; public opinion polls that accurately reflect voter sentiment may lead politicians to become subservient to that sentiment rather than do what they believe is right or strive to change that sentiment in a desirable direction. Nevertheless, aptitude tests do predict, opinion polls do reflect public sentiment, of course, on a statistical rather than a certain basis.

In psychology, however, knowledge that does cumulate, cumulates slowly. We do know some things about some conditions. We know that phobias and specific anxieties are not simply symptoms of a "deeper" disturbance, and hence that they can be addressed directly through behavioral means without the emergence of new symptoms. We know there is a strong genetic component in schizophrenia[42] and alcoholism.[43] The general wisdom based on actual scientific *studies* is that mild or moderate depression is best treated by a combination of behavioral, cognitive and drug approaches,[44] although evidence is accumulating that cognitive styles of blaming oneself for failure and crediting "luck" for success play a vital role in depression.[45] (We cannot be yet certain, however, that helping people to get over depressive symptoms as quickly as possible will be beneficial in the long term, in part because judgments of what is valuable or beneficial in life cannot be made on the basis of the standard categories for mental health or illness.)

Psychology has also developed a number of effective measurement devices and ways to predict future behavior. These devices are of the type that can be administered without much training, however, and do not require doctoral-level skill to interpret. Moreover, the best predictions are made on the basis of past overt behavior. It's not that people don't change—they do, sometimes profoundly. Rather, no personal skill has yet been developed—or assessment instrument devised—that allows us to predict who will change, when, and how.

Ironically, the reasons for this invalidity may be found in another area of psychological research—in human judgment and decision making. That area investigates people's systematic biases in making judgments and reaching decisions.[46] (These biases are covered in Chapters 3 and 4 of this book.) Such biases are particularly strong *either when judgments are made in the absence of a well-validated scientific theory or when they are evaluated without systematic feedback about how good they are.* Unfortunately both those conditions characterize the art of clinical prediction in professional psychology. These biases lead not only to invalid judgments of the type the APA president claims to be invalid but to an inference that the type of "portrait" judgments he espouses will be invalid as well. One particular problem stems from reliance on retrospective memory in the self-evaluation of judgmental accuracy when no careful records are kept or no scientific comparisons made. Retrospective memory is not only kind but also "makes sense" out of both the past itself,[47] for example, that if one is depressed, one's parents have been aloof, uncaring, and demanding, and our thoughts about past events and predictions made from them, for example, that "I knew it would happen all along"—even when past evidence indicates a contrary prediction was made.[48] Consequently, it systematically distorts the past and our past judgments in a way that makes the course of events appear to have been predictable and our judgments appear to have been good. The research supporting this generalization is so strong that editors of the National Academy of Sciences reports summarizing advances in psychology (see note 41) specifically exclude studies and "evidence" based on retrospective memory as providing nothing of scientific value.

It is precisely the biased judgments associated with lack of a well-validated theory, lack of systematic feedback, and reliance on retrospective memory that leads to what David Faust, himself a clinical psychologist, terms "the delusions of clinical psychology."[49]

STATISTICAL GENERALIZATION IN PSYCHOLOGICAL STUDIES

Some "scientific" studies in the mental health area involve experimental comparisons (see Chapter 2). Others are "correlational" studestablished the trend. "Significant" means simply that we have reason to believe that the trend did not arise on a chance basis. (Ironically, the way we establish "significance" is to assume that only chance variation is operating, then prove that under those circumstances it is unlikely that we would have obtained a result as strong as the one we actually obtained. A phrase like "significant at the 1 in 1000 level" means simply that if the results were due to random processes, we would obtain a result as strong as the one we actually obtained only one time in a thousand.)

A further analysis of scientific studies of mental health practice indicates that there is always more unexplained variation in the results than there is variation that can be explained by the trend we believe the study has supported. Thus, when we assert that there is genetic influence in certain mental disorders, the basis of our assertion is that we can predict these disorders on the basis of genetic factors more reliably than we could on the basis of chance fluctuation. For example, studies find that the incidence of alcoholism is related to the alcoholism of the biological parents of children who have been adopted and is unrelated to the alcoholism of the children's adoptive parents. That does not mean that a child with an alcoholic biological parent, or even two, is more likely than not to become an alcoholic. Far from it—most children of alcoholic biological parents do *not* become alcoholics themselves. The study's conclusions mean simply that there's a trend, involving genetic constitution. Even in the face of well-established statistical trends, a child of alcoholic biological parents is most likely to turn out normal. The same result is found for children of schizophrenic parents, even though they are more likely to become schizophrenic than are people whose parents are not schizophrenic. In this context of variability (technically termed "unexplained variance"), about the closest we can come to making a prediction that a particular person will have a particular disorder is that identical twins of a schizophrenic individual have about a 50 percent chance of suffering from schizophrenia themselves.

The same principle of finding more unexplained variability than variability explained by an established trend also characterizes studies of the alleviation of emotional distress. We just don't know all that much about the causes of emotional distress, which is not to say that if we stick with what we do know, we cannot help people. When, for example, we find a trend that behavioral approaches are better at helping people overcome phobias than psychoanalytic ones, someone who is phobic would be well advised to try behavioral rather than psychoanalytic therapy—even though the differences in success rate are not great, and a particular individual might actually be better off with psychoanalysis. (Knowing which people would be better off with psychoanalysis would require further research—which might or might not yield positive results.) This unexplained variability is preponderant, and it is important to understand its existence. It is, in fact, the basis for a common argument from a vacuum that some psychologists make: "Because the research results indicate a great deal of uncertainty about what to do, my expert judgment can do better in prescribing treatment than these results." This judgment is then claimed to have "arisen from experience," without any evidence that the judgment yields more certainty than careful studies indicate is there. (In fact, such judgments that are opposed to research findings do worse; see Chapter 3.)

The statistical nature of generalizations in psychology and other social sciences can be masked by study conclusions indicating high reliability and significance. Both these factors are, however, a joint function of the effect size and the sample size in a given study.[50] Consider a hypothetical medical finding, based on a study of two samples of 10,000 people, that those who eat bacon at least once every two weeks have twice the rate of some dire consequence than do those who don't eat bacon—and that this finding is "highly significant." First, we don't know whether the two rates themselves are quite high or quite low in proportion to the general population, a result that is absolutely crucial to our decision about whether to enjoy eating bacon. Even if the rates are presented, however, they might be quite inconsequential for our decision making. For example, 8 dire consequences in the bacon-eating group (for a rate of 8/10,000 = .0008) versus 2 in the other group (for a rate of 2/10,000 = .0002) would yield a ratio of 4 to 1, a result of boardline significance. But we might not care much about a rate of .0008. (We have a 1-in-50,000 chance of being seriously injured or killed *every time* we go on an automobile trip.[51]) On the other hand, if the number of dire consequences were 10,000 (for a rate of 1.0) in the bacon-eating group and 2,500 (for a rate of .25) in the other, we might care a great deal about the rates.

It is important to keep in mind that even effects that are proudly proclaimed to be "highly significant" may be slight. It is especially important to keep that in mind when psychologists and their publicists tout results that have serious implications about life and desirable ways of living. One researcher group states that "it is generally accepted that a positive view of the self and positive mood state are necessary for adaptation and for persistence toward goals."[52] This statement is based on results of multiple studies that indicate a positive statistical relationship. But those findings do not imply that no one adapts, persists, or succeeds without being positive; some people may actually succeed because they believe that whatever they are trying to achieve is extremely valuable, or because they believe that putting forth the effort is the "right thing to do" or the only thing to do. (The term *necessary* in the foregoing quote is unfortunate, in my view.) In fact, the authors of this study themselves, in another section of the same paper, discuss the effects of unanticipated success,[53] which would be almost impossible if the former statement were interpreted to mean that feeling positive about an endeavor is absolutely necessary for its success.

Given all the variability in "the science of mental health and illness," responsible assertions must be of a ceteris paribus nature. Despite claims that pervade the popular media, for example, there is absolutely no scientific evidence that feeling good about oneself is a *necessary* condition for engaging in desirable behavior. Nor is there any evidence that feeling bad about oneself is *necessary* for engaging in undesirable behavior. There is a statistical correlation—that's all. But we are not even sure how to interpret that correlation. The behavior might lead to the feelings, or *vice versa*; or the influence behavior has on feelings might explain the whole correlation. Moreover, the correlation could change as people's beliefs and attitudes change, or even as their beliefs about how to interpret the same correlation change.

More important, the uncertainty of knowledge and its application in the mental health area means that *responsible professionals should practice with a cautious, open, and questioning attitude*. "Knowing" within ten minutes from the way a client walks that she was an incest victim as a child can easily lead a psychologist to ask questions that suggest to her that she must have been a victim. This suggestion in turn can lead the client to reinterpret inaccurately recalled instances of benign behavior toward her as indicative of abuse, which can lead her to conclude that abuse occurred when it didn't—perhaps based on a fully reconstructed memory of such abuse. That belief can lead the client to be alienated from her family and to adopt a stance of incompetence in the face of her own "recalled" childhood traumas. The resulting distress reinforces the therapist's conclusion that the client has suffered greatly from her childhood trauma, which she must now "live through"—occasionally with more improbable and bizarre details added—before she can function as an adult. "Authoritative" beliefs and statements about particular individuals are inappropriate and—because they are so often wrong—can be harmful. Those who seek the services of psychologists should be wary of any professionals who do not proceed cautiously.

Experts in court in particular should point out the statistical nature of psychological generalizations rather than paint a "portrait" of an individual or make causal statements about what led to what in an individual's life. Such causal statements are particularly common in civil suits, because the courts often demand proof of psychological harm. Psychological harm may be every bit as devastating as physical harm, but the question is how to establish it. To assess physical harm, we have well-validated theories about how the individual human body works and techniques for establishing malfunction—such as X-rays and blood tests—that transcend the self-report and behavior of victims. In contrast, to assess psychological harm, the evidence consists of the behavior and self-report of the victim, and the intuitive "art" of the examining psychologist or psychiatrist.

Does that mean we cannot assess psychological harm, even if it is considerable? No. In civil suits concerning exposure to harmful substances, we use statistical analyses to establish that the substances "caused" harm. We make an inference *from* the aggregate, such as an increase in the cancer rate of those exposed, *to* the individual, that is, to his or her cancer. In litigation about such exposure no responsible expert witness would claim to be able to tell exactly what led the particular victim's cancer to develop, or why it developed in one area of the body rather than another. Rather, he or she would testify based on these inferences from the aggregate. Similarly, a responsible psychologist or psychiatrist could cite statistical knowledge—saying "in general . . ."—and let the court apply the same rules for determining harm that it does in medical areas where the evidence is statistical rather than concrete.

Unfortunately, as indicated by the forensic psychologist's presidential address, that type of testimony is not what courts accept from professional psychologists, and it is not what professional psychologists present. Instead, a psychologist will most likely present a mythic statement about what caused what for the individual being evaluated, which will often be disputed by a professional psychologist on the other side. If the expert is not very persuasive in court, or if the opposition's expert is more persuasive, then what? Since the judgment is at best dubious in the first place, style of presentation may become of utmost importance.

ADVICE ABOUT HOW TO LIVE

Professional psychologists not only claim to have expertise they don't have, they claim to have insight into how people should think, feel, and behave. For example, the APA president who referred to the "jewel in the crown" stated in the fall of 1988, "We are all teachers. I think my clients see me as primarily a teacher. We have taught the whole culture. We didn't invent the woman's movement, but we have been among its most ardent supporters. I would like to get away from the medical model entirely. Our job is to bring knowledge to the world."[54] We can well ask what it is that professional psychologists have to teach. Without the "medical model," with its presumed expertise, what is left in mental health? The answer might be scientific knowledge. But the practice of professional psychologists is often *not* based on scientific knowledge, in fact flies in the face of such knowledge. What then do psychologists have to teach?

The answer is a belief system. Under the guise of advancing "positive" mental health—which certainly sounds fine and is consequently hard to oppose—the profession of psychology has propounded a simplistic philosophy of life. This philosophy maintains that the purpose of life is to maximize one's mental health, which is dependent wholly on self-esteem. Some psychologists, like Shelley Taylor, have "discovered" that self-esteem is more important even than realism: "Every theory of mental health" she asserts, "considers a positive self-concept to be the cornerstone of a healthy ego."[55] If the point of life is to maximize self-esteem, it follows that a positive self-concept is everything.

explained away, in terms of their illness or their adaptation to the illness.[57] "Mental illness is just like any other illness," like an ulcer, perhaps, but someone labeled as mentally ill is best advised not to tell others, because such "illness" carries a stigma. In fact, a plethora of mental health "experts" have to a large extent convinced the public that whether or not one is emotionally distressed can actually *define* whether or not one is a good person or is leading a desirable life.

Psychologists' contempt for the individual autonomy of their own clients can also be seen in their explanations of the lives of everyone else. They explain social problems in psychological terms, often to the detriment of solving or ameliorating them; it is the poor self-image of the impoverished American child that leads to a lack of academic skills, we are told, not the fact that these skills are taught for roughly half the number of hours per year in American schools than they are taught in a country such as Japan.[58] Mental illness is used to explain the problem of impoverished and homeless Americans, even though the broadest possible definition of mental illness would classify at most only one-third of them as "mentally ill."[59]

One result of the prevalence of these unwarranted assertions and theorizing has been to weaken people's trust in their own autonomy, in their own abilities to deal with the problems they confront in life. Once such belief is weakened, it is necessary to provide people with compensation in some way. One form of compensation is to encourage them to have positive illusions to enhance their self-esteem (see Chapter 9). Without such illusions, people are supposedly so childlike that they could not function in the face of the uncertainties of the world (again, the derogation of autonomy). I will propose at the end of this book (Chapter 10) that we are actually a lot freer than the pronouncements of professional psychologists would have us believe we are, that we can decide what to do in our own lives, and that autonomy grows by exercising it. We do have choice. It is perfectly possible to function without illusions—and it must be kept in mind that the "teachers'" claims involve statistical generalizations about things as they are, not as they could be.

AND FINALLY, DRUGS

Another compensation that mental health professionals provide is drugs—again, to make people "feel good," as if that were a necessary

CHAPTER 2

PSYCHOTHERAPY

The Myth of Expertise

> I had therapy cases I just botched, and yet they got better. Other cases I did great, and yet the patient deteriorated. I wondered what was going on here.
>
> —Lee Sechrest[1]

Psychotherapy works overall in reducing psychologically painful and often debilitating symptoms. The reasons it works are unclear, because entirely different approaches may work equally well for the same problem or set of problems. Recovery is a *base rate* phenomenon. That is, in predicting the likelihood that a particular individual will recover, we can do little better than by predicting from the overall rate of recovery; we have no insight into exactly why some people get better while others don't. We do, however, know something about psychotherapist characteristics that make it work. Therapists in verbally oriented therapies, we know, should be "empathetic," while those using primarily behavioral techniques should have some knowledge of behavioral principles.

We also know that *the credentials and experience of the psychotherapists are unrelated to patient outcomes*, based on well over five hundred scientific studies of psychotherapy outcome. In fact, it is partly because psychotherapy in its multitude of forms is generally effective that I am writing this book. Having it more generally available is socially desirable.

THE NEED FOR SCIENTIFIC STUDIES TO EVALUATE EFFICACY

The scientific evaluation of psychotherapy is a fairly recent activity. The profession's own resistance to evaluating itself stemmed partly from its psychoanalytic origins. Freud's basic idea was that distressing psychological symptoms result from "the return of the repressed" in a debilitating form (*not* from repression per se). Adults defend against unacceptable needs, wishes, and feelings—which in childhood may have been quite conscious—and keep them from consciousness by means of "defense mechanisms," which are themselves unconscious, specifically, an unconscious part of the "ego." If during childhood these defense mechanisms are not well constructed, or if certain experiences such as sexual seduction by an adult or persistent fantasies about it make it particularly difficult for the developing person to prevent these needs, thoughts, and wishes from impacting him or her, they may express themselves as psychiatric symptoms. For example, Freud's patient "Dora's" coughing fits were thought to express both her wish to engage in oral sex with her father (perhaps after observing him engaging in that activity with his mistress) and her revulsion at the wish.[2] (Freud, ever the Victorian moralist, believed that the father's engaging in oral sex provided conclusive proof that he was impotent, which supported Freud's further conclusion that some hereditary constitutional factors were involved in the development of neurotic symptoms.)

Only through prolonged psychoanalytic sessions leading to a "transference" to the therapist of the patient's childhood reactions to parents and other significant adults can the defense mechanisms, and the impulses they are attempting to keep from consciousness, be understood. The therapist is extremely passive during psychoanalytic sessions, both to encourage this transference and to avoid premature "interpretations" of either the defenses or the needs; premature interpretations would lead to a "resistance" from the patient that would discourage rather than encourage insight. With eventual insight comes an ability to "sublimate" the unacceptable impulses in socially constructive ways (sublimation itself being a type of defense mechanism). Such sublimation can occur without the help of a therapist or psychoanalyst; for example, when the modern psychiatrist George Vaillant[3] discussed the men whose lives he followed in terms of the "maturity" of their defense mechanisms, sublimation was considered the pinnacle. Unlike Freud, Vaillant concluded that defense mechanisms evolve throughout adult life as well as childhood, and that final maturity cannot be well predicted from earlier life. An example the impotence went away at all, they said, as long as the transference was successful. Only a truly expert therapist could evaluate whether it had been successful.

It is indeed very difficult to "get inside" someone else to determine unambiguously whether that person has benefited from any form of psychotherapy. But eliminating a symptom that is of crucial concern to the client really does matter. I agree with Hans Strupp that "a global judgment of [psychotherapy] outcome, which is analogous to a still photograph of an object in motion, must always remain exceedingly difficult and elusive" and that outcome judgments are "contingent on *values* placed on human behavior." But it is exactly a value-laden outcome for which the client (or insurance company, or government) is paying. Procedures—whether they are medical techniques or social programs—must be evaluated on the basis of certain indicators like blood pressure or infant mortality rate that nonetheless do not in and of themselves tell us the "whole story." A perfect one-to-one relationship between observable indicators and the global process in which we are interested is difficult to establish, but not having one available should not be used as an excuse to avoid evaluating what is happening by assessing these indicators. As Eugene Meehan points out, these indicators must be chosen wisely; infant mortality rates cannot be used to evaluate the quality of nursing home care to use his example.[5] But we must insist that a procedure such as psychotherapy be assessed in a way that allows the outside observer to reach a conclusion about its effectiveness. In psychotherapy, symptom remission is a prime candidate for such an indicator, as it is in medicine. What medical doctor would proclaim a patient "cured" without relief of painful symptoms? In fact, after my conversation in the coffee shop, the studies that have been conducted used symptom remission as the primary criterion of cure or improvement.

These studies are very important simply by virtue of the fact that they involve outside observers evaluating the efficacy of psychotherapy. The philosophy that only the individual therapist can tell whether improvement in a client has occurred is flawed. In the first place, the unsystematic judgment of a therapist is as subject to bias as is the unsystematic judgment of anyone else, including clients. The problems of unsystematic judgment are well documented, especially those of retrospective memory. Therapists, thoroughly committed to a profession—and perhaps to a particular technique—may well be "the last to know" when their efforts are ineffective. In the second place, evaluation without involving the patient's own feelings and behavior (that is, symptoms) ignores lessons from the history of mental health treatment. Mental patients were long treated in a cruel and unusual manner when practitioners disregarded their protestations that they did not want a "treatment" thrust on them; only later did the practitioners conclude in hindsight that many of these treatments were indeed poor or cruel. "Treatment" has included the use of chains, scalding baths, lobotomies, insulin shock, and now narcoleptics that can lead to tardive dyskinesia. Given this history, mental health professionals should be extremely careful before deciding that a treatment is "effective." The current approach is, as we know, much more enlightened—just as every previous generation knew that its approach was much more enlightened than previous ones. This conclusion is valid only when based on evidence.

It is even possible that treatments are biased by two attitudes that psychologists themselves hold but don't want to admit even to themselves: They don't like what the emotionally disturbed do and want to distance themselves from them. These attitudes could well bias psychologists' unsystematic evaluations of cures. Mental health professionals must reach their conclusions with extreme care, in a way that would convince a skeptic (again, a criterion implicit in almost all demonstrations that we term "scientific"). They owe that extreme care to people who seek professional help for themselves and alleviation of their problems, who trust those claiming to have expertise in mental health.

The question, then, is how to reach legitimate conclusions about the effectiveness of therapy. A number of deficient ways have certainly led to bad conclusions. One method that at first glance does not seem deficient is to search among therapy clients for examples of "success" and argue for the efficacy of therapy on the basis of this success. The problem with this method is that some people who experience distress will get over it whether they are in treatment or not. Having improved, these people will search for a reason for the improvement, as will their therapists. Not surprisingly, both client and therapist may well agree that the reason was psychotherapy. There is no way of evaluating that conclusion, because the client cannot compare what did happen with what would have happened if therapy had been unavailable. This simple objection illustrates an important logical fallacy. We do something in a particular situation, and something else follows. Was that something else caused or even influenced by what we did first? If we conclude that it was, we must have in mind a *hypothetical counterfactual*, or some idea of what would have happened if we had not done what we did. It is hypothetical because we can never be certain "what would have happened if." Nevertheless, we can evaluate the effect of what we did do only by comparing it with what we believe would have happened if we hadn't done it.

In most of our everyday functioning, we don't use hypothetical counterfactuals to confirm simple beliefs that we consider self-evident. Most of us believe that we know where our home is located, for example, because when we go there, it is there. We never check our belief by going to some other location—perhaps randomly chosen—when we wish to go home to verify that our home *isn't* at this other location. This absence of a check illustrates a "confirmation bias" in our everyday beliefs, a bias that generally serves us quite well. Most emotional disturbances are not fatal, and as our life situations change, we develop new ways of coping and thinking about them; our feelings of distress or happiness also vary. Where is the hypothetical counterfactual in the examples of "successful therapy"? Nowhere. Both therapists and clients nonetheless often cite such successes as "proof" of the effectiveness of therapy.[6]

Despite their uselessness, however, instances of success continue to be cited as "evidence" for the effectiveness of therapy, even by psychologists. For example, a writer maintains in *American Psychologist*: "Suppose you test artistic ability before and after therapy. Should you predict a difference between treatment and control [that is, no-treatment] groups? Not at all! Predict that your measure will increase only for the successful subgroup. After all, you do not want to predict an increase for the failure cases."[7] I leave it to the reader to figure out the validity of "predicting" success only for those cases later found to be successful.

A more common way to establish a claim for therapeutic effectiveness is to treat a group of people and find that in general they are better off after treatment than they were before. Here, the hypothetical counterfactual is that they would have remained the same without therapy. But there are serious flaws with this. For starters, we don't know that they would have remained the same. A much more subtle flaw is technically termed a *regression effect*. That is, processes appear to "regress" from less likely states to more likely ones simply because the more likely ones are likely to occur at later points in time. For example, people are not often extremely happy (or extremely unhappy). It follows that when they are, they are less likely to be as extremely unhappy (or happy) later—no matter what happens in the meantime. Because most people enter therapy when they are extremely unhappy, they are less likely to be as unhappy later, independent of the effects of therapy itself. Hence, this "regression effect" can create the illusion that the therapy has helped to alleviate their unhappiness, whether it has or not. In fact, even if the therapy has been downright harmful, people are less likely to be as unhappy later as when they entered it.[8]

To understand regression effects in general, suppose we toss a coin twice, and it falls heads both times. We toss it two more times. We expect fewer heads on these second two tosses; specifically, we expect that the probability that we will get two heads again is only 1 in 4; we will get one head and one tail (in either order) with a probability of 1 in 2, and two tails with a probability of 1 in 4; thus, our expected number of heads on the second two tosses is only one. Does that mean that coins "catch up to themselves" (a belief termed "the gamblers' fallacy")? No. It simply means that when we get an unusual result one time in a random process, we are unlikely to repeat it. This regression to 50 percent heads is, of course, probabilistic, because the probability is only 1 in 4 that the two heads will be repeated in the second two trials, but they *can* be. Similarly, if a fair coin has landed heads 9 times in 10 trials, the probability that there will be fewer than 9 heads in the next 10 trials is 99 in 100, but there is still a chance of roughly 1 in 1000 that it will land heads all 10 times in this subsequent set of 10 trials.

I am not claiming that life is a toss of a coin (any more than it is a river). The point of the example is that when there is any random component whatsoever, and we pick a group on the basis of being unusual in some way or other, we get a regression effect. It is, moreover, not even necessary to hypothesize a random component in what is being observed. All that is necessary is that the variables studied are not perfectly correlated. Not everyone realizes that illusions can result from regression effects. The best way to receive an award for "noted improvement" in academic work in some grade and high schools is to do terribly the previous semester; for example, an Israeli fight instructor has protested to psychologist trainers[9] that he "knows" punishment works better than reward, because: "I've often praised people warmly for beautifully executed maneuvers, and the next time they almost always do worse. And I've screamed at people for badly executed maneuvers, and by and large, the next time they improve."

The direct relevance of regression effects to evaluating psychotherapy is that people often enter therapy at times when they are particularly unhappy and distressed. But if their problem is one that varies over time rather than having a consistently downward course, regression effects alone could result in "improvement"—and an illusion that the improvement is due to psychotherapy: "If treated, a cold will go away in seven days, whereas if left alone, it will last a week." Emotional distress is certainly more serious than a cold, but even serious emotional distress will vary over time. Since such variability implies an imperfect relationship between outcomes at two different points in time, regression effects are to be expected. In fact, they occur even within a condition: "Of particular significance was the fact that those scoring highest on symptom reduction after SD were those whose symptoms were initially more severe, and who were less promising candidates for conventional types of therapy."[10] Of course.

The best way to evaluate the efficacy of therapy, however, is to compare a group of people who receive therapy with a group who don't. That is to say, as in any such scientific experiment, there must be an *experimental* group and a *control* group. The two groups must be equivalent when they begin therapy, moreover, so the comparison cannot be between people who seek out therapy and others who don't seek it out but who all have the same symptoms. People who seek out therapy will likely be more motivated to get over their problems than those who don't, which will skew the results. Even a highly sophisticated statistical control cannot establish equivalence on this most given the treatment, and a control group, which is not; the outcomes for these two groups are then compared. Such random assignment of individuals to groups does not guarantee that the two groups will *not* differ in ways relevant to the outcome. It merely creates a statistical expectation that the two groups will not differ. Larger sample sizes produce more likelihood that the experimental and control groups will be alike. This approach is called a *randomized experiment* in the social science literature and a *randomized trials experiment* in the medical literature, where it is most commonly used. What happens is that the control group provides the hypothetical counterfactual against which the outcome for the experimental group can be compared. The logic is explained well in Sinclair Lewis's novel *Arrowsmith*; the most widely publicized randomized trials experiment was that on the Salk polio vaccine in 1954.[13]

In psychotherapy, randomized experiments often involve randomly selecting people for a control group and promising them that they will receive treatment after a specified period of time—a "wait list control." Classical medical randomized control experiments do not do this. Another difference is that many medical experiments involve a *placebo control*, in which subjects in the control group are given a placebo and neither group is told whether they are the experimental or the control group. Such experiments are termed *double-blind* experiments because both the people in them and the people evaluating them are "blind" to whether they are in the experimental group or in the control group.

It is hard to develop a double-blind experiment in psychotherapy. Both the subjects and those examining the subjects are generally aware of whether they have received therapy. In psychotherapy, moreover, many of the criteria used to assess the success of treatment rely on the self-report of the subject, for the simple reason that much treatment is aimed to alleviate the emotional distress that the subject has experienced. Such self-reports could easily be biased by subjects' knowledge that they had or had not received psychotherapy. For this reason self-reports are rarely used without considering other outcome criteria as well.

Such randomized experiments are very necessary in evaluating treatments for emotional disorders. Studies that did not conform to the principles of randomized experiments have had dubious results. In one study the investigator concluded that of 136 people given a ject to scrutiny by Harry Bakwin.[20] He found that although a majority (61 percent) of children in the New York School System in the 1930s and early 1940s had had their tonsils removed, there was no correlation whatsoever between the estimate of one physician and that of another regarding the advisability of tonsillectomy when a sample of the remaining 39 percent of the children were examined,[21] of whom 45 percent were said to be in need of having their tonsils removed (as were 46 percent in a later screening of those children who passed *this* screening). Eighty children had died each year as a result of the anesthesia administered for tonsillectomies. Findings such as these have led medicine, especially medical school professors, to appreciate the importance of systematically checking clinical judgment, for example through employing randomized clinical trials. Mental health professionals would be well advised to follow medicine's example.

STUDIES THAT EVALUATE EFFICACY

Randomized experiments evaluating the efficacy of psychotherapy began appearing occasionally in the scientific journals during the 1960s. One impetus for them came from psychologists' increased use of behavioral techniques, in which specific behaviors were targeted for change through the use of reinforcement principles. Since the whole point was to change these behaviors, the efficacy of the techniques was easily evaluated, and randomly selected (usually wait-list) control groups could be easily evaluated as well. As professional psychologists proliferated—and their fees soared—the "only the therapist knows" philosophy became increasingly difficult to maintain.

In 1977, Mary L. Smith and Gene V. Glass published a famous article in *American Psychologist* that concluded that psychotherapy is very effective. They summarized the results of 375 studies of psychotherapy effectiveness that had purported to use random assignment to experimental and control groups.[22] The summary technique they used, termed *meta-analysis*, first determined the average difference in each study between the experimental and control groups on some outcome variable that the therapy attempted to address (like behavior, self-report of anxiety or depression, or assessment of psychological functioning by "blind" observers). These differences were measured in terms such as subjects' well-being or reduction of symptoms. Each difference was assessed after therapy had ended for the people in the would like to restrict practice to those who are licensed. In the years after the Smith and Glass article was published, many attempts were made to disprove their finding that the training, credentials, and experience of therapists are irrelevant. These attempts failed. The abstract of a review by Jeffrey S. Berman and Nicholas C. Norton summarized such results:

> [A recent review] concluded that patients treated by paraprofessionals [people minimally trained] improved more than those treated by professionals. However, this provocative conclusion is based on inappropriate studies and statistical analyses. The present review omitted problematic studies and organized the data to permit valid statistical inference. Unlike [earlier authors listed] we found that professional and paraprofessional therapists were generally equal in effectiveness. Our analyses also indicated that professionals may be better for brief treatments and older patients, but these differences were slight. Current research evidence does not indicate that paraprofessionals are more effective, but neither does it reveal any substantial superiority for the professionally trained therapist.[28]

In other words, the professionals are no different from the paraprofessionals in the effectiveness of their treatment. Furthermore, consistent with earlier summaries of studies they and other authors had examined:

> In a first set of analyses, we examined whether the relative effectiveness of professionals and paraprofessionals might vary for different types of problems and treatments. When we classified studies according to the four most commonly occurring categories of patient complaint (social adjustment, phobia, psychosis and obesity), we found no reliable differences [between professionals and paraprofessionals] among the separate effect sizes. . . . We also failed to detect any systematic differences when we divided the studies into five forms of treatment (behavioral, cognitive-behavioral, humanistic, crisis intervention, and undifferentiated counseling).

And:

> Similarly, there were no statistically significant differences [again, between professionals and paraprofessionals, *not* between experi-

mental and control groups] between four different sources of outcome (patient, therapist, independent observer, and behavioral indicator).

Perhaps the most famous study supporting this conclusion was performed by Hans Strupp and Suzanne Hadley.[29] They recruited as therapists university professors who had no background in psychology and randomly assigned clients either to them or to professionally trained and credentialed psychologists. In all, they assigned fifteen clients to the professionals and fifteen to the professors. The clients were those whose problems, as Strupp and Hadley put it, "would be classified as neurotic depression or anxiety reactions. Obsessional trends and borderline personalities were common." The professionals charged higher fees, but they were no more effective as therapists than the professors. The only slight difference was that after therapy the clients of the professionals tended to be a bit more optimistic about life than those of the untrained professors, but they didn't function any better on any of the multiple measures the investigators evaluated. While this difference may result from the current professional belief that optimism is an important criterion in mental health (perhaps *the* criterion, see Chapter 9), it could also have arisen on a chance basis.

Other reviews indicate that the level of experience of professional therapists is unrelated to their efficacy.[30] Consistent with such "it doesn't matter" findings, William Miller and Reid Hester published a highly influential review indicating that the intensity of professional treatment does not matter even for people with the problem of alcoholism.[31] Miller and Hester summarized all the studies in which alcoholics were randomly assigned to inpatient or outpatient treatment. Some of the inpatient programs involved prolonged stays in institutions devoted to radical changes in lifestyle, beliefs, and attitudes. But there were no differences in outcomes between inpatients and outpatients. Nor did Miller and Hester find any relationship between the length of treatment and outcome. In fact, nothing worked better for alcoholics than a minimal treatment involving detoxification and one hour of counseling!

This result contradicts results of studies or other types of therapy, in which a "dose-effect" relationship between length of psychotherapy and outcome has been established; approximately 50 percent of

CHAPTER 3

PREDICTION AND DIAGNOSIS

More Myths of Expertise

> There is no controversy in social science which shows such a large body of qualitatively diverse studies coming out so uniformly in the same direction as this one. When you are pushing 90 investigations [as of 1991 closer to 140], predicting everything from the outcomes of football games to the diagnosis of liver disease and when you can hardly come up with a half dozen studies showing even a weak tendency in favor of the clinician, it is time to draw a practical conclusion.
>
> —Paul E. Meehl[1]

Much of the success of verbal therapy is influenced by the personal qualities of therapists and how they relate to clients. Much of the success of behavioral psychotherapy is influenced by therapists' understanding of the basic principles of behavior change, which are not too difficult to grasp. Much of the success of all therapy may be influenced by the fact that the client is taking action and no longer feels helpless in the face of disruptive emotional pain. Clear findings that psychotherapy works in general and that the training, credentials, and experience of the therapist are irrelevant to its success give rise to these speculations.

Nevertheless, a well-trained and experienced professional psychologist or similar professional may better understand what people—particularly distressed ones—are like, why particular individuals act and feel as they do, and how to diagnose individual problems, however much of a hodgepodge the resulting classification system may be. If so, then the professionalization of the mental health field, the fees, its status, and its public acceptance may all be justified.

Professional psychologists in particular behave as if they understand. Thirty-five percent of them appear in court.[2] Many have hospital admissions privileges, including involuntary hospitalization. They are deeply involved in diagnosing people in mental health facilities and in their offices, and they are often seen in the media providing explanations of why someone (from the latest serial killer to Saddam Hussein) did this or that, or offering advice to listeners about what to do and how to feel, and above all, when to seek psychotherapy. Moreover, they are well remunerated for such services.

The claim is that professional training yields understanding, not just about people in general but about the single individual in all her or his uniqueness. Statistical generalizations can be found in textbooks but an understanding of single individuals in all their complexity cannot.

UNDERSTANDING AND PREDICTION

To evaluate this claim, we must first decide what it means to understand another individual. Certainly, it means more than creating a "good story" about why particular people do what they do and feel as they feel, or about why this or that happened or is likely to happen. Good stories may be psychologically compelling, but they are not necessarily valid. Going beyond the good-story criterion of understanding requires some knowledge of the world and its workings. How do we obtain that knowledge? This complex philosophical question can be transformed into a slightly simpler one: How do we know that we know?

The question of establishing the validity of our knowledge may appear equally complex, but answering it does allow us to establish criteria for knowledge. If we truly know something, these criteria must be satisfied. The criterion with which this chapter will be concerned is the *ability to predict*; that is, we know something is the case if we can predict that in given situations it will hold true. Predictability is not synonymous with knowledge—a horribly complex, ad hoc system that could predict would not involve as much knowledge as a theoretically justified simpler one that didn't predict quite so well.[3] Nor is predictability the only criterion we use to assess whether we have knowledge. Such an aesthetic notion as "beauty in one's equations" may be a criterion. (See Dirac's discussion of Schrödinger's wave equations.[4]) But a crucial test of understanding is the ability or lack of ability to predict. Prediction need not be perfect, but it is found in all branches of science. Even sciences not involved with the future, such as paleontology, make predictions of what will be discovered when certain evidence is examined, such as new fossil evidence. Making sense of evidence already gathered is an extremely important activity in science, but it alone is not enough.

In the last chapter I discussed the crucial role of randomly controlled experiments in evaluating a treatment or therapy. A person who claims that a treatment is effective must demonstrate that it has an effect in comparison to a hypothetical counterfactual, obtained through construction of a randomly constituted control group. This "show me" criterion may have been extremely important to the development of scientific demonstration in Western civilization, beginning with the Renaissance rejection that an assertion about the universe or people could best be proved by referring to the Bible or to Aristotle. "Show me" is basically translated into "show me what will happen next" when we require prediction.

The demand to "show me" can also be quite subtle. In the late 1840s for example, Dr. Ignaz Phillipp Semmelweis noted that the rate of death from "childbed fever" among mothers who had given birth in a ward serviced by physicians was almost four times as high as mothers in a ward in the same hospital serviced by midwives.[5] The deaths tended to occur in women in the same rows of beds. Semmelweis wondered whether the reason was that they were attended by the same doctor. The doctors didn't clean their hands, even after returning from dissecting a cadaver in the morgue, because such practice was considered to be unmanly. Or perhaps the effect was psychological, since after a priest administered last rites to a dying patient, he went down the line of beds ringing the "death bell." At Semmelweis's request, the priest stopped ringing the death bell in the hospital, but the mothers continued to die in rows. Semmelweis then demanded that his colleagues and assistants wash their hands in a solution of chlorine of lime before they examined a woman or delivered her baby. Over the next fifteen months, the death rate fell from 12 percent to 1.2 percent. After participating in a republican street demonstration in 1848, however, Semmelweis was fired from his hospital post. His successor stopped the silly requirement of hand washing, and the death rate rose to 15 percent. We would be more certain that the changes in death rate were due to hand washing if he had required the doctors to wash their hands in some randomly picked rows but not in others. Semmelweis happened nevertheless to be correct, and he tested it in a way that allowed him to present evidence to a person who demanded "show me." Unfortunately, the medical people at the time were not as impressed as we believe in retrospect that they should have been. The old practices were retained until the 1880s, when Dr. Joseph Lister understood the importance of Semmelweis's experiments. In the meantime, Semmelweis had lost his sanity, begun accosting people on the streets to warn them to stay away from doctors who didn't clean their hands, and died in a mental institution in 1865.

A happier example of the "show me" approach is provided by R. W. Wood, a professor of physics at Johns Hopkins University and "an inveterate perpetrator of pranks and hoaxes."[6] After Wilhelm Roentgen's 1895 discovery of X-rays—which could be broken down into alpha, beta, and gamma rays—physicists were eager to discover other sorts of radiation as well. In 1903 one of the most distinguished physicists in France, René Blondlot, announced that he and his laboratory colleagues had discovered a new type, which he labeled an N-ray in honor of the University of Nancy, where he was a professor. These rays, he announced, were emitted from the sun; others found that they were emitted from the human body as well. Wood and others were unable to see the N-rays when they attempted to duplicate Blondlot's experiments, and eventually Wood visited the laboratory at the University of Nancy, presumably to find out what he was doing wrong. When he was shown various ways N-rays could be created, he couldn't see them, although his hosts could. The N-rays had to be generated in the dark; one device involved passing light through a prism. After one demonstration Wood—the prankster—surreptitiously removed the prism from the device and asked for a repeat demonstration. It was repeated. He again failed to see the N-rays, but his hosts again did. So much for N-rays.

The Semmelweis and Wood stories illustrate an important point: The test of a claim requires a *comparative* demonstration. Death rates should decrease when the doctors clean their hands; the scientists

SAMMEN LIGNENDE

in isolation—hence, in all her or his "complexity"—rather than as a member of an aggregate.

This approach was recently summarized in a 1992 *American Psychologist* article.[9] The authors say that experts' knowledge consists of experience and practice that "involves accommodating previous understanding to the uniqueness of a particular clinical situation." This accommodation is *not*, however, explicit—that is, "a compilation of independent facts or sets of rules. Rather, it is a dynamic and contextualized understanding that is the result of the interaction of cognitive patterns or meaning gestalts with environmental cues." The authors justify this approach by referencing work on medical and chess expertise. In its extreme and less responsible form, this approach is expressed as "on the basis of my experience, I just know" (see Chapter 1).

The problem with this argument is that it *begins* by assuming that practicing clinicians have an expertise similar to that of medical diagnosticians and chess grandmasters, rather than by establishing this similarity empirically. But this similarity is not at all self-evident. Medical diagnosticians use a great deal of explicit knowledge—resulting from diagnostic tests—to "build intuitive expertise," and most chess grandmasters have studied roughly fifty thousand chess games.[10] We define "expertise" in terms of what experts accomplish, not in terms of how they go about their task. How well do these mental health experts do in comparison to actuarial predictions? Since understanding is not equivalent to prediction but necessarily implies it, we can ask whether professional psychologists make predictions that are better than predictions based on statistical models not involving professionals. A related question is whether the clinical approach is superior to the actuarial approach in a number of fields—including medicine, business, criminology, accounting, livestock judging, and so on. These questions have been extensively studied by psychologists themselves. The answer to them is no.

ACTUARIAL VERSUS CLINICAL PREDICTION: THE RESULTS

The first comprehensive review of whether statistical prediction or clinical prediction is superior appeared in 1954 in Paul Meehl's *Clinical Versus Statistical Prediction: A Theoretical Analysis and Review* *of the Literature*.[11] Meehl reviewed approximately twenty studies that compared the two methods for predicting such outcomes as academic success, response to electroshock therapy, and criminal recidivism. In no comparison was the clinical prediction superior to the statistical prediction. In predicting academic performance, for example, a simple linear weighting of high school rank and aptitude test scores outperformed the judgments of admissions officers in several colleges. In predicting the success of electroshock therapy, a weighting of marital status, length of psychotic distress, and a rating of the patient's "insight" into his or her condition outperformed one hospital's medical and psychological staff members. In predicting criminal recidivism in several settings, past criminal and prison record outperformed expert criminologists.

TRBAKE-
FALL

Meehl was concerned primarily with the statistical versus clinical methods for integrating information; thus, he primarily compared instances in which both types of prediction had been made on the basis of exactly the same data. (He also insisted that the accuracy of the statistical model not be checked on the same data on which it was derived—or that the sample size be so large that it not appear superior due to chance fluctuations.) Twelve years later, Jack Sawyer published a review of about forty-five studies; again, in none was clinical prediction superior.[12] Unlike Meehl, Sawyer also emphasized studies in which the clinician had access to more information than that used in the statistical model—such as studies that included interviews of people about whom the predictions were made conducted by experts who had access to the statistical model information prior to the interview. But such interviews didn't improve the clinical predictions. In fact, the predictions were better when the opinions of the interviewers were ignored. Moreover, in the few studies where the professional clinicians were given the actuarial predictions and were asked to "improve" on them, they did worse than the actuarial predictions; that is, prediction was better if their "improvements" were ignored. Sawyer concluded that even if some inputs of the clinicians were found to be valid, they should be incorporated within a statistical model, along with the other predictors, in what he termed a "mechanical" way. After Sawyer's review, similar evidence continued to mount. That led Paul Meehl in 1988 to reach the conclusion quoted at the head of this chapter. Later, he, David Faust, and I summarized even more studies in an invited article for *Science*, which was published in March 1989.[13]

A topic covered since the publication of Meehl's book is the prediction of whether the final diagnosis for an inpatient in the Minnesota hospital system will be one of "psychosis" or "neurosis." A patient diagnosed as psychotic is one who has lost touch with external reality (as in schizophrenia); a patient diagnosed as neurotic is one who is in touch with external reality but suffers from possibly immobilizing internal emotional distress.

Upon entering a Minnesota hospital, each patient filled out the Minnesota Multiphasic Personality Inventory (MMPI), a test consisting of 567 items with which the patient must agree or disagree. Some of the items have clearly psychological content—for example, "at times I think I am no good at all" and "my sex life is satisfactory" and "at times I have been so entertained by the cleverness of some criminals that I have hoped they would get away with it." Others have no apparent psychological implications, like "I like mechanics magazines" and "I believe in law enforcement." The items were chosen on the basis that their answers would differentiate between patients with a clear problem and "normal" people (often, unhappily, chosen from people visiting the patients, such as relatives). Thus, depressed people are more likely to answer "yes" to the statement about feeling worthless than are normal people. It also turns out that paranoid people are more likely than others to like mechanics magazines. The patient's answers to all the questions lead to an MMPI "profile" (to be distinguished from "portrait") of ten scores, each of which indicates the degree to which the patient's answers are consistent with one of ten different types of pathology. Constructed in the late 1940s, the MMPI quickly became—and has continued to be—the most widely used test given to psychiatric patients and others for "screening" purposes, both in and outside mental institutions.[14] Numerous studies have been conducted to establish the statistical relationship between the resulting profiles (or profile "types") and various types of psychological problems and personality and behavioral disorders.[15] In addition, the "clinical art" of profile analysis has been practiced and taught by professional psychologists.

In the early 1960s, Lewis Goldberg obtained access to the results of more than a thousand MMPI tests that had been given to patients in several Minnesota mental hospitals upon admission. Goldberg also had access to their final diagnosis as neurotic or psychotic. He developed a simple statistical formula based on the ten MMPI scores to predict this final diagnostic categorization. The formula, applicable to all the patients in all the hospitals, was roughly 70 percent accurate when applied to equal-size groups. Goldberg then presented sets of these profiles to professional psychologists with varying credentials and experience and asked them to judge whether each patient would have been diagnosed neurotic or psychotic. These people ranged from graduate students in clinical psychology to experienced professionals with a reputation for being expert in MMPI profile interpretation. None of them could surpass the 70 percent accuracy mark; occasionally, some did on some samples, but they could not repeat their superior performance on other samples. In one study, Goldberg and Len Rorer even presented professionals with the results of the statistical formula to help them in their judgment, but they did worse than the formula itself.[16]

Goldberg's studies, in which the statistical formula and the clinical judgment were based on the same data (the ten scores), have been criticized on the grounds that in some of the hospitals the MMPI results themselves could have been influential in determining the final diagnosis. But no reason has been presented as to why that possibility should be more helpful to the statistical formula than to the professional clinicians implementing their "art." At the Ann Arbor VA Hospital in 1966 and 1967, I myself encountered another type of criticism. I had instituted a procedure whereby the profiles of all entering patients were automatically scored using the Goldberg formula. Whenever the clinicians in the hospital found a patient who had clearly been misclassified by this formula, they pointed that error out to me, sometimes gleefully—such as when it classified an actively hallucinating, psychotic individual as neurotic. They were silent about the errors they made that the formula didn't; perhaps they did not even note them. The result was that their memory was biased against the formula and in their own favor. I was confidently assured that the formula didn't work as well as I had maintained, at least at the Ann Arbor VA Hospital—as if the clinicians' memory of a small sample of patients were a better basis for establishing the formula's validity than a sample of more than a thousand patients analyzed systematically. (When I pointed out this possible bias in their evaluation, my colleagues would good-naturedly agree that it presented a problem, but none were motivated to do a systematic study of the accuracy of their *own* judgment, even on the small sample available.)

Cases where the professional psychologist has information in addition to that used in the statistical formula—but still makes worse predictions—may be found in almost any evaluation of unstructured interviews. In one Second World War study, personnel psychologists predicted performance of navy recruits in the military elementary school that these recruits attended before receiving specialized training. The personnel officers had access to the recruits' high school records or aptitude test scores, or both. They made predictions about how well the recruits they interviewed would do in the elementary schools they attended. R. F. Bloom and E. G. Brundage studied a sample of more than 37,000 recruits attending various schools and discovered that the predictions of the personnel psychologists were consistently worse than predictions based on the high school ranks, or aptitude test scores, or a combination of the two, to which these same psychologists had access.[17] This consistently poorer prediction of interviewers compared with statistical models based on predictive information available to them has been replicated again and again. Certainly the interview is valuable, but only as a way of discovering information that is truly predictive—which is best then analyzed by using a statistical model. The "clinical art" of interpreting interview results yields poorer accuracy than is obtained by combining this information "mechanically." Yet "experts" continue to interview, make predictions, and express great confidence in the validity of their predictive judgments. ("The more I do this, the more I learn and the better I am.") The practice of basing selection for jobs and academic or professional programs on such interviews is especially popular, again despite the evidence. By now, however, the results of new research on this practice are quite predictable. At a recent convention of the American Psychological Society, Thomas Gehrlein and Robert Dipboye presented a paper whose abstract reads as follows:

> Interview research has largely ignored differences among interviewers and incremental validity [the degree to which an interviewer may improve upon the information in the interview statistically

> combined]. These issues were examined in the context of college admissions. SAT's and high school rank were the best predictors of freshmen GPA. No evidence of [incremental] validity was found for the interview at the aggregate level or at the level of the individual interviewer. Contrary to expectations, experienced interviewers were no more valid than inexperienced interviewers. Results cast doubt on recent suggestions that interviewer-level analyses provide higher estimates of validity.[18]

This paper was not news. It would have been news only if the results had turned out differently. It simply refuted the suggestion that previous studies had underestimated the predictive validity of interviews by pooling results across interviewers.[19] Using the argument from the vacuum "logic" that by now should be familiar to the reader, the previous studies had made this suggestion to criticize results that the authors didn't like, but they provided no positive evidence that their suggestion might be correct. Gehrlein and Dipboyle presented evidence that it wasn't.

In fact, entire programs of interviews have been evaluated and found to be invalid. In April 1979 the Texas state legislature required that the University of Texas Medical School at Houston enlarge the size of its entering class from 150 to 200 students from Texas. The previous 150 had been selected by first examining the credentials of approximately 2,200 students and determining which 800 were best qualified. These 800 were then invited to the Houston campus, where they were interviewed by a member of the admissions committee and one other faculty member. The interviewers had submitted written assessments to a central committee, each member of which rated the applicant on a scale of 0 (unacceptable) to 7 (excellent). These rankings were averaged to obtain a combined ranking of all 800 students; the Houston ranking together with the rankings of the other three medical schools of the University of Texas were compared with the applicants' rankings of these schools by a computer program that guaranteed mutually highest choices. All 150 applicants who ended up coming to Houston were in the top 350 as ranked by the interview procedure. About ten dropped out and were replaced by applicants from lower ranks, to obtain the 150 students desired. When the school was required to add an additional 50 students to its entering of intellectual functioning correctly identified 83 percent of the new cases. But groups of inexperienced and experienced professional clinicians working from the same data correctly identified only 63 percent and 58 percent of the new cases respectively. When the clinicians were given the results of the formula, they did better (68% and 75% correct identifications respectively), but neither group matched the 83 percent accuracy of the formula. The clinicians' improvement appeared to depend on the extent to which they used the formula.[28]

In fact, in a series of studies David Faust and his colleagues discovered that professional psychologists could not even detect young adolescents who were faking brain damage on standard intellectual tests after being given virtually no instructions about how to do it other than "to be convincing."[29] Even when the faked results were sent to the professionals with an equal number of results from truly brain-damaged individuals and the professionals were truthfully told that there was a 50 percent chance that the test results they saw were faked, they still could not detect the fakes.[30] These professionals listed themselves (in the American Psychological Association directory or in the *National Register of Health Service Providers in Psychology*) as specialists in "neuropsychology"; many of them had had advanced training; and some of them had been awarded a special status of expertise called a "diplomate." Yet less than 10 percent recognized the faked results. Moreover, if anything there was a *negative* relationship between experience and ability to recognize the fakes—not, however, a statistically significant one. The usual criticism was made that the studies were flawed because neuropsychologists do not usually interpret test results without seeing the clients—yet this critique, as usual, lacked any positive evidence that the neuropsychologists would have done any better if they had seen the clients.[31] A more interesting criticism came from a famous neuropsychologist who, when told that the proportion of diplomates in the study was roughly equal to the proportion in the field, said to Faust, "Well, they couldn't really have been good neuropsychologists. Anyone willing to participate in your study could not have been competent."

Despite such poor showings and despite consistently poorer predictive and diagnostic performances than statistical analyses of test results make, a majority of neuropsychologists indicated in a 1988 survey that they preferred to use nonstandard methods—that is, intuitions—to reach judgments about intellectual deficit over statistical formulas.[32] These results have profound implications for court testimony about intellectual deficits that are alleged to have arisen from accidents or chemical exposure that might have resulted in brain injury, and for psychological "portraits" of the parties involved.

Studies of medical judgments are more mixed in their results, although when clinical judgments and statistical formulas are based on exactly the same input information, the formula once again makes superior predictions. Before Hodgkin's disease was controllable, for example, the late Hillel Einhorn studied how well judgments of the severity of the disease process as established from biopsy predicted survival time.[33] All 193 patients in the study died; the number of days of survival after the biopsy was the criterion studied. Three doctors, one an internationally recognized authority and the other two his "apprentices," rated nine characteristics of each biopsy that they believed were related to severity. They also made overall ratings of the severity of the disease process. While severity judgments are not identical to judgments about how long patients will survive, they should certainly be strongly related (in a negative direction, i.e., the greater the severity the less the survival time). Einhorn developed actuarial formulas—ones involving weighted averages of the numerical ratings of the doctors—to predict survival time from the nine characteristics for a sample of 100 patients the doctors had examined and then checked the accuracy of these predictions using these same formulas on the remaining 93. The doctors' overall judgments of severity were totally unrelated to survival time, but the formulas were. Einhorn's study demonstrated that the doctors' ratings of the biopsy characteristics provided potentially useful information in predicting survival time, but that only the statistical combination of these ratings actually predicted it. (Once when I talked about this study at a formal lecture, a dean of a prestigious medical school suggested that if only Einhorn had studied Dr. So-and-so, the recognized "world's expert," he would have discovered that a doctor's overall ratings could be quite accurate. I couldn't say so there, but the doctor was in fact Dr. So-and-so.)

A similar outcome has been found in studies comparing diagnoses of heart attack made respectively by doctors and a computer program in an emergency room. The doctors and the program were equally good at spotting a heart attack when it was actually present, but the program was superior to the doctors in diagnosing the absence of a heart attack when there was in fact none.[34] Statistical formulas have also been found to be superior to clinicians in predicting future heart attacks.[35] On the other hand, the predictions of a statistical formula (the APACHE-2) have been found to be inferior to predictions made by doctors who were board-certified in internal medicine, who were the "critical care fellows," and who "had seen the patient, obtained a history, and conducted a physical examination, as well as reviewed the pertinent laboratory and roentgenogram data available."[36] In another medical study, doctors were superior to a formula when they had more information than that used in the statistical models and when they had personally examined the patients.[37]

In the business context of predicting bankruptcy, a formula has been found to be superior to the judgment of bank loan experts, some of whom were highly paid by banks that loaned billions of dollars a year.[38] In a study predicting sales, however, managers outperformed the statistical formula.[39] In the bank loan study, the predictions of both the actuarial formula and the bank loan experts were based on the same information; but in the sales study, the managers "also had inside information" in addition to the information used in the statistical prediction. Thus, in both the medical and the business contexts, exceptions to the general superiority of actuarial judgment are found when clinical judges have access to more information than the statistical formulas used. Perhaps if this information had been incorporated into the formulas, they would have again been superior. In fact, that has happened. Unlike APACHE-2, a new statistical formula, APACHE-3, outperforms doctors in predicting death within 24 hours on an intensive care unit.[40]

These findings are not compatible with our intuitions about the validity of our intuitions. They challenge the expertise of professional predictions, and if professionals cannot predict the future well, how can the rest of us? Moreover, the findings appear to be "dehumanizing" in that they "reduce people to mere numbers." (But nothing in the statistical approach makes claims about what it means to be human; rather, the question is how to predict. In fact, the valid statistical approach involves a greater recognition of the role of autonomous choice than does the invalid clinical approach which is based on the

But why would we then believe in high predictability in general in such contexts? First, people have good cognitive reasons for seeking predictability in the world, and success in this search is, according to many theorists, the "function" of cognition. A belief that predictability exists when it does not may often create more harm than a belief that we can't predict when we can,[51] but a general bias to believe that the predictability is present in the world may certainly be adaptive. Moreover, we apparently have a compelling emotional need to believe in such predictability. A world that is not predictable cannot be a "just" one that provides us (good people) with the "entitlements" (good outcomes) we deserve.[52] (But no one wants a world that is *perfectly* predictable, which would be a dull one. Nor do we want a world that is *overly* just, in which everyone guilty of a bit of bad behavior or suffering from a touch of neurosis would be haunted with a fear of retribution.) Having cognitive and emotional needs for predictability, however, does not imply that it exists in every context we seek it.

AN EXPLANATION OF THE FINDINGS

Why is statistical prediction superior to clinical prediction in the contexts studied? Some of the reason have to do with the desirable characteristics of such formulas. They are specifically designed to discover a pattern in contexts of variability—the signal distorted by noise. The statistical formulas combine the information optimally to detect the pattern. Moreover, small differences in weights due to variability do not result in large differences in the predictions the formulas make. In fact, as long as the predictive variables themselves are positively related, small differences in combination rules (e.g. the weights applied in constructing weighted averages) result in predictions very similar to those provided by the optimal combination rules. When, for example, weighted averages are used to predict, random weights applied to standardized variables yield predictions very similar to those provided by the best possible weights.[53] Finally, the weighted averages provided by statistical formulas automatically involve the comparison of psychologically incomparable predictors.

People, in contrast, have great difficulty combining qualitatively distinct or incomparable predictors. How, for example, does someone reviewing an applicant for medical school combine information about a past college record with a score on the Medical School Aptitude

Test? In order to do so well, it is necessary to know about both the distributions of these predictors and their predictability—information that is not available to the judge on an intuitive basis, but that forms the basis of the statistical prediction. Similarly, how can an interviewer integrate information about past job history with a self-reflective statement about ambitions, talents, and goals? How does a clinical judge integrate a positive test result in a medical test or an unusual response to a Rorschach Ink Blot Test with knowledge that a disease indicated by such results is extremely rare?

Such integration cannot be done on an intuitive basis. Instead, clinical judgment is often based on a number of cognitive "heuristics" rules of thumb. The first heuristic is to search one's memory (including memory of one's training) for instances similar to the one at hand. This heuristic is termed *availability*.[54] Unfortunately, availability can be quite biased by selective exposure, selective recall, vividness of the instance or category recalled, and so on. A second heuristic is to match the cues or characteristics with a stereotype or a set of other characteristics associated with a category—a heuristic termed *representativeness*.[55] The degree to which something matches a category, however, does not indicate how probable it is. For example, our stereotype of someone addicted to intravenous drugs is that such a person smokes marijuana; hence, marijuana-smoking is a characteristic that matches our stereotype of an intravenous drug addict—even though people who smoke marijuana are far more likely *not* to use intravenous drugs than to use them, let alone be addicted to them.

Availability and representativeness are the heuristics that most commonly lead us to make poor judgments, but they are not the only ones. Since these heuristics have *some* validity in the judgments reviewed in this chapter, the clinical judges generally do better than chance, but they do not do as well as a careful choice among possible relevant factors and determination of how they should be combined, which is done automatically by a statistical model. (For a bitingly humorous description of these heuristics in action, see Paul Meehl's essay "Why I Do Not Attend Case Conferences."[56] In such conferences people spend a great deal of time in "free association," making judgments about patients being discussed by comparing them to a previous patient or a prototypical patient—or sometimes even a relative—on the basis of a single common characteristic, combining biased availability and biased representativeness.)

As illustrated in the Einhorn study involving Hodgkin's disease and longevity, however, the expert human judge does have a very important role to play in making predictions: choosing the variables that *might* be predictive and coding them. Just as people without medical training cannot code the characteristics of a biopsy, people without some training in psychology cannot devise tests that may be predictive of success in a job or academic setting. Here is the valid role of expertise. Once the variables are chosen and then constructed or coded, they should be studied to discover exactly how good they are at predicting outcomes. That's what statistical "science" is all about: subjecting ideas to public scrutiny in a way that will convince the critic who demands "show me." The resulting statistical formulas, moreover, need not be "rigid" (a common criticism of them); they may be modified to incorporate new information as it becomes available. That's the antithesis of the "only I can tell and I can't explain how" approach of much expert testimony in court settings.

OVERVIEW AND IMPLICATIONS

The superiority of statistical formulas in predicting gives rise to what can be termed a "base rate" psychology. People's behavior and feelings are best predicted by viewing them as members of an aggregate and by determining what variables *generally* predict for that aggregate and how. That conclusion contradicts experts' claims to be able to analyze an individual's life in great detail and determine what caused what. Unfortunately, it is exactly the individualized-causality type of analysis that is most expected of professional psychologists and other mental health professionals. This expectation arises not only from our intuitive beliefs about the world but from these psychologists' own declarations about their abilities. As David Faust once phrased it, such declarations should be viewed *versus* the demonstrations of what professionals can actually do.[57]

Moreover, as we have seen, the inability to predict implies a lack of understanding—not because understanding and prediction are synonymous but because a claim to understanding implies an ability to predict. Evaluating the efficacy of psychotherapy has led us to conclude that professional psychologists are no better psychotherapists than anyone else with minimal training—sometimes than those without any training at all; the professionals are merely more expensive. Moreover, in predicting what people will do, clinicians are *worse* than statistical formulas, and statistical formulas are *a lot* less expensive; even developing them is now no great expense, given the availability of inexpensive computer time. One criticism of the statistical formulas is that they may have to be constructed, modified and tested in each separate context in which they are used, but there is evidence that across similar situations there is "validity generalization."[58] It is a great pity that so much effort has been expended in repeating the same result, and raising and countering the same objections to that result, effort that could have been expended on using that result to develop better statistical formulas that will make better predictions. When our *Science* paper appeared, one critic (whose letter to the editor was not published) concluded that what the results implied was that psychologists and psychiatrists should be trained to use the standard psychiatric diagnostic manuals more reliably. Then maybe these clinicians would do better. Why not instead put our efforts into improving the method we know to be superior by developing better statistical models? That should benefit almost everyone—except, of course, the people who are being highly paid to make inferior predictions. But the general public is more important than they are.

THE ETHICS OF PREDICTION

One objection to these conclusions that I personally find particularly distressing—in fact, infuriating—is that making predictions about people using statistical formulas is "dehumanizing," that it treats people as "mere numbers." There is nothing in the approach that implies a judgment about what people *are*; the point is to make the best possible predictions, which can then be used to everyone's benefit. Moreover, the statistical model can be made public, open to scrutiny, and modified appropriately. It can even be shared with the people about whom predictions are made, so that they know how it is they will be judged.

Let me give an example of this openness. A study at the University of Oregon's Psychology Department indicated that a simple (weighted average) statistical model combining past record, test scores, and a crude rating of the selectivity of the students' undergraduate institu-

Past behavior provides an indicator of the former, and aptitude or achievement test performance an indicator of the latter. Unlike personality or honesty tests, the subject taking an aptitude or achievement test understands that there is a correct answer (such as to a short mathematics problem) and strives to obtain it. But what is the correct answer to a question about how one should respond to a valued employee who steals five dollars? According to the test scoring, the correct answer is to fire that employee. But if the person taking the test doesn't think that is the best solution, that person is scored for dishonesty.

The bottom line is a happy finding. In a majority of situations, an individual's past behavior is the best predictor of future behavior. That doesn't mean that people are incapable of changing. Certainly many of us do, often profoundly. What it does mean is that no one has yet devised a method for determining who will change, or how or when. Professional psychologists cannot predict that. (If any have been able to do so, it has been kept secret from the research literature.) But if we are responsible for anything, it is our own behavior. Thus, the statistical approach often weights most that for which we have the greatest responsibility.

CHAPTER 4

EXPERIENCE

The Myth of Expanding Expertise

> [It is] important, perhaps imperative, that psychology begin to assemble a body of persuasive evidence bearing on the value of specific educational and training experience.
>
> —American Psychological Association, 1982[1]

The empirical data indicate that mental health professionals' accuracy of judgment does not increase with increasing clinical experience, just as their success as psychotherapists does not. There are good logical and empirical reasons why experience does not help in this context, even though we may all "learn from experience" in other contexts. Moreover, there are good psychological reasons why the professionals incorrectly believe that experience does enhance their purported expertise, when it doesn't. The major reasons involve selective recall, selective interpretation, and assumptions about what is likely to be true even though it isn't observed.

Why does the American Psychological Association believe that assembling "persuasive evidence" is imperative, as expressed in the quotation that opens this chapter? The reason is not that evidence wasn't assembled, but that the evidence assembled was negative. The body of evidence at the time about psychologists in particular indicated that there was little of any value in their training and experience for their practice. In 1989, Howard Garb summarized the evidence in a *Psychological Bulletin* article[2]: Professional clinicians make somewhat better judgments than do nonprofessionals, but that can easily be explained in terms of differences in such characteristics as intelligence and by the fact that people who have learned how to use valid techniques employ them better than people who haven't learned to use them. That's not a surprising conclusion, but what may be surprising is that once the rudiments of the techniques have been mastered, their accuracy does not increase with additional experience using them. That is a very important finding. Any selective advantage that the professional has over the nonprofessional lies in their mastering the basics of a valid technique or two. They learn, for example, that the proper source books about the *statistical* evaluation of MMPI profiles will help them evaluate those profiles. They learn the principles of a particular behavioral technique, which leads to its proper use, and such techniques work.[3] *The accuracy of the judgment of professional psychologists and other mental health workers is limited, however, by the accuracy of the techniques they employ.* That's no different from any other applied field, but what has happened in psychology is that for intuitively compelling reasons, the myth has arisen that through experience per se a professional can develop accurate use of a "pet" technique that research has shown to be invalid, such as the Rorschach Ink Blot Test. Moreover, however often professional psychologists disavow the "medical model," the myth has arisen that the continued practice of a valid technique results in improvement, by analogy with medical procedures such as surgery. The research evidence supports neither of these myths.

Garb's generalization that experience does not improve performance was based on a survey of the research literature evaluating the performance of clinicians who employed a broad variety of techniques. One area in which we might expect there to be an exception is in evaluating neurological impairment. Measures of specific types of intellectual functioning have been carefully devised over many years. Ever since the success of tests of general intelligence used to screen United States military recruits in World War I, psychologists have been interested in differentiating various types of intelligence and intellectual functioning. Along with tests of overall "level" of intelligence (IQ), tests to evaluate specific types of intellectual abilities have proliferated, and many of them have been well validated. The use of such tests to evaluate the results of brain injury has likewise proliferated, and indeed many tests considered in isolation do determine abilities that are associated with such injury. Using such tests to determine how someone's abilities have *changed* as the result of brain injury is much more difficult, except in those few cases where the same test was administered prior to the damage—and even then any changes in performance must be evaluated with reference to the inherent degree of instability in the test results and to any factors other than the injury that might have occurred in the meantime. Given that they generally use such valid tests, neuropsychologists evaluating brain injury might well be expected to benefit from experience in their trade.

Not so. In a recent study (published after Garb's review) Faust and his colleagues asked "a nationally representative sample of clinical neuropsychologists" to evaluate the written results of tests of ten people known to have suffered from specific types of brain injury, or known to have suffered none.[4] Faust and colleagues concluded:

> Except for a possible tendency among more experienced practitioners to overdiagnose abnormality, no systematic relations were obtained between training, experience, and accuracy across a series of neuropsychological judgments. Comparable results were obtained when analysis was limited to the top versus bottom 20% (in terms of experience) of the sample. This and other studies raise doubts that clinical neuropsychologists train and practice under conditions conducive to experiential learning.

Why did the American Psychological Association's committee believe it "imperative" to "assemble a body of evidence" for something that isn't true? The reason I propose is that the success of psychology and often mental health professions stems from the public's belief that experience *does* enhance professionals' performance. After all, it does in many other professions, and therefore it must in the mental health professions as well. But although professional psychology is proclaimed to be based on "the science of psychology," it nonetheless sees a need to provide evidence that experience enhances performance, rather than admit that it doesn't and implement changes in its practices accordingly. In fact, as the profession proliferates at an ever-increasing rate, providing this nonexistent evidence becomes "imperative."

Sadly, the association's statement has not yielded the initiation of a broad research program oriented toward findings that may help practitioners and their clients. Rather, it stands as a public admission that the profession has been rolling merrily along in the absence of such findings, and it reflects the degree to which the profession has lost its research base. The statement definitely does *not* take the form: "Research evidence has shown . . . [say, that the Salk vaccine works]. Therefore, we . . . [recommend its use]." It reflects the opposite approach to gathering evidence: "We do it . . . [say, recommend laetrile]. Therefore, it is imperative to assemble evidence . . . [that it works, even though it doesn't]."

I could end this chapter here, because the empirical bottom line has already been established: Appeals to experience per se are invalid because experience per se does nothing to enhance accuracy. There is not even a hint in the research literature that it does—just selective anecdotal evidence. But I would like to specify *why* experience per se does not enhance accuracy and then discuss why we might be incorrectly convinced that it does.[5] These sections will provide the reader with an understanding of why false claims are false, and perhaps an understanding of the nature of learning from experience as well. The reader interested only in the bottom line may wish to skip—or simply skim—the remainder of this chapter.

We all, of course, learn from experience. That is, we learn some things about certain matters from some types of experience. It is tempting to conclude that we therefore always learn from experience, independent of what is to be learned and the nature of the experience. Didn't Ben Franklin say that "experience is the best teacher"?

Actually, he didn't. He said that "experience is a dear teacher," following a discussion of the Book of Job, and he added, "and fools will learn from no other."[6] It is clear from both the context and the addition that "dear" in this context meant "expensive." (In the 1940s and early 1950s, the owners of some small farms in New Hampshire whom I knew were fond of misquoting Franklin and contemptuous of "book learning"; their land is now owned by a neighbors' sons who went to college instead of acquiring "experience.")

LEARNING MOTION SKILLS VERSUS LEARNING HOW TO CATEGORIZE AND PREDICT

Learning is a term so broad that it refers to a multitude of activities. To "learn from experience" means to develop a particular skill as the result of a particular type of experience. Skills learned may be intellectual (like medical diagnosis), physical (like walking), or a combination of the two (like athletic or musical performance, or even driving a car). We learn simple motor skills differently from how we learn such skills as categorizing, predicting, and differentiating what is important from what is unimportant in a complex pattern, all of which are critical to psychological practice. What sorts of skills are developed as a result of what sorts of experience? What are the characteristics of the experience necessary to develop various skills?

It is clear that we learn many motor skills from practice. Accomplishing some of these skills requires coaching, like swimming and playing the piano; others, however, are acquired through practice alone, like walking and sitting in a chair. In fact, some motor skills that are acquired "automatically" through practice can be seriously disrupted by coaching. It is very difficult to coach someone about how to sit in a chair, for example; an amusing exercise is to explain the process to someone in words, insist that they follow your instructions exactly, then watch both the person and the chair collapse on the first attempt to follow your instructions. Some skills, such as driving a car, consist both of parts that develop automatically and of parts that are coached. Steering in a straight line, for example, is accomplished by making tiny discrete adjustments of the steering wheel that are not made consciously.[7] (The "weaving" behavior of drunk drivers is often due to impairments in making these adjusting movements rather than to any visual problem.) The skill in making these adjusting movements is developed only through experience in driving; in fact, on the first driving lesson most complete novices alternate between going toward the ditch and almost crossing the center line—much to the surprise and consternation of their novice teachers, who themselves are often unaware of their own "tremorous" movements of the steering wheel. Driving skill is developed as the result of continuous and immediate feedback. Like sitting in a chair, driving a car could result in disaster for the person who follows explicit verbal instructions exactly.

Learning to sit in a chair and to drive in a straight line epitomizes our idea of how skills are learned intuitively and are improved with practice. Is clinical skill in the mental health professions of that nature as well? People often explain that clinical skill, too, is based on experience that leads to an ineffable feel about how to proceed, but it is not. Rather, it is a cognitive skill that most often involves conscious decision making on the part of the psychologist. Occasionally, a psychologist may make an impulsive or "intuitive" response to what a client does, but these occasions are rare. Moreover, because the psychologist does not experience feedback about the effects of such responses on either an immediate or a continuous basis, they cannot be "shaped" automatically, as in they are in motor skills. The skills of a clinician are more akin to concepts and categorizing (diagnosing) instances. In order to understand how experience can aid in the development of clinical skills, it is necessary, therefore, to consider how people learn to identify concepts and hence to categorize instances. Work in the area of how people learn to identify concepts has spanned several decades and shown *why* experience per se does not enhance accuracy of clinical judgment. Two conditions are important for experiential learning: one, a clear understanding of what constitutes an incorrect response or error in a judgment, and two, immediate, unambiguous and consistent feedback when such errors are made. In the mental health professions neither of these conditions is satisfied.

Consider first learning the skill of categorizing people—say, distinguishing between child abusers and non–abusers—on the basis of their psychological characteristics rather than their histories. The problem is to decide which people are in each category. Presumably, the person learning this skill is first presented with a number of people who have already been identified as one or the other; the learner is then asked to make judgments about subsequent people. Supposedly the person "learns" the correct assignment through a process of finding out which of these judgments are correct and which are incorrect. Learning how to diagnose cancer is done this way; first the medical student is presented with patients who have cancer and with others who have overlapping symptoms but who do not have cancer; subsequently he or she is asked to make a judgment about new people, then receives feedback about which judgments were correct and which were incorrect.

CAREFUL STUDIES OF LEARNING TO CATEGORIZE

How people learn to make such categorizations accurately has been studied in psychological laboratories. Simplified problems have been invented in which people are asked to distinguish between two cateinvolves the ability to distinguish the important characteristics that define a concept from the unimportant ones. This is precisely the problem facing a clinician in making a diagnosis or categorization.

Experiments were conducted that asked subjects to sort instances into categories. The standard outcome measure of the experiments was the proportion of correct sortings they made across all the trials. The outcome involved pooling correct versus incorrect responses across subjects to obtain this proportion; initial analysis of the outcome led to the conclusion that subjects learn gradually, much as we learn how to drive a car. For any given problem, the subjects' proportions of correct sortings increased on each trial. Moreover, the improvement in this proportion correct was itself proportional to its distance from 100 percent—for example, increased twice as quickly when the subjects were sorting at a 60 percent correct rate as when they were sorting at an 80 percent correct rate. This proportional improvement is the classic form of the "learning curve," which describes gradual learning; such gradual learning is consistent with the *law of effect*, which postulates that responses are shaped automatically through reinforcement without being influenced by the subject's ideas or hypotheses. Thus, early investigators concluded that concepts might be grasped automatically through reinforcement contingencies, much as a motor skill is gradually learned through consistent feedback about the results of practice. If this is how concepts are learned, then professional clinical psychologists may gradually come to form the concepts necessary for their practice and learn through experience how to categorize people and behaviors correctly.

There is, however, another interpretation of the results, one that can be described as involving *terminal insight*. The subjects in a sorting experiment may have ideas about the possible categorical sorting rule that the experimenter has in mind. The subjects guess what the rule is and sort accordingly. When they are told they are incorrect, they abandon that hypothesis and try out another ("maybe it's the large figures that belong on the left"). Thus, if their set of ideas about possible rules includes the one the experimenter has in mind—and the experimenter consequently reinforces as "correct"—they eventually identify the right rule. Once they have identified it, they stick with it, because from that point on they will be making correct sortings. Hence the phrase "terminal insight."

Note that the ideas of the subjects about what categorization rule the experimenter is employing is extremely important in this interpretation; that is, the subjects and the experimenter must share a "common ground" about how the instances might be categorized and about the messages conveyed by the experimenter's statements "correct" and "incorrect."[8] Moreover, and equally important, the terminal insight explanation implies that subjects will try new ways of sorting by distinguishing characteristics *only after they receive feedback that they have made an error* (as when a doctor discovers that what she had believed was a stomach cancer turned out to be an ulcer on closer examination). It follows that if the terminal insight model is correct, then mental health experts can learn correct categorizations only after they have discovered that they have made a mistake. Given the probabilistic nature of knowledge in this area, however, how can they ever be certain that they have made an error? In addition, the type of feedback that mental health experts actually obtain—as opposed to that of medical practitioners—tends to be chaotic, even in terms of the time interval that elapses between judgment and feedback, and sometimes it is nonexistent—the client just disappears. If the gradual learning model for categorization were correct, such ambiguous and chaotic feedback could be surmounted; given enough experience, it could yield learning. If, however, the terminal insight model is correct, then such feedback virtually precludes learning. Thus, for understanding how people learn to categorize from experience, it is critical to differentiate between these two models. I will now present the research evidence that compares them.

COMPARING THE GRADUAL LEARNING VERSUS TERMINAL INSIGHT MODELS

What would the pattern of learning look like according to the terminal insight model? When pooled across subjects, the proportion of correct sorting on each trial would be much the same as those predicted by the gradual learning model. In fact, if the subjects who haven't yet guessed what categorization rule the experimenter had in mind have a constant probability of doing so on each subsequent trial, the results are identical. Subjects who had already guessed what it was by the previous trial would continue to sort correctly, and a constant proportion of those who hadn't would now guess correctly. Once again, at the back of some playing cards and attempting to decide whether the design—for example, a vase with flowers—is unitary or multiple might convince the reader that the distinction is extraordinarily vague.

Dulany and O'Connell's debunking experiment points to a very important conclusion. It is crucial in these concept identification experiments that subjects have a set of clear ideas about which rules the experimenter is apt to use *prior* to receiving the information about whether a particular choice is correct or incorrect. Hypotheses that do not exist can not be grasped. (In fact, this very principle led Plato to propose a heaven consisting of "pure forms" that must be recalled in order for a person to develop an understanding of such concepts as "virtue"—which must be defined in the same manner for a race horse as for a person.) A concept that is unknown to the subject—or that is sufficiently obscure that it does not "come to mind" as a possibility—cannot be understood by the subject simply by being told which stimuli belong to it and which do not.

IMPLICATIONS FOR "LEARNING FROM EXPERIENCE" IN THE MENTAL HEALTH FIELD

Now consider whether a psychological or psychiatric expert can learn a categorical distinction based on pure "experience." Generalizing from these careful experiments, we can conclude that such learning can occur only if categorical membership is based on a well-defined rule that can be understood prior to observing *anybody*, that is, prior to attempting to test this rule by applying it to instances. Otherwise, the professional is in the position of a medical diagnostician who is attempting to diagnose cancer on the basis of previous experience with cancer patients and those free of cancer without knowing what cancer is. Categorization—hence diagnosis and prediction—can be no better than the theoretical knowledge leading to the construction of well-defined categories.

Another important finding from the concept identification experiments has direct relevance to the learning through experience question. Recall that in these experiments *subjects identified the correct concept only after being informed that they were wrong.* Such explicit accurate feedback led them to reject an incorrect hypothesis and understand the correct one. The application to professional psychology is that even when a rough categorization does exist—as, for example, between "paranoid" versus "simple" schizophrenia—knowledge of what these categories are like and how to categorize individuals in them is gained through experience only when the learner judges incorrectly that an individual belongs to one category and then discovers through feedback that the individual belongs to another.

Now consider the statement that "I can identify child abusers because I have had experience working with fifty [or one hundred, or even five hundred] of them." Child abuse may have a fairly precise definition on the basis of actual behavior, but professionals who attempt to learn from experience to distinguish between abusers and nonabusers must—according to the learning-from-errors principle—have experience with people who *appear* to be child abusers but are *not*. Where does such experience come from? It is extraordinarily difficult to obtain; in fact, it is empirically impossible to obtain if one's contact is limited to people who actually are child abusers. It is also impossible to obtain by definition if the professional's conclusion that someone is a child abuser is assumed to be correct in the absence of supporting evidence.[11] Medical experts can claim to be able to recognize cancer without *extensive* experience in making erroneous judgments, but a clear definition of this "natural" condition (category) exists prior to the experience, as does the biopsy test to confirm whether it is present. This knowledge, in turn, results from previous research and biological knowledge. Compare that with knowledge about "the abusive personality." Categorical learning can occur in the absence of feedback about actual negative instances only when it is based on a "well-corroborated theory to make the transition from theory to fact (that is, when the expert has access to a specific model)," as Dawes, Meehl, and Faust put it.[12] Such models simply do not exist in the areas in which professional psychologists or psychiatrists most often make confident judgments—in courts—"as based on my years of experience." This experience is definitely *not* analogous to that in learning how to drive in a straight line or to sit in a chair.

Another characteristic of the concept identification experiments is that the subjects must receive immediate and correct feedback in order to identify the concepts. In fact, such feedback is important if people are to learn anything at all from experience—whether it is a concept, a general idea of how to deal effectively with people, or a motor skill. Even in learning to drive a car or sit in a chair, feedback must be immediate, systematic, and subject to a minimum of probabilistic distortion. But *the feedback most professional clinicians receive about their judgments and decisions is neither immediate nor systematic nor free of probabilistic distortion.* The immediacy problems are obvious: Lacking definitive procedures such as a medical biopsy, the correctness of the diagnoses and predictions that mental health professionals make often cannot be determined for years. Some are intrinsically impossible to evaluate, such as a judgment about who would be a better custodial parent. Knowing whether a good judgment was made would require not just feedback about what happened when one parent was granted custody but knowing what *would* have happened *if* custody had been granted to the other parent. But such hypothetical counterfactuals are unavailable.[13]

Even easily interpretable feedback, however, is probabilistic in three ways. First, the professional may or may not receive it, and its very existence may be biased in a particular direction—as when mental health workers notice people who return to an institution but do not notice those who do not; thus, "as any psychiatrist can testify, 'success' among the long-term mentally ill is a sometime thing."[14] (Somehow, I always thought that success was a "sometime thing" for any of us.) Even for practitioners outside institutions, as Courtenay Harding, Joseph Zubin, and Joseph Strauss point out, "there is no built-in system about eventual outcome success. They receive only the negative messages signaled by the reappearance of patients who have relapsed and often simply assume that those who leave 'uncured' are leading a 'marginal existence' somewhere."[15]

Second, feedback that is received lacks clarity or is rendered ambiguous by the operation of various confounds. Hence it is *intrinsically* probabilistic from the perspective of the professional. Human behavior and feelings are influenced by a multiplicity of incomparable factors, many of which are not known at the time when a judgment is made or that may exert an important influence only after that time.

Third, "self-fulfilling prophecies" arise when conditions that the professional may or may not have assessed correctly are influenced by the professional's own judgment. When a person is judged to be irredeemably violent and sentenced to death row, for example, this judgment itself may be a factor in facilitating later violence. When a seriously disturbed person is judged to be "in need" of hospitalization, this judgment may be a factor in worsening the person's condition. In such instances, no comparison is possible since the hypothetical counterfactual is missing—only here it is due to the professionals' own behavior (in judging the prisoner to be "irredeemable," and the seriously disturbed person to be "in need" of hospitalization), while the subsequent feedback is easily interpreted as indicating what would have happened anyway in the absence of the judgment.[16] (It is important to note that not all prophecies are self-fulfilling; for example, the prophecy that nothing bad will happen to me even though I take chances when I drive is a self-negating one.)

Such probabilistic feedback is a problem. Careful laboratory research has demonstrated that subjects cannot learn even the simplest task if the feedback about how well they are doing is sufficiently probabilistic.[17] Feedback "error" always hampers learning; how much depends on the nature of the task. Professional judgment in psychology is a difficult task, and the probabilistic component of the feedback after the judgment is made is enormous.

BUT THERE IS AN ILLUSION OF LEARNING

The fact that judgmental accuracy does not increase with experience is now (1992) being acknowledged in many journal articles; so is the efficacy of paraprofessionals as psychotherapists. Garb's review is an example.[18] Why, then, are so many professionals still convinced that their judgmental abilities are "honed" and enhanced by experience? So, unfortunately, are many courts, where a statement about years of experience is accepted as evidence of expertise. The answer lies in the feedback problem discussed above—specifically, in the biased availability of the feedback that the professional receives. These biases, as we have seen, include the lack of hypothetical counterfactuals, the probabilistic nature of the feedback, and the possibility of self-fulfilling prophecies.

But isn't constructing such categories (or mental illness) exactly what the diagnostic manuals are attempting? Yes, they are attempting it. Recall, however, that these categories have some natural characteristics and some artifactual characteristics—some more natural, some more artifactual. Those categories most easily classified as "natural" are those that involve genetic or other organic tendencies or "susceptibilities" that sometimes appear in overt form and sometimes do not. It is, for example, very difficult to diagnose schizophrenia or manic-depressive psychosis that is "in remission" other than on the basis of a previous diagnosis (and most probably a hospitalization). In contrast, those categories most easily classified as artifactual (like "histrionic" or "situational depression") are precisely those that have the largest amount of fuzz in their fuzzy boundaries. Much more important, however, is the problem that "by their fruits shall ye know them." If such categorization is important in treatment, why are untrained therapists as effective as those who have been trained in this categorization system? Why do simple statistical models predict better?

BIASES ENHANCING THE ILLUSION OF LEARNING

Some of the biases in feedback availability are also psychologically compelling and hence lead to a belief in the validity of learning, even as they result in its invalidity. The first bias is that instances where judgment turns out to be correct (even if on a purely chance basis) are often quite vivid. Such vivid instances are easily recalled by the professional and are shared with others. As such instances are "collected" throughout a career, their relative frequency is overestimated. Ironically, the more statistically improbable it is that a judgment would be correct, the greater the vividness of its success (perhaps yet another reason for derogating statistical formulas). Hence the greater impact of such instances over time as compared with more mundane judgments, which would potentially form a better basis from which to receive feedback, were it to be provided. Let me give an example of such a vivid instance.

Years ago a middle-aged man walked into a hospital complaining that he was growing breasts. The intake doctor noted that the man appeared depressed and asked why. The man answered that his mother had committed suicide earlier that week. The man was quickly referred to a hospital psychiatrist and subsequently placed on a locked psychiatric ward. The staff members on that ward, many of them psychoanalytically trained, were fascinated by the delusion of growing breasts following a mother's suicide. (The delusion was possibly related to the fact that the man, while married, had no children.) The man was given the usual psychological tests, including an intelligence test, the Rorschach Ink Blot Test, the MMPI (discussed in Chapter principle that we shouldn't take seriously what a disturbed person has to say. Recently, when I mentioned to a friend who is basically sympathetic to my view that people should feel as free to "shop around" for psychotherapists as for those providing any other service, she said that she thought this conclusion was questionable because "how can you expect someone who is seriously distressed to make a good choice?" My answer is, why shouldn't they? There is absolutely no evidence that emotional distress necessarily implies incompetence or an inability to judge what is helping or hurting in an attempt to alleviate that distress—any more than there is evidence that someone who has a severe physical problem cannot judge whether medical treatment is doing any good. Even many people who are seriously psychotic are out of touch with reality only *some* of the time about *some* aspects of reality.[20]

The second availability bias enhancing the illusion of learning from experience arises because professionals often can recall or cite specific instances in a way that creates feedback that is *consistent* but irrelevant. A professional notes that a majority of people who have a problem (say, neurosis) also have a characteristic "diagnostic" of that problem (say, recall of unhappy incidents in childhood). Another professional notes that women with breast cancer previously had high-risk breasts; still another notes that dyslexics had difficulty spelling. The relatively consistent association between the problem and the characteristic leads to the invalid conclusion that those with the characteristic will probably have the problem. No. The frequency with which people who have the problem also have the characteristic is *not* equivalent to the frequency with which people with the characteristic also have the problem. A vast majority of people who can recall unhappy incidents in childhood are *not* neurotic. Most psychotic individuals brushed their teeth as children, but tooth brushing does not predict psychosis, just as a vast majority of women with high risk breasts do *not* develop breast cancer, and a vast majority of those who can't spell are *not* dyslexic. Continually finding association by sampling only those who have the problem, however, often leads professionals to generalize from this select sample of people with the problem, which Paul Meehl, Gary Melton, and I—among others—have pointed out is an irrational generalization.[21] Careful laboratory studies have indicated that we all are prone to making this mistake, all child abusers. (I am not maintaining that a large number of child abusers, or even any particular proportion, stop on their own. I have no idea how many do. What I am maintaining is that experience limited to those who haven't stopped on their own is an irrational basis from which to conclude that child abusers "always keep doing it until they are caught.") Moreover, it was by definition impossible for these abusers to stop without therapy, because they were already in therapy.

When I pointed out these experts' flaws in generalizing, they agreed with me that there was a "logical" problem with reaching the conclusion they did, but they would then produce the same argument with equal vehemence at the next meeting. Our own experience is indeed extraordinarily important in reaching conclusions; it can overwhelm logic, and we would in general be in great difficulty if we didn't pay attention to it. When, however, our experience is systematically biased, it forms a poor basis on which to learn. Again, there's nothing unique about professionals making such inappropriate conclusions. The same problem—and the same vehemence—can be found, say, among distressed children of alcoholics who attend groups of other distressed children of alcoholics, and who subsequently maintain that any children of alcoholic parents who don't admit to being very distressed are simply denying their feelings.

Ultimately, the confidentiality law was not changed, but perhaps more because it infringed on the prerogatives of a number of professional groups—including lawyers, psychiatrists, and ministers—than because the arguments for changing them didn't stand up to logical scrutiny. But whether it was changed or unchanged, the law was of dubious value anyway, because therapists had an ethical obligation not only to report suspicions of child abuse but to warn clients in advance that they would report these suspicions. They thereby effectively warned their clients not to say anything to arouse their suspicions. Although current evidence indicates that the area of sexual abuse may be an exception to the generalization that psychotherapy is effective (few controlled studies have been attempted[23]), this exception should not be generalized to the area of physical abuse.

The third bias toward the illusion of learning from experience arises when people inadvertently create their own experience and hence their own feedback. As we have seen, it is difficult to "learn" the validity of a judgment that someone is "irredeemable" or violent when that judgment itself leads to circumstances prone to encouraging violence, like placement on death row. What is "learned" in such circumstances is due to a self-fulfilling prophecy. A waiter, for example, "learns" that only well-dressed people give good tips, because having developed that initial hypothesis for whatever reason, he gives well-dressed people better service than poorly dressed ones.[24] These self-fulfilling prophecies can be particularly pernicious. For example, when I was head of a department, a friend in an important departmental position created immense chaos and ill-feeling through his extremely aggressive behavior. He would denounce people in public, threaten to resign, and so on. In one conflict he finished a telephone conversation with me by screaming that "the only way you get anywhere in this world is to push, push, push, push!" But that was true for *him*. His own aggressiveness had so alienated most of his colleagues that they wouldn't cooperate with him unless they were coerced to do so. But from his own perspective, he was reacting to them in the only reasonable way. (Fortunately, he quit his position eventually, and his much more positive characteristics—which were always evident in his behavior outside the university setting—came to predominate.)

On a broader level, self-fulfilling prophecies are part of some mental ailments. Depressed people often suffer from "negative cognitions" about themselves and their lives; they feel socially inferior and rejected by others. Unfortunately, these people are often correct—their sense of inferiority and rejection is brought about by their own "dysphoric" behavior, as pointed out and documented by James Coyne and his colleagues.[25] Depressed people are not a lot of fun to be around. Efforts of friends to express sympathy and to "talk them out of" their depression fail; subsequently, friends avoid them. Thus, an initial judgment of one's own inferiority can lead to behavior that provides compelling feedback that this judgment was correct. This sad process can end up like the famous cartoon of the therapist informing a mother that her child does not suffer from an "inferiority complex": "I'm sorry, Mrs. Jones, your son really is inferior." Sufficiently strong belief (for example, in one's inferiority) can in some circumstances create the reality (for example, of inferiority).

Self-created feedback can occur in professional practice as well. For example, I have an acquaintance who can be described in nontechnical terms as self-indulgent, conceited, and precious. He is a psychotherapist who has apparently achieved some sort of inner peace through labeling himself as having a "narcissistic character disorder." When he was talking about his private practice a few years ago to a group of other professionals, he made the startling assertion that he had concluded that the major psychological problem with American males today is narcissistic character disorder. This syndrome is tied to their being told as boys to be brave, he said, and in particular not to cry. After he realized that he himself had such a disorder, he had become increasingly aware that many of his male patients suffered from it as well. Virtually all of them had been told by someone at some point in their childhood that they should be brave and not cry. (What boy hasn't?) Once he had reached this conclusion about his clientele, he had developed a specialty in helping male narcissistic characters, and subsequently more and more of them had come to him for help. Hence his discovery that this syndrome is the major psychological problem faced by American males today.

All these availability biases are magnified by retrospective memory, on which most people—professionals included—rely to "learn from experience." Memory is, however, basically a *reconstructive* process, as we demonstrated in laboratory experiments as far back as 1930 by Sir Frederick Bartlett.[26] We attempt to "make sense" out of our recall of bits and pieces of our past ("memory traces") in terms of what we "know" to be true of the world today, by "filling in the gaps." Moreover, the general ideas that we evolve after filling in some gaps will influence our search for the other traces that we end up recalling. That means, for example, that our recall of stability and change in our own life is highly influenced by our implicit or explicit theories of human stability and change at present. If we are currently depressed, for example, and we believe that adult depression is brought about by the childhood experience of having aloof and demanding parents, it is easy for us to recall instances where our parents were aloof or demanding. We thereby form a judgment that our parents were generally aloof and demanding, which supports our judgment that we are currently depressed because of the way they treated us. Or, to take another example, if we are quite politically liberal as older adults, and we believe that people tend to get more politically conservative as they grow older, it is easy for us to recall some political attitudes in our youth that were even more liberal than our current ones. Such memories reinforce our belief that people tend to get more conservative as they grow older—even if assessments of our attitudes taken at various times in our lives demonstrate that we have actually grown more liberal.[27] That also means that we have a *hindsight bias* in which we not only conclude that we "knew it all along," but are unable to recall what we actually believed before an outcome was known.[28] We now have become experts in what did occur and hence suffer from the *expert's curse* in reconstructing what we believed before we knew the outcome.[29] (It is often hard for an expert to understand what someone with less expertise will think or judge, such as about a wine or an automobile.)

While our memories may be vivid, that does not mean that they are accurate; I myself often have a precise visual image of where something is, only upon checking to discover that this image is incorrect. Study after study has indicated how easy it is to manipulate what people "recall," studies that include careful checks to make sure that the subjects are not deliberately distorting.[30] The fact that something is recalled does not necessarily imply that it has happened. The naive idea that someone recalls something accurately, or can't recall it, or recalls it accurately but lies about it, is incorrect. We reconstruct the past. (This problem will be covered in greater detail in Chapter 6.)

The reconstructive nature of memory serves to enhance all the availability biases outlined above, because the past is recalled in a way that "validates" our current judgments. To be sure, professionals take notes, and charts of patient behaviors and feelings are available. But the professionals do *not* include in such notes or charts their *current* judgments concerning their clients. When they review such notes, it is all too easy to be subject to biased availability and to conclude that something that happened later "made sense" in the light of what the client said in the notes. As Baruch Fischhoff points out, we cannot learn unless we are *surprised* at what has occurred (again, the importance of the "error" in learning). The "creeping determinism of hindsight" makes us unsurprised, and we therefore cannot learn even in those contexts where some learning may be possible. Professionals relying on retrospective memory are no exception. This problem is exacerbated when—as is too common—the notes themselves are "upgraded" retrospectively, sometimes weeks or even months after the actual interaction with the client occurred.

inhibits honesty. In particular, as Lee Sechrest (whose qualifications are described in Chapter 1) points out, it reinforces a human failure from which most of us suffer at least some of the time: a lack of "the courage to say 'we do not know how.'"[15] A license implies that its holder does know how. The license implies expertise in whatever the licensed person has been "trained" to do within the domain of "psychology." As Lee Sechrest points out,

> Court testimony is an example of what has happened. We [psychologists] drifted into it as a field. It started with psychologists talking about matters where they *did* have some expertise: measurement of intellectual functioning, descriptions of cognitive and behavioral impairment, and so on. Now psychologists can be "expert" on anything that can be defined as "psychology." That doesn't follow. Just because there are all sorts of things that are part psychological in nature—they involve behavior, beliefs, attitudes and so on—doesn't mean that we can claim to be experts in an area that involves these things without having to generate a scientific data base. Sexual abuse is an example. There are very few scientific data on the validity of the opinions that psychologists are giving. And psychologists can't just give an "opinion." Expert witnesses are forced to go beyond the data. Their fees depend on making these kinds of statements with a level of confidence that can't be justified given the state of our knowledge.[16]

A LICENSE TO USE TECHNIQUES THAT DON'T WORK

One pernicious effect of licensing is the pretense of knowledge in the absence of evidence. Worse yet, however, is the pretense that knowledge exists when there is evidence that this "knowledge" is incorrect. The history of the use of biofeedback to alleviate various disorders provides an example of such pretense. This history is recounted by Alan H. Roberts,[17] one of the first researchers to propose that biofeedback might be beneficial in alleviating such problems as headaches. He begins his history by pointing out that two articles published in 1969 raised the possibility that people could control their own autonomic emotional states through feedback involving systematic reinforcement methods, without any movement on the part of the person controlling these states. The first paper was written by the world-more than 2000 publications about the test" (italics in the original; p. 154). McCall pointed out that in particular the number of human movement responses, which had been thoroughly investigated by then, is unrelated to anything.

In the 1978 edition, Richard H. Davis concluded: "The general lack of predictive validity for the Rorschach raises serious questions about its continued use in clinical practice" (p. 1045).

What did the Rorschach supporters write in the *Yearbook*? They cited no evidence whatsoever. Instead, they justified use of the technique on the basis that it is a "very novel interview" or a "behavior sample." Yet more recent reviews show only that the Exner System, the antithesis of "projective" uses of the Rorschach, has some validity. Why then does use of the Rorschach continue among licensed professionals? One obvious reason is that they are paid well for administering it. Another, however, was mentioned in the favorable 1972 *Yearbook* review by A. G. Bernstein: "The view that recognition [*sic*], the act of construing an unfamiliar stimulus, taps central components of personality functions is one that will remain crucial to any psychology committed to the understanding of human experience." I agree that the view is crucial to the *practice* of using projective tests, but is it a correct view? It yields a plausible belief that it should work. Such a belief provides a good rationale for seeking positive evidence that the Rorschach actually works. The evidence is that it doesn't work.

This evidence has finally had some impact. Professional psychologists are now being warned against using the Rorschach and other projective tests as a basis for court testimony. In the February 1991 issue of the *Pennsylvania Psychologist Quarterly*, the members of the ethics committee of the Philadelphia Society of Clinical Psychologists warn against its use in a court setting: "any psychologist who chooses to use instruments whose validity has not been demonstrated as predictive of desirable arrangements (for example, projective tests), should be prepared to be challenged on ethical grounds."[40] I ask, however: *If the use of an instrument (a projective test or any other) can be challenged on ethical grounds in a court of law, how can its use be ethically justified in any context at all?* My answer is that it can't.

I would like to offer the reader some advice here. If a professional psychologist is "evaluating" you in a situation in which you are at risk and asks you for responses to ink blots or to incomplete sentences, or explaining *others'* behavior. That again is compatible with what most of us do: "He is an angry person, you are touchy, I just look out for my rights." Here, aggressive behavior is ascribed as a stable personality characteristic to "him," variable behavior to "you," and a reasonableness to "me." As noted in that famous declination, the causes we often hypothesize to explain behavior can be quite different when applied to others from the ones we attribute to ourselves.

These causal attributions vary in a simple but systematic way. As Ned Jones and Richard Nisbett have pointed out, we have a *fundamental attribution error* or *bias* to attribute the behavior of others to personality factors and that of ourselves to situational factors.[13] This bias is not merely self-serving, as in the above example of explaining aggressive behavior. The person who engages in a heroic act—helping others off a burning airplane, diving under the ice to rescue a child—and then maintains that "anyone would have done the same thing" is not merely being modest or even false modesty. When people act themselves, they attend to the environment—to the plight of the people needing help, to an insult, to a competitive threat. The hero or villain is not attending to his or her own heroism or villainy. In fact, it isn't clear that there are internal cues at all about what heroism or villainy feels like, certainly not cues that differentiate my feelings from the feelings of others, to which I do not have access. In contrast, the observer of the heroic or villainous act is primarily attending to the person who is acting, not to that person's environment. The actor seeks causal explanations in the environment, while the observer seeks them in the actor. The observer of an heroic act will quite likely attribute the actor's claim of attending only to "what needed to be done" as due to modesty or false modesty. (It follows, incidentally, that we often worry that the military capacities of a potential enemy country may lead it to take aggressive action, while we rarely worry that our own military capacities might have a similar effect on *us*.) We are biased to attribute causality to the focus of our attention. The professional mental health expert delivers a message perfectly consistent with this bias. What other people do—in particular, the things they do that we don't like—is due to their personality. They suffer from such problems as low self-esteem or repressed hostility. They are not mentally healthy. Thus, we derogate them without having to acknowledge our own judgmental attitudes or anger—as we would, for example, if we were to label them "evil" or "depraved." The icing on the cake is that when *we* do things that we don't like, we attribute our acts to environmental factors, specifically the way in which our parents raised us and our consequent experiences that have stunted our growth up through the hierarchy of mentally healthy states.

THE BELIEF IN THE TYRANNY OF CHILDHOOD EXPERIENCES

But do our childhood environments and experiences really have such a profound causal effect on the rest of our lives? When Alexander Pope wrote that "Just as the twig is bent, the tree's inclin'd,"—changed to "the child is father to the man" by Wordsworth, not Freud—Pope was discussing education, not early childhood experience.[14] As compelling as his analogy is, it is generally not even literally true; trees are bent by consistent forces such as rocks, other trees in the way of their growth, or wind, and in absence of such forces will orient straight toward the sunlight. Moreover, Freudian psychology's claim that adult problems are "caused" by childhood experience can be interpreted in two ways. The first is that some experiences are pathological and will inevitably yield problems; for example, people who have been abused in some way as children *must* suffer as adults. The second is that when problems occur, they take the form of "regression" to childish ways of coping and hence mirror these experiences. For example, if—for whatever reasons including medical ones—people become distressed as adults, the form and type of their distress is likely to mirror the form and type of distress they experienced as children. Compatible with this second interpretation, Freud observed that he knew many people who had experienced the same childhood traumas and pathologies that his patients had but who nevertheless did not become disturbed as adults—a fortunate outcome he often ascribed to "constitutional factors." Nevertheless, our culture has come to accept the primacy of childhood experiences in yielding adult character and personality problems, and this acceptance is widespread. A recent book about the painter and sculptor Max Ernst tells us that "we now *know*, in the light of twentieth-century psychological research, that childhood experiences of the type Ernst suffered are a decisive factor in personality development" (italics added).[15] In a more popular outlet, *TV Guide*, the director of a special television ders a "romance of childhood"—not evidence that childhood experiences radically constrain adult functioning. I suggest a more dramatic phrase: *tyranny of childhood.* Americans marvel at how people from "primitive" cultures accept absurd beliefs on the basis of little evidence, and at how the Germans in an "advanced" culture could have believed all the nonsense about "Aryan superiority." Yet our belief in the tyranny of childhood has little more foundation than a belief in a mountain god. Yes, the professional mental health authorities propagate this belief, but the authorities about mountain gods also propagate beliefs about mountain gods. Acceptance of this belief is to their advantage; after all, if the locus of problems is childhood and its effects are tyrannical, then interminable talking about childhood while paying a handsome fee to the listener must be the only way to escape the tyranny. Again, I am not claiming that mental health professionals propagate this belief deliberately to make money, but again, what works is reinforcing and evolves. Many scientifically oriented mental health professionals, by contrast, have come to understand the misguided intuitions and the distortions of retrospective memory that underlie this belief. They have come to slough it off, and I am suggesting that it is time for the rest of us to do so as well.

In summary, we believe in the authority of mental health professionals because we have continually heard that they are experts, because we are prone to accept what people claiming to be authorities say anyway, because these particular authorities tell us what we already believe, and because they reinforce our bias to attribute undesirable behavior in others to personality characteristics ("mental illness") and in ourselves to environmental circumstances—particularly to the environment in which we were raised as young children and the tyrannical effects it has had on us throughout our lives. These needs and beliefs do not stand up to rational or empirical scrutiny, but they are there. Consequently, we accept even contradictory assertions, we agree, and we license.

We shouldn't. We don't have to. Instead, we should believe the recommendations and research findings of those psychologists and psychiatrists who believe that psychological knowledge should be "given away," to use a phrase from the APA presidential address of one of the last distinguished research psychologists to be elected to that post, George C. Miller.[46]