2

The Clinician as Subject

Practitioners are Prone to the Same Judgment Errors as Everyone Else

JOHN RUSCIO

In the *New York Times* bestselling book *Word Freak*, Stefan Fatsis (2001) chronicled his journey into the world of competitive Scrabble players. The tale he tells about the development of expert judgment holds lessons that extend well beyond the realm of Scrabble. Players must memorize a tremendous amount of information, beginning with game rules and the frequencies and point values of the letters in a set of Scrabble tiles. This much is fairly simple, but studying the lists of acceptable words presents more daunting task: There are about 120,000 words allowed in U.S. tournaments, and the addition of about 40,000 British words yields a total of 160,000 words allowed in international tournaments. It takes many years of devoted study to approach complete word knowledge, and even the leading experts engage in a continual struggle to retain this information and create multiple, complex interconnections so that as many words as possible can be retrieved quickly in different game scenarios.

As impressive as these feats of memory may seem, successful expert-level play also demands sophisticated information processing. Increasingly thorny judgments and decisions must be made as one learns to master such strategic issues as rack and board management and the handling of the end game. Experts do much more than scan their memory stores for

possible word plays. For example, many of the words played by experts are unrecognizable to laypersons, and even competitive players can be uncertain whether a particular play is an allowable word. This raises the question of whether to gamble a challenge of an opponent's play: If the word is invalid, it is removed and the opponent forfeits that turn. If the word is valid, one loses his or her own turn. If one opts not to challenge, another decision is whether to play a word or exchange one or more tiles. Particularly if no high-scoring or defensively important plays can be identified, it can be wise to forfeit a turn to exchange some unwanted letters for new ones. If one opts to make a play, this forces the decision of when to terminate the search for the best available play. These decisions, along with many others, must be made using limited information. One's retrievable word knowledge is incomplete, and information regarding an opponent's tiles and those that remain in the bag is bounded by probabilistic constraints. Likewise, decisions must be made rapidly, as there is a penalty for running over the 25 minute limit each player is allotted per game. With massive amounts of study and practice, some Scrabble players achieve a state in which their command of strategic decisions and generation of optimal or near-optimal plays appears effortless. Through a rigorous course of training and experience, the deliberative, short-sighted, and relatively foolhardy style of play exhibited by novices is replaced by the wisdom and automaticity characteristic of experts.

The process by which Scrabble players hone their judgment provides many useful clues about how to improve clinical judgment. Clinical practitioners must acquire and retain a wealth of factual knowledge as well as decision-making strategies for applying this knowledge effectively. Learning and using the full breadth and depth of theory and research related to the assessment, classification, and treatment of mental disorder within the constraints of applicable ethical and legal codes certainly does not constitute a game, yet many of the challenges of clinical work are analogous to those of an intricate game. A broad array of potentially relevant client characteristics, alternative interventions, and therapeutic goals constitutes the panoply of variables to consider. Relations among variables, especially causal relations, are seldom established unequivocally by previous research or experience. In light of available assessment tools and techniques, it can be difficult to obtain pertinent information in a reliable and valid manner. For a number of reasons, one will often have to make probabilistic inferences regarding gaps or apparent inconsistencies in the data. The nature of the judgments and decisions to be made, and the available options, are often open-ended. Ethical and legal codes proscribe some courses of action, but the breadth of tolerable practices remains vast. Tough choices must be made, and they can have significant consequences.

The complexity of the situation faced by clinical practitioners often demands the use of shortcuts to make critical judgments and decisions. Otherwise, the cognitive limits of human information processing could

easily be exceeded. Likewise, inattention to potential cognitive biases can lead to judgment errors that might otherwise have been prevented. Although people vary in their aptitude for memorization and strategic thinking, the formidable knowledge base and skill set involved in competitive Scrabble or clinical practice must be built through training and experience. In what follows, suggestions for the development of expert clinical judgment will be drawn from an examination of cognitive limitations and biases, the disproportionate influence of personal experience, and the requirements for successful experiential learning.

Before proceeding, it is worth underscoring the approach and emphases of this chapter. Rather than attempting to catalogue exhaustively the types of errors that have been identified in the judgment literature, I have selected a handful of exemplars based on their applicability to clinical practice. Likewise, I have presented illustrative instances of judgment errors instead of descriptions of relevant research studies. I have provided citations for readers interested in pursuing additional reading, but the emphasis here is on the detection and prevention of judgment errors in clinical practice. Finally, and perhaps most important, this chapter's focus on judgment errors is not intended in a pejorative sense. Human fallibility stems from universal cognitive limitations and biases, not from foibles unique to practitioners. As the chapter subtitle states, clinicians are prone to the *same* judgment errors as everyone else. In everyday life, individuals are relatively free to use flawed reasoning. In the role of an expert, however, one assumes an added responsibility to "get it right." Training and experience are expected to correct errors in experts' intuitive understanding of their disciplines, including both the factual knowledge base and the implementation of appropriate techniques through sound reasoning. The examination of error in this chapter is intended to introduce students in the mental health professions to the sources, types, and prevention of common judgment errors to which everyone is susceptible but that can adversely impact clinical work.

COGNITIVE LIMITATIONS AND BIASES

One of the most fundamental principles guiding research on judgment and decision making is that human information processing is constrained by certain cognitive limitations. For example, there are limits to the amount of information that can be retrieved into and held in working memory (e.g., Miller, 1956), the complexity of the operations that can be performed on this information (e.g., Halford, Baker, McCredden, & Bain, 2005; Ruscio & Stern, in press), and the speed with which information can be processed (e.g., Sternberg, 1969). Whereas a computer will be unable to solve a problem when its memory capacity is exhausted or will spend as long as necessary to work out a solution when its memory is sufficient and its processing speed is the limiting

factor, clinicians seldom have the option of either reaching no judgment or taking longer to make a decision. When working with a client, many provisional judgments must be made rapidly, on the basis of a wealth of information of mixed or ambiguous validity, to proceed with an assessment or treatment during an ongoing session.

When a judgment must be reached, cognitive limitations often necessitate the use of mental shortcuts, or *heuristics* (Turk & Salovey, 1988; Tversky & Kahneman, 1974). By simplifying the task, these strategies afford a judgment—even if a normatively suboptimal one. Usually, there is an inherent trade-off between accuracy and efficiency (but see Gigerenzer, Todd, and the ABC Research Group, 1999, for exceptions in which both accuracy and efficiency can be improved). Of particular interest is that the errors resulting from the use of heuristics are not always random. Predictable types of mistakes are sometimes observed, in which case the use of a mental shortcut can be understood as causing a cognitive bias.

Representativeness and Availability Heuristics

Two heuristics have received the lion's share of attention in the literature, as they manifest themselves in myriad judgment errors. The *representativeness heuristic* produces similarity-based judgments made on the superficial basis of "like goes with like" (Kahneman & Tversky, 1972). For instance, effects are presumed to resemble their causes. Such relationships often, but do not always, hold. Consider the popular notion that mental disorders with a "biological basis" are more appropriately treated with medication than with psychotherapy, whereas psychotherapy should be reserved for disorders with no biological basis. Setting aside the often vague meaning of "biological basis"—here it will be used to signify that biological factors play a role in the etiology of a disorder—the underlying assumption appears to be that a biological problem suggests the need for a biological solution (and vice versa). This clear case of representative thinking gives rise to a number of logical problems and conceptual puzzles.

Pitting interventions against one another in this way creates a false dichotomy between different levels of analysis (biological and psychological) at which one can conceptualize and test theories of psychopathology. There is no logical inconsistency between the existence of biological bases for a disorder and an understanding of that disorder in terms of psychological mechanisms. Unless one is a mind-body dualist, it should be easy to see that all mental functioning, normal or abnormal, must have a basis in the brain (see also Chapters 13 and 15). However, even though all mental disorders are biologically *mediated* (i.e., situated somewhere in neural tissue), this does not guarantee that either the original cause(s) or the successful treatment of a disorder is biological in nature. Thus, the notion that some disorders have a biological basis whereas others do not

is logically flawed. Instead, it is more appropriate to ask about the nature of the biological basis for each disorder and to pursue possible treatments based on promising knowledge at any level of analysis.

In addition, the apparent correlation between the existence of biological bases for disorders and the availability of biological treatments may be spurious. Whereas the discovery of biochemical anomalies among individuals suffering from a particular mental disorder often prompts the development and testing of new medications, the absence of known biological anomalies prohibits such focused research on biological interventions. Thus, present *knowledge* of biological bases may be associated with the availability of biological treatment options, with no causal connection between the nature or extent of biological bases and the utility of biological interventions. In the end, of course, efficacy and effectiveness research are required to evaluate the appropriateness of any treatment. The naïve, "like goes with like" belief that disorders with known biological bases are most appropriately treated using medications may hinder the search for fruitful treatments.

Whereas representative thinking uses similarity as a cue, the *availability heuristic* produces judgments of frequency or probability on the basis of the ease with which instances can be retrieved from memory (Tversky & Kahneman, 1973). Whereas the ease of recall generally provides a useful clue to how common or rare a class of events is, this heuristic can sometimes lead to biased or erroneous judgments. Unusual occurrences often attract greater attention than more mundane happenings, with the result that one might be able to retrieve instances of these relatively rare events more easily than objectively more frequent events. This can be especially true of vivid, emotionally compelling events that seem noteworthy in large part because of their rarity (Nisbett & Ross, 1980). For example, when Schreiber (1973) published *Sybil*, few (if any) individuals diagnosed with multiple personality disorder (MPD, now listed in *DSM-IV* as dissociative identity disorder) had reported childhood abuse or as many as 16 alternate personalities. Highly unusual features such as these not only helped to captivate a large audience but also served as models for future reports because many people—including professionals and laypersons—formed an MPD schema on the basis of this exceptional case. Many (if not most) subsequent MPD reports included childhood abuse and increasing numbers of alters (Spanos, 1996). Despite the absence of compelling evidence that childhood abuse is correlated with diagnoses of MPD (Lilienfeld et al., 1999; Spanos, 1996)—let alone etiologically relevant—when clinicians rely on the availability heuristic in evaluating this putative association, they can retrieve many instances consistent with an abuse-MPD link.

Even if there is no statistical association between abuse and MPD, such an *illusory correlation* (Chapman & Chapman, 1967; see also Chapter 1) may persist due to the operation of the availability heuristic. A clinician who specializes in the diagnosis and treatment of MPD can expect to encounter a number of patients who report incidents of childhood abuse during a life history interview. After all, childhood abuse is not uncommon

among clinical patients (or, for that matter, among mentally healthy individuals; Renaud & Estess, 1961). The availability of these instances in memory may be mistaken as evidence to support the abuse-MPD link. What is *not* available in memory are the frequencies with which individuals not diagnosed with MPD do and do not report abuse. Potentially available, but not especially salient, in memory is the frequency with which individuals diagnosed with MPD do not report abuse. Without comparing the relative frequencies of abuse histories among individuals diagnosed with MPD to individuals not diagnosed with MPD, one cannot determine whether these variables covary (see also Chapter 16).

The operation of the availability heuristic explains how illusory correlations can be formed from equivocal observations, and additional research suggests that such illusions can persist in the face of contradictory evidence. Chapman and Chapman (1967) demonstrated that laypersons and clinical psychologists share many false beliefs about relations between characteristics of human figure drawings and the personality traits of the individuals who drew them. For example, the empirically unfounded belief that people who draw large or exaggerated eyes tend to be suspicious or paranoid is one illusory correlation used in the Chapmans' work. When provided with evidence of a *negative* relationship (e.g., a series of drawings and personality traits paired such that paranoid individuals tend to be less likely to draw large or exaggerated eyes), individuals still reported that they "learned" from these data that a *positive* relationship holds. The fact that laypersons and clinicians share many illusory correlations regarding projective tests, coupled with the fact that these illusions can persist despite experience with contradictory evidence, may help to explain the popularity of projective test indices of limited validity (Chapman & Chapman, 1969; Wood, Nezworski, Lilienfeld, & Garb, 2003; see also Chapter 4).

AU: Garb not listed in References; please confirm.

*Ba*d Habits: Confirmation and Hindsight Biases

The representativeness and availability heuristics are mental shortcuts that sacrifice accuracy for efficiency, yet they only result in biased judgment under certain circumstances. Other aspects of the normal cognitive repertoire, however, include more intrinsically biased ways of thinking, which Faust (1986) labeled "bad habits" (see also Chapter 3). One such bad habit, known as *confirmation bias*, involves selectively seeking, attending to, and attaching greater weight to information that supports rather than refutes one's own beliefs (Nickerson, 1998; see also Chapter 1). For example, some clinicians who work with victims of trauma use techniques to recover allegedly repressed memories (Poole, Lindsay, Memon, & Bull, 1995), and Sagan (1995) suggests that the nature of the material obtained using these techniques often bears an uncanny resemblance to the expectations of the practitioner. There are at least three specializations within this niche, each of which involves

belief in the high frequency and pathogenicity of a particular type of trauma: child sexual abuse, satanic ritual abuse, and alien abduction. Patients whose therapists emphasize alien abduction tend to recover memories of being abducted by aliens, seldom of being sexually abused as a child or of being abused by satanic cults. To the extent that a similar correspondence holds for clinicians in each of these specializations, this would place considerable strain on coincidence as an explanation, even after one acknowledges the potential influence of referral biases (i.e., patients may seek out or be referred to practitioners who share their core beliefs). The most parsimonious explanation may be that confirmation bias guides the memory recovery process, which proceeds in the service of strongly held preconceptions rather than in a more objective search for veridical information (Lynn, Lock, Loftus, Krackow, & Lilienfeld, 2003).

When confirmation bias goes unchecked, open-minded consideration of multiple perspectives can become the exception rather than the rule: Support for a single working hypothesis is sought and incoming information passes through filters that operate to distort or remove potentially troublesome data. Whether intentionally or not, we expose ourselves to situations and environments that favor our prior beliefs. For example, we tend to associate with people who think as we do, read books and articles that support our views, and join professional organizations and attend conferences to interact with others who share our beliefs. Information is often packaged in ways that will most appeal to people who hold certain beliefs—and that will not challenge those beliefs. Different chapters in the same edited book, like different presenters within a symposium at a conference, seldom take opposing positions. By choosing which book to read or which session to attend, one can avoid dissonance-provoking confrontations with alternative viewpoints. More generally, consumers of information are increasingly able to select information sources that share their preconceptions. Although it can be comforting to experience agreement on positions regarding important issues, there are serious drawbacks to consider.

First, one might mistake a carefully selected survey of opinion—a highly biased sample—for genuine, generalizable agreement. It is easy to overestimate the extent of support for a position, or the expertise of fellow supporters, when one only consults articulate, like-minded individuals. For example, in the fall of 2004, *National Public Radio* aired a story on the skyrocketing sales of political books during the U.S. presidential campaign. A number of book publishers observed that sales were brisk, yet none believed that these books were influencing readers' political views. Instead, they suspected that people were buying and reading books by authors that shared their views to gain ammunition—in the form of the authors' credentials as well as the readers' favorite anecdotes or factoids—for political discussions and debates.

Second, to avoid discrepant views is to squander valuable opportunities to learn, especially when one holds mistaken beliefs that are correctable.

Often, one stands to benefit far more from engaging rather than evading the expertise of those with whom one disagrees. If the best arguments and evidence, presented in the most compelling fashion, fail to adequately support an opposing position, one can place greater confidence in one's own. In contrast, the case for an alternative stance may warrant changing one's position. Without giving a fair hearing to those who hold different views, one might foolishly cling to misguided beliefs.

The bad habit of confirmation bias manifests itself in many judgments and decisions that clinicians are called on to make routinely. For instance, when gathering information to reach a diagnosis, a preliminary hypothesis is often formed remarkably quickly (Garb, 1998). This working hypothesis can steer one toward a search for supportive information rather than the more normatively appropriate testing of competing hypotheses (Faust, 1986). Assessment performed in a confirmatory mode is likely to yield information that is consistent with an initial hunch, but this consistency is interpretationally ambiguous because the same information may be equally consistent with other, unconsidered hypotheses. The failure to adequately consider alternative hypotheses is known as *premature closure*. A clinician aware of this danger could pose multiple hypotheses and determine how to tease them apart most effectively. Performing assessment in a more explicitly hypothesis-testing mode is more likely to yield evidence that genuinely supports correct ideas and contradictory evidence that serves to rule out false ones.

Another bad habit of human judgment, *hindsight bias*, involves mistaking a perceived understanding of the past for an ability to predict or control future events (Hawkins & Hastie, 1990). Once knowledge of an event's outcome becomes available, one has a feeling of having "known it all along" (Fischhoff, 1975). This phenomenon has also been described as "creeping determinism" (Fischhoff, 1980), as a chain of events can appear to have unfolded in an inevitable sequence. Because it is easy to construct plausible explanations for events after they have occurred, it is unwise to place much confidence in such accounts, much less to deem an outcome inevitable. The remarkable ability to recognize patterns, which enables us to craft a good story by imposing order on chaos, is a perceptual skill of inestimable adaptive value. However, an apparatus adept at organizing information into coherent patterns carries with it the liability of occasional mistakes, patterns that are only apparent and not real. Given the survival imperative of successfully learning environmental contingencies, one might expect human beings to be imbued with a positive bias toward the recognition of potential patterns even when this entails frequent false positive identifications. The frequency with which people commit the *post hoc ergo propter hoc* fallacy (B follows A, therefore B was caused by A) attests to such a hypersensitivity of our pattern recognition faculties. For example, reasoning that "I tried this treatment and felt better, therefore the treatment works" is to commit this fallacy. Beyerstein (1997) described many alternative explanations that cannot be

ruled out when attempting to draw conclusions on the basis of personal experience, testimonials, or other anecdotal evidence.

Similarly, Meehl (1973) described as a common fallacy observed in clinical case conferences the "assumption that content and dynamics explain why this person is abnormal" (p. 244). Engaging the services of a clinical practitioner establishes that the client is currently experiencing problems that, even if not diagnosable as mental disorder, involve at least some of the symptoms. The individual's present mental state constitutes an outcome in need of an explanation, and one's therapeutic orientation often guides the conceptualization of the case. For example, clinicians who believe that traumatic exposure is the root cause of most mental anguish tend to search for trauma in a life history interview. Because even most mentally healthy individuals have experienced events that can be described—whether by client or therapist—as traumatic, a sufficiently effortful search will nearly always yield information that is consistent with the clinician's etiological theory. Confirmation bias can be influential in guiding the selective search for this information, but hindsight bias is the culprit when one concludes that the uncovered trauma explains the client's current mental problems. This outcome only seems inevitable in hindsight, and there may be either no causal connection between the trauma and present mental state or a connection that is more subtle or complicated than presumed. Either way, the premature acceptance of the first plausible narrative may preclude a more thorough assessment of other factors necessary for the most accurate case formulation or the best treatment plan. The true test of understanding is not the construction of a plausible explanation for past events, but the successful prediction of future events (Dawes, 1993).

THE DISPROPORTIONATE INFLUENCE OF PERSONAL EXPERIENCE

Mental health disciplines such as psychology, psychiatry, and social work grant professional degrees that certify expertise in clinical practice. In an article aptly titled "Credentialed Persons, Credentialed Knowledge," Meehl (1997) considered the evidential support required to substantiate such claims to expertise. Any field of study necessarily begins with the anecdotal evidence of its practitioners' personal experiences. In clinical work, experience can include training exercises as well as supervised and independent practice; the term "personal experience" does not mean "single case" (see Chapter 7 for a discussion of the inferential value of single cases). Of course, anecdotes all too readily suggest faulty conclusions and unwarranted generalizations, especially when parsed impressionistically (Faust, 1984; Meehl, 1992). To overcome the short-comings of human judgment, pioneers of a new discipline must promote a balance between open-minded speculation and skeptical inquiry within

an atmosphere of dispassionate investigation. Recognizing that scientific methodology—including research design and data analysis—has been crafted to counter cognitive limitations and biases in teasing apart fact and fiction, Meehl (1997) emphasized the importance of collecting data systematically and testing relationships between variables using appropriate statistical analyses.

For a variety of reasons, clinicians' personal experience often exerts a strong influence on their judgments even when more reliable and valid information is available. Because it is acquired firsthand, knowledge gained through personal experience in clinical practice is often more emotionally resonant than the comparatively pallid reporting of research results that one encounters in the literature. Because more vivid information is more easily retrieved from memory, the application of the availability heuristic provides one avenue by which personal experience can be assigned substantial weight in reaching clinical judgments and decisions.

To grant center stage to one's personal experience, however, can be to devalue the more informative collective experience of many other clinicians who have worked with a much larger and broader sample of clients. Acknowledging the informational value of clinical experience does not give privileged status to *personal* experience relative to the experience of everyone else. Systematic research, for example, constitutes the synthesis of many people's experiences, often a much larger and more representative sampling of pertinent experiences than one has encountered firsthand. In addition, knowledge obtained through personal experience is seldom subjected to adequate statistical testing. As a result, illusory correlations may take root and actual relationships that are in any way subtle or counterintuitive may escape notice. Although theory and research on mental health are far from satisfactory—much less complete—in many important respects, the extant literature can often provide sounder guidelines for practice than a comparatively narrow consideration of one's personal experiences. An exercise such as the following might reveal a double standard of evidence skewed toward the acceptance of one's own experience and the rejection of others' experience:

> Suppose that rather than having had certain experiences and reached a certain judgment myself, someone else presented me with the same conclusion on the basis of the same evidence. That is, the haphazard nature of the sampling, the unavailability of an unknown portion of the original data due to memory limitations and biases, the nonrandom assignment of clients to conditions that vary nonsystematically, the reliability and validity of objective and subjective outcome data (as it is recalled, not as it was initially assessed), and the steps in the reasoning process would be identical to what is going through my mind right now. The only difference would be that I did not personally experience any of this. Rather, I would be learning about the fully equivalent experiences of someone else, stated in unambiguous detail. Would I accept the judgment on these terms?

Through an exercise of this sort, one might remove the *personal* aspect of the relevant experiences and more objectively accord them the weight they merit in the judgment process.

In addition to the potential roles that availability bias and evidential double standards may play, a widespread misunderstanding within the mental health community can serve—intentionally or otherwise—to dismiss the knowledge available from research literature. When the collective experience of clinical investigators is discredited in this way, practitioners are forced to rely more heavily on the anecdotal evidence of their personal experience.

The misunderstanding at issue is captured in the maxim that "probability is irrelevant to the unique individual." Variants of this claim involve the substitution of "statistics" or "research" for "probability." Regardless of its precise phrasing, the idea is that knowledge of the long-run frequency of occurrence for many similar people, under similar circumstances, is of no bearing in a specific situation that is not to be repeated. For example, statistics reported in the research literature suggest that the probability of successfully alleviating an individual's specific phobia is maximized through exposure-based treatment (Barlow, 2002). It is not unusual, however, for a mental health expert to disregard this finding, administering some other treatment (e.g., long-term psychoanalysis) on the grounds that a particular client's case is special—that the probability/statistics/research do not apply to this unique individual. There are two ways of understanding such a claim.

First, one might interpret this as a claim that, despite the clinician's awareness that exposure therapy best addresses specific phobias, he or she perceives something sufficiently probative *in this instance* to countervail the prescribed treatment. Following Meehl's (1954) classic treatise on prediction, this is referred to this as a "broken leg" case: An otherwise sound statistical prediction that a certain professor is likely to attend a movie one evening *should* be modified in light of the fact that the professor had just broken a leg and is in a cast that cannot fit in a movie seat. Despite the existence of such cases, research has revealed that practitioners overidentify "broken leg" counterexamples, departing too frequently from the predictions of a statistical formula derived from real-world outcome data and making more errors in the process (Grove, Zald, Lebow, Snitz, & Nelson, 2000). Meehl (1998) noted that this fact is predictable from the more general finding that, when given the same pool of valid information and evaluated against the same criteria, statistical predictions derived from outcome data are as or more accurate than clinical predictions *even when the clinicians are provided with the statistical predictions and are allowed to copy them*. If clinicians adopted the statistical predictions except in those instances where they could correctly identify exceptions, then their accuracy would be higher than that of the formula. Because this does not happen, the clinicians must be identifying too many exceptions. It is important to recognize what this means: Appeals to the uniqueness of the individual as grounds for

AU: Not in References; please supply.

countervailing the dictates of probability will, on balance, *increase* judgment errors (see Chapter 3 for more on statistical prediction).

Second, one might interpret this as a claim that, *in general*, probability is irrelevant to understanding or predicting the behavior of an individual. A simple thought experiment, originally presented by Meehl (1973), exposes the speciousness of this interpretation. Suppose that you are to play Russian roulette once, meaning that you will put a revolver to your head and pull the trigger. Would you prefer that there be one bullet and five empty chambers in the revolver, or five bullets and one empty chamber? You are, after all, a unique person who will either live or die, and this event will not be repeated. The only basis for preferring that there be just one bullet is that the *probability* of death is one in six rather than five in six. Clearly, probability is extremely relevant despite any unique aspects of this event.

The same reasoning applies when making clinical judgments—present knowledge (based on personal experience or more systematic research) can only establish the conditional probabilities of various outcomes given a certain decision. The rational way to reach important judgments is to choose the option with the best probability of success. Granted, actual clinical work complicates the subjective assessment of probabilities, as it is extremely challenging to identify, gather, and integrate the wealth of information pertinent to making many of the important decisions that arise, and knowledge of the relations between predictors and outcomes is usually quite modest. Nonetheless, the obstacles faced by practitioners do not negate the basic principle—carefully considering probability is *essential* for minimizing the chance of making a judgment error in each unique case.

THE CHALLENGE OF EXPERIENTIAL LEARNING

Expertise in any endeavor requires, among other things, a considerable amount of dedicated practice. Some skills, such as the motor coordination involved in playing a musical instrument, can be improved through repetitive practice exercises. Over time, the automaticity of performance increases and less effort is required to avoid making amateurish mistakes. Other types of skills, such as the creativity involved in composing new works of music, would not benefit from the same sort of repetitive practice. Instead, useful exercises might incorporate trial-and-error explorations of potential melodies, harmonizing, instrumentation, tempo, and so forth. With tasks as multifaceted and open-ended as this, there is no guarantee that experiential learning will occur. Certain requirements must be met, and there may be ways to structure practice sessions to maximize the rewards reaped for a fixed commitment of effort.

Some aspects of the earliest stages of clinical practice, when a large volume of information must be memorized, may bear greater similarity

to the development of motor coordination than the development of musical creativity. An aspiring practitioner must learn about the signs and symptoms of a large number of mental disorders, an ever-expanding collection of assessment tools and treatment techniques (and, ideally, the empirical support for each), and the ethical and legal codes that apply to practitioners in a given locale, for example. Whereas the working vocabulary of mental health practice is acquired through rote learning, many interpersonal skills are honed through experiential learning in supervised training with actual clients and (later) through independent practice. With respect to the development of expert clinical judgment, how effective is experiential learning?

Reducing judgment errors by learning through experience requires attention to concrete, immediate, and unambiguous feedback on the accuracy (or inaccuracy) of prior judgments. Much of the feedback typically available to practitioners, however, is intrinsically ambiguous and temporally distal. For example, if a client does not arrive at several scheduled appointments and remains unreachable thereafter, one could interpret this outcome as a personal failure to form a strong therapeutic alliance. Or, one could assume that the client moved away on short notice and either lost his or her therapist's contact information or forgot to contact the therapist's office. Or, perhaps the client was cured. In either case, the feedback accumulates long after the sessions with this client have ended, and it becomes increasingly difficult to draw firm conclusions about what specific actions may have led to the early termination of therapy.

Practitioners also are exposed to and attend to more positive than negative feedback. Because it can be considerably more interpersonally awkward and difficult, displeased clients can be less likely to communicate blame to their therapists than pleased clients are to express gratitude. At least as important, even when feedback is available, the normal self-serving biases of human judgment can mount a variety of defenses against ego-threatening information while allowing more flattering information to arrive unfettered (Faust, 1986). Moreover, hindsight bias can make poor outcomes seem inevitable rather than the result of judgment errors. Even if a case is handled badly and therapeutic change is either nil or negative, there are many ways that a clinician can deflect this otherwise negative feedback. For instance, one might console oneself with the fact that the prognosis is poor for individuals suffering from chronic posttraumatic stress disorder, and especially poor for those with a comorbid substance abuse disorder. Even the most honestly self-critical therapist may not be able to distinguish the effects of some subtle errors in judgment from the effects of prior difficulties that ordinarily are not amenable to treatment. The net result of ambiguous feedback, time delay in the receipt of feedback, the scarcity of negative feedback, and hindsight bias is that there may be precious few opportunities to learn through experience.

Given these factors, it should not be terribly surprising that the accuracy of clinical judgment tends not to improve with clinical experience (Garb,

1989; see also Chapter 3 for a discussion of clinical judgment). This is consistent with what Dawes (1994) refers to as the *myth of expanding expertise.* Many people—clinicians included—simply assume that skills improve with experience and fail to consider the requirements for such learning to occur. Whereas skills acquired through rote memorization can be assessed relatively directly and easily, those built through experiential learning are considerably more difficult to assess. In place of reliable and valid measures of genuine improvement in clinical judgment, the myth of expanding expertise may fill the void with the presumption of gains attributable to experiential learning.

PROMOTING EXPERIENTIAL LEARNING

A return to the world of competitive Scrabble suggests some strategies that might be adapted to promote more effective experiential learning in clinical practice. As in chess, Scrabble players are provided with numerical ratings of their skill level. These ratings, updated with each game played, are calculated based on such factors as the outcome of the game and the skill level of the opposing player. Given the psychometric proficiency of psychologists and others in related disciplines, it is not inconceivable that a rating system could similarly be devised to quantify therapists' track records. Although clients are not directly analogous to opposing players, a good rating system could account for clients' current mental health, history, and complicating factors so that therapists who succeed with more difficult cases earn higher ratings. It is easy to imagine abuses of a rating scheme, but it also is possible to imagine beneficial uses of a well-constructed system, especially if access to ratings is appropriately restricted to those with educational, training, research, or other approved purposes. Clinical trainees and less effective therapists could seek opportunities to learn from expert mentors, and researchers could study expert therapists for clues about how they achieve their success. Particularly if such a system were developed and maintained by mental health professionals themselves, much might be learned about therapeutic success and truly expert clinical judgment. At the same time, the increasing demand for health care accountability suggests that the imposition of a rating system on therapists by insurers or govern-ment agencies is not out of the question. This possibility may provide some incentive for clinicians to devise a satisfactory system of their own before being forced into one that they find less palatable.

Perhaps more striking than the quantification of Scrabble players' expertise are some of the behavioral differences between Scrabble novices and experts observed by Fatsis (2001). Whereas beginners tend to clear a board and begin a new game quickly after one has ended, presumably believing that the best way to improve their play is through practice, experts study each game for opportunities to prevent the repetition of

suboptimal plays in the future. In addition to studying the board itself at the end of the game, an expert takes meticulous notes on each play so that it can be evaluated in the context of the game at that moment. As Meehl (1997) and others argued, clinicians might learn more effectively if they tabulated and quantified their experiences. This practice could be useful for the generation and testing of hypotheses in real-world contexts. Even if not done formally as a research project, more informal tallies of the frequencies with which certain types of hunches or approaches do and do not bear fruit, or with which certain variables do and do not co-occur, could be highly informative for oneself or others.

Another tool that is increasingly used by current and aspiring Scrabble experts is to compare actual or hypothetical plays with the "optimal" plays generated by an expert system. These plays are optimal in the sense that a computer program—provided with complete Scrabble word knowledge and algorithms to score plays—can determine, probabilistically, what play is likely to yield the best final game score margin across a large number of games that all begin with precisely the same specifications (e.g., layout of tiles on the board, each player's current score, one or both players' racks of tiles). One can use such a program to ask whether a certain play is optimal or whether the computer can devise a better play, or one can compare two or more alternative plays (e.g., playing a word, playing a different word, exchanging certain tiles) to learn which would have been best. The ability to simulate follow-up data to evaluate every judgment is a powerful tool for Scrabble players to exploit. Clinicians do not have the same opportunity, but just as they could tally observations for subsequent analysis, they could take better advantage of opportunities to gather systematic data on various criterion measures with which to evaluate critical judgments retrospectively. Such criterion data could be collected on an ongoing basis, at termination, or subsequent to termination.

A final recommendation for improving judgment is not only consistent with observations of expert Scrabble players but also strongly supported by the literature on correcting judgment errors and overconfidence (e.g., Arkes, 1991). Scrabble experts are continually searching for weaknesses in their own play, striving to grow as players through ruthless self appraisal. A key component of their success in learning through experience is the use of hypothetical counterfactuals such as "What mistakes have I made?" and "How might I prevent similar errors in the future?" In clinical work, one could examine cases with especially poor outcomes (e.g., the death of Candace Newmaker during rebirthing therapy; Mercer, 2002) to formulate hypotheses about how to prevent harmful judgment errors. Janis (1972) used this approach to identify the groupthink phenomenon as a culprit in many disastrous foreign policy decisions and recommended the institutionalization of a "devil's advocate." Of course, one can adopt that role with regard to one's own judgment. Like everyone else, clinicians are in a position to learn more about their trade by habitually asking themselves "Why might I be wrong?" (see also Chapter 1 for a discussion of science as a process of error-elimination).

CONCLUSION: TIPS FOR REDUCING JUDGMENT ERRORS

An understanding of the cognitive biases and logical fallacies discussed in this chapter suggests a number of concrete steps that can be taken to minimize judgment errors in clinical practice.

1. *Scrutinize similarity-based arguments.* Because the representativeness heuristic can make claims taking the form "like goes with like" appear quite reasonable, one must be especially careful to evaluate the logic and evidence bearing on such assertions.
2. *Conceptualize problems in multiple ways.* The availability heuristic can lead one astray when the instances most easily retrieved from memory provide a biased sample of data. Reconceptualizing an issue may provide new memory cues that elicit complementary information that reduces the initial bias and provides a firmer basis for reaching a judgment.
3. *Formulate and test multiple working hypotheses.* To prevent the premature closure that can result from the operation of confirmation and hindsight biases, it is important to generate multiple hypotheses and to tease them apart rigorously. Deliberately constructing and evaluating plausible alternative explanations can prevent many of the judgment errors resulting from a search for information to support an impression that was formed quickly.
4. *Recognize that personal experience is anecdotal evidence.* It is all too easy to allow personal experience to disproportionately influence clinical judgments. Whereas research systematically aggregates the experience of many practitioners with many clients, one's personal experience may involve a smaller, more haphazard, and less rigorously evaluated knowledge base. Considering whether one's own conclusions would be acceptable if presented by someone else may help to identify instances in which personal experience is being given undue weight (see also Chapter 7 for a discussion of anecdotal evidence in clinical science).
5. *Learn and apply basic principles of probability.* Because clinical work involves probabilistic relationships between variables, practitioners need to recognize that probability, statistics, and research evidence do apply to unique individuals. At least as important is learning the basic rules of probability and knowing when and how to apply them (e.g., using Bayes' Theorem to combine base rates with individuating information).
6. *Identify exceptions to statistical trends with caution.* A statistical trend represents a "signal" that can be detected despite the "noise" of individual differences and contextual variables. Although judgments informed by such trends will not be accurate in all cases, the literature strongly suggests that practitioners identify too many exceptions. Judgment errors can result from attaching too much significance to a client's uniqueness, which is often of little predictive value precisely because it is impossible to establish statistical associations involving truly unique characteristics. Discovering meaningful ways in which a client's

case shares features with others enables a savvy practitioner to more successfully play the odds by taking advantage of statistical trends.
7. *Play "devil's advocate" to one's own judgments.* Finally, asking why one might be wrong can suggest the need for additional information, help to differentiate between relevant and irrelevant information, or lead to a more appropriate way to integrate the available information when reaching a judgment. The more one learns about the limitations and biases of human reasoning, the more opportunities are afforded to prevent judgment errors by actively checking for mistaken premises or faulty logic in one's own thinking.

KEY TERMS

Availability heuristic: A mental shortcut for judging of the probability or frequency of an event by using the ease with which instances can be retrieved from memory as a guide.

Confirmation bias: The tendency to selectively seek, attend to, or attach greater weight to information that supports rather than refutes one's beliefs.

Experiential learning: The development of expert knowledge or judgment through a process that requires concrete, immediate, and unambiguous feedback.

Hindsight bias: The presumption that the ability to construct a plausible explanation of past events implies a causal understanding that can be used to successfully predict future events.

Representativeness heuristic: A mental shortcut for reaching judgments based on perceived similarity or "goodness of fit" rather than actual probabilistic or causal relationships.

RECOMMENDED READINGS

Faust, D. (1986). Research on human judgment and its application to clinical practice. *Professional Psychology: Research and Practice, 17,* 420–430.

Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment.* Washington, DC: American Psychological Association.

Meehl, P. E. (1973). Why I do not attend case conferences. In P. E. Meehl (Ed.), *Psychodiagnosis: Selected papers* (pp. 225–302). Minneapolis: University of Minnesota Press.

Meehl, P. E. (1997). Credentialed persons, credentialed knowledge. *Clinical Psychology: Science and Practice, 4,* 91–98.

Turk, D. C., & Salovey, P. (1988). *Reasoning, inference, and judgment in clinical psychology.* New York: Free Press.

REFERENCES

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin, 110,* 486–498.

Beyerstein, B. L. (1997). Why bogus therapies seem to work. *Skeptical Inquirer, 21,* 29–34.

Barlow, D. H. (2002). *Anxiety and its disorders: The nature and treatment of anxiety and panic* (2nd ed.). New York: Guilford.

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

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic observations. *Journal of Abnormal Psychology, 74*, 271–280.

Dawes, R. M. (1993). Prediction of the future versus an understanding of the past: A basic asymmetry. *American Journal of Psychology, 106*, 1–24.

Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York: Free Press.

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

Fatsis, S. (2001). *Word freak: Heartbreak, triumph, genius, and obsession in the world of competitive Scrabble players*. Boston: Houghton Mifflin.

Faust, D. (1984). *The limits of scientific reasoning*. Minneapolis: University of Minnesota Press.

Faust, D. (1986). Research on human judgment and its application to clinical practice. *Professional Psychology: Research and Practice, 17*, 420–430.

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 288–299.

Fischhoff, B. (1980). For those condemned to study the past: Reflections on historical judgment. In R. A. Shueder & D. W. Fiske (Eds.), *New directions for methodology of behavioral science: Fallible judgment in behavioral research* (pp. 79–93). San Francisco: Jossey-Bass.

Garb, H. N. (1989). Clinical judgment, clinical training, and professional experience. *Psychological Bulletin, 105*, 387–396.

Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association.

Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19–30.

Halford, G. S., Baker, R., McCredden, J. E., & Bain, J. D. (2005). How many variables can humans process? *Psychological Science, 16*, 70–76.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107*, 311–327.

Janis, I. L. (1972). *Victims of groupthink: A psychological study of foreign-policy decisions and fiascos*. Oxford, England: Houghton Mifflin.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. C*ognitive Psychology, 3*, 430–454.

Lilienfeld, S. O., Lynn, S. J., Kirsch, I., Chaves, J. F., Sarbin, T. R., Ganaway, G. K., & Powell, R. A. (1999). Dissociative identity disorder and the sociocognitive model: Recalling the lessons of the past. *Psychological Bulletin, 125*, 507–523.

Lynn, S. J., Lock, T., Loftus, E. F., Krackow, E., & Lilienfeld, S. O. (2003). The remembrance of things past: Problematic memory recovery techniques in psychotherpy. In S. O. Lilienfeld, S. J. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. 205–239). New York: Guilford.

AU: Please confirm that this is cited in text.

Meehl, P. E. (1973). Why I do not attend case conferences. In P. E. Meehl (Ed.), *Psychodiagnosis: Selected papers* (pp. 225–302). Minneapolis: University of Minnesota Press.

Meehl, P. E. (1992). Cliometric metatheory: The actual approach to empirical, history-based philosophy of science. *Psychological Reports, 71*, 339–467.

Meehl, P. E. (1997). Credentialed persons, credentialed knowledge. *Clinical Psychology: Science and Practice, 4*, 91–98.

Meehl, P. E. (1998, May). *The power of quantitative thinking.* Invited address as recipient of the James McKeen Cattell Award at the annual meeting of the American Psychological Society, Washington, DC.

Mercer, J. (2002). Attachment therapy: A treatment without empirical support. *Scientific Review of Mental Health Practice, 1*, 105–112.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175–220.

Nisbett, R. E., & Ross, L. 1980. *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice-Hall.

Poole, D. A., Lindsay, D., Memon, A., & Bull, R. (1995). Psychotherapy and the recovery of memories of childhood sexual abuse: U.S. and British practitioners' opinions, practices, and experiences. *Journal of Consulting and Clinical Psychology, 63*, 426–437.

Renaud, H., & Estess, F. (1961). Life history interviews with one hundred normal American males: "Pathogenicity" of childhood. *American Journal of Orthopsychiatry, 31*, 786–802.

Ruscio, J., & Stern, A. R. (in press). The consistency and accuracy of holistic judgment: Clinical decision making with a minimally complex task. *Scientific Review of Mental Health Practice.*

Sagan, C. (1995). *The demon-haunted world: Science as a candle in the dark.* New York: Random House.

Schreiber, F. R. (1973). *Sybil.* Chicago: Henry Regnery.

Spanos, N. P. (1996). *Multiple identities and false memories: A sociocognitive perspective.* Washington, DC: American Psychological Association.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance II* (pp. 276–315). Amsterdam: Elsevier-North Holland.

Turk, D. C., & Salovey, P. (1988). *Reasoning, inference, and judgment in clinical psychology.* New York: Free Press.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*, 207–232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

Wood, J. M., Nezworski, M. T., & Lilienfeld, S. O. (2003). *What's wrong with the Rorschach? Science confronts the controversial inkblot test.* San Francisco: Jossey-Bass.

AU: Please see earlier query: Garb?