The Journal of Mind and Behavior
Autumn 1993, Volume 14, Number 4
Pages 297–316
ISSN 0271–0137

Diagnostic Reasoning and Reliability: A Review of the Literature and a Model of Decision-making

Jonathan Rabinowitz

Bar Ilan University

A review of mental health practitioners' decision-making biases is presented that integrates the diverse literature in the area. Previous reviews have considered only the effects of single biases and have not looked at multiple biases in mental health practitioner decision-making. The biases are reviewed relative to a four stage schema of clinical judgment (a) input, (b) processing, (c) output–action, and (d) feedback. Each stage is influenced by background variables that are also reviewed, including the effects of client biases, oversights, clinician mood, theoretical orientation, values and setting on diagnosis and clinical judgment. This review points to the immediate need for changing practice and for additional research.

Studies over the last 30 years have identified shortcomings in diagnostic reasoning of mental health practitioners. Such shortcomings can result in inappropriate treatment, unjust institutional confinement, poor custody arrangements, and unjustified monetary awards. They can also result in death, as in failure to diagnose mental disability in capital-punishment cases or in failure to diagnose fatal physical illness. Questions of reliability and validity of judgments by mental health experts have led researchers (Faust and Ziskin, 1988; Ziskin and Faust, 1988) to critical evaluation of the scientific and forensic credibility of their diagnoses.

Mental health practitioners frequently disagree with each other and with actuarial methods in diagnosis, disposition, and prediction of outcome. Research shows that clinicians often give unreliable and invalid diagnoses and predict behavior as inaccurately as lay people (see section on effect of training, below). Plag and Arthur's (1965) study of U.S. Naval recruits recommended for psychiatric discharge illustrates clinicians' difficulty in predicting

Requests for reprints should be sent to Jonathan Rabinowitz, DSW, School of Social Work, Bar Ilan University, Ramat Gan, ISRAEL.

behavior. For research purposes, these 134 recruits were not discharged, but retained in the Navy for two years of study. After two years seventy-two percent of these "psychiatrically unfit" soldiers remained on active duty compared to 85.8% of the psychiatrically "healthy" group (see Matarazzo [1978] who contends that this study demonstrates that clinicians can predict behavior). Problems with reliability of diagnosis are illustrated in Boxer and Garvey's (1985) study of Cuban refugees initially denied immigration to the United States on psychiatric grounds. In that study, 57% of 109 refugees denied immigration were allowed to enter after reexamination by another psychiatrist found them mentally healthy.

The shift to behaviorally based DSM-III, Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 1980) and the revised DSM-III (American Psychiatric Association, 1987), has not substantially improved levels of diagnostic agreement. Levels of agreement among clinicians using the DSM-III and revised DSM-III rarely reached the "generally good level" established by the authors of the DSM-III. (For a review of the reliability of Axis I and Axis II see Drake and Vaillant [1985]; Flaum, Arndt, and Andreasen [1991]; Fuhrer, Rouillon, Lellouch [1986]; Heumann and Morey [1990]; Hjortso, Butler, Clemmesen, Jepsen, Kastrup, Vilmar, and Bech [1989]; Lieberman and Baker [1985]; Mellsop, Varghese, Joshua, and Hicks [1982]; and for children see Newcorn and Strain [1992]; Vitiello, Malone, Buschle, Delaney, and Behar [1990]. For a review of the reliability of Axis IV see Skodol [1991]; and for a review of the reliability of Axis V see Fernando, Mellsop, Nelson, Peace, and Wilson [1986].)

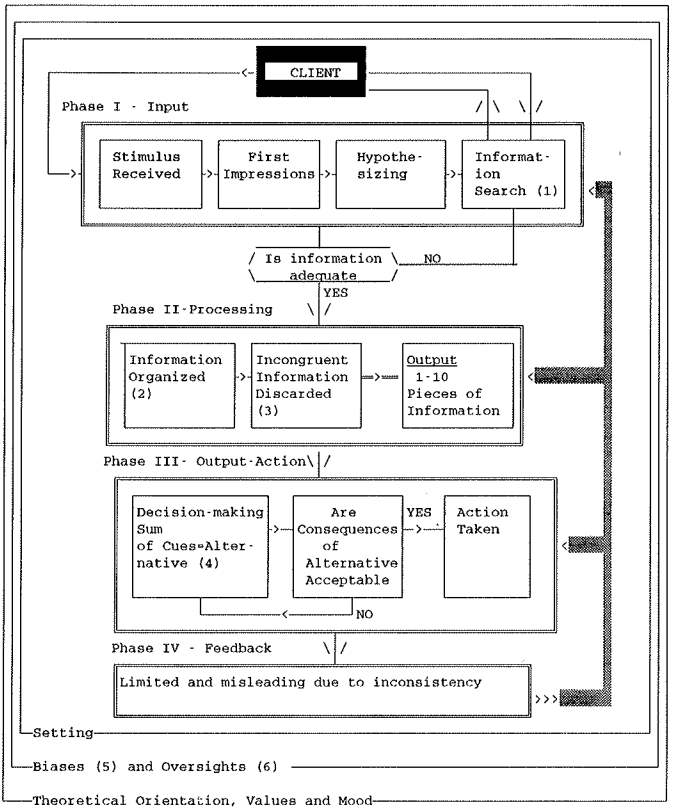
Kutchins and Kirk (1986) reexamined levels of agreement reported in DSM-III and found them to be excessively high. They also reviewed the studies from 1980 through 1985 and concluded that most diagnostic categories have poor reliability. More favorable reviews, like that of Matarazzo (1990), argue that the situation is improved from 30 years ago, when "almost all" studies found poor reliability. In his review of literature prior to 1983, Matarazzo found that only half of the studies had adequate levels of reliability, though he does not define "adequate." Curiously, the revised DSM-III (1987) does not report reliability coefficients.

This paper presents an integrated review of biases in clinical judgment. Other investigators have reviewed only single biases to clinical judgment, for example, cognitive limitations (Arkes, 1981; Faust, 1986; Nurius and Gibson, 1990; Turk and Salovey, 1987), and gender (Brown, 1990). This article is designed as a springboard for furthering understanding of biases in order to improve clinical judgment.

Schema of Clinical Decision-making

This review integrates mental health clinical-judgment research and highlights shortcomings. A flow sheet summarizing shortcomings is presented in Figure 1. Clinical judgment is divided into four phases: (a) input, (b) processing, (c) output and action, and (d) feedback. (The phases are similar to Hogarth's [1980] conceptual model of human judgment.) The outcome of each phase depends on variables such as setting, clinicians' values, theoretical orientation, mood, biases, and oversights.

Figure 1
Flow Sheet of Clinical Decision-making.



(1) supporting hypothesis, counterevidence overlooked; (2) typically around the initial hypothesis; (3) anecdotal information preferred over empirical; (4) problematic strategies used include: representativeness, availability, conjunction, ignoring probability, mistaken sign/disorder co-variation; (5) e.g., collegial opinion, clients': race, SES, gender, culture, physical attractiveness; (6) i.e. conditions overlooked, e.g., physical illness, sexual abuse, addictions.

Background

Setting: Decision Milieu

Settings (see background of Figure 1) influence judgment because practitioners expect to see specific types of clients in specific places. For example, in the emergency room at a psychiatric hospital clinicians expect to find a seriously disturbed person. Such typical cases can anchor clinical judgment, with clinicians initially assuming that the person is disturbed, and thus they are not likely to change their clinical judgment. Anchoring is not always harmful. For example, when clinician expectations approximate base rates of a given condition, anchoring can improve accuracy.

Different settings have different possible response alternatives (Bieri et al., 1966). In most settings pathology is expected and tends to be found. Clinicians tend to seek evidence supporting pathology. This can lead to locating non-existent pathology. In Temerlin and Trousdale's study (1969; see also Temerlin, 1970), every psychiatrist diagnosed a well-adjusted job applicant as having a disorder. When the subject was presented as a patient, respondents diagnosed even more severe psychopathology. Langer and Abelson (1974) report similar findings.

Biases

Biases (see background of Figure 1) influence all phases of judgment, including filtering and amplifying information. Typical sources of bias include patients' race, socioeconomic status, gender, culture, physical appearance, and the opinions of colleagues about the patient.

Collegial opinion. The opinion of a colleague or supervisor about a patient can lead to preconceived first impressions (MacKinnon and Michels, 1971). In one study, clinicians who viewed a filmed interview after being told that colleagues had diagnosed psychosis, subsequently diagnosed the filmed subject with significantly more pathology than clinicians not told others' opinions. The same interview, shown as a screening for employment, yielded even less pathology. Thus, aside from being influenced by another expert's opinion, judgments were also anchored by setting (Temerlin, 1970; Temerlin and Trousdale, 1969).

Client variables. L'opez (1989) thoroughly reviewed studies of patient-variable biases. His analysis reveals that there is as much evidence for bias as there is against bias. He broadens the concept of bias from overpathologizing and overdiagnositic biases to include minimizing and underdiagnostic biases. In order of most to least consistent evidence of bias, he found: (a) consistent evidence for underdiagnostic bias of mentally retarded people; (b) social class bias — over three fourths of the studies show possible clinical error for sever-

ity and diagnostic judgment, typically overpathologizing lower social class people; (c) race bias was found in six of nine cases; (d) gender bias, according to L'opez, had the least consistent evidence. Yet in a review of gender effects Brown (1990) concludes that gender bias leads to inaccurate assessment and diagnosis.

Cultural factors. Normative behavior is a yardstick for measuring deviant behavior. Because different cultures have different behavioral norms, behaviors considered pathological in one culture may be normal in another. Time and official sanction can change the definition of deviant behavior. Before the American Psychiatric Association voted on changes in the DSM-III, homosexuality was deviant behavior and a psychiatric disorder. After a vote to replace homosexuality with ego dystonic homosexuality, many homosexuals were no longer considered behaviorally deviant or mentally disturbed. Rosenhan (1975, p. 464) notes that "psychiatric diagnoses are maintained by consensus alone," not on scientific criteria.

The role of culture bias was demonstrated in a study comparing Lao ba folk diagnosticians and two groups of psychiatrists. One group of psychiatrists was conversant with the Laos culture and the other group had no Laotian experience. The diagnoses of psychiatrists conversant with Laotian culture showed greater agreement with those of folk diagnosticians than with those of the psychiatrists unfamiliar with Laotian culture (Westermeyer and Zimmerman, 1981). This underscores that cultural norms can be a basis for making clinical judgments. (See also Baskin, 1984, who compared diagnoses from different countries to identical vignettes.)

Physical attractiveness. Physically attractive individuals are judged as having better self-concept (Hobfoll and Penner, 1978), more favorable prognosis (Barocas and Vance, 1974), and less psychopathology (Cash, Kehr, Polyson, and Freeman, 1977) than less attractive patients. This is similar to Schoefield (1964) who found that preferred clients were "young, attractive, verbal, intelligent, and successful" (YAVIS) clients.

The presence of physical disabilities can bias clinical judgment. Dickert (1988) compared treatment decisions concerning hearing versus hearing-impaired patients. Clinicians rated the most severely disturbed hearing-impaired patients less ill than their hearing counterparts. Yet, more supervision and medication were recommended for the least disturbed hearing-impaired patients than for the hearing patients.

Oversights

Physical illness. Studies suggest that mental health practitioners tend to overlook physical illness. They also occasionally overlook psychiatric symptoms in the physically ill and disabled. In fact there are many legal cases of

inappropriate psychiatric confinement of the physically ill, e.g., New York Association for Retarded Children vs. Carey (Perlin, 1989). Physical illness can divert attention from psychiatric symptoms (Flaherty and Flaherty, 1983) or be mistakenly attributed to psychogenic causes. Koranyi (1979) medically examined 2,090 mental health outpatients. He found that 43% had at least one physical illness. Nearly half of their illnesses had been undiagnosed. Bartsch, Shern, Feinberg, Fuller, and Willet (1990) report similar findings and add that 16% of the patients had conditions that could cause or exacerbate mental disorders.

Sexual abuse. Despite the recognized traumatic effects of sexual abuse, it is often ignored in clinical practice (Gelinas, 1983). In a study by Craine, Henson, Colliver, and MacLean (1988) 51% of 105 female state-hospital patients told a trained interviewer that they had been sexually abused as children (at least once before age 18). More than half of these women had not been identified as abuse victims by staff (Jacobson and Herald, 1990, report similar findings). Only when abuse was a cause for hospital admission (e.g., adolescents admitted because of abuse at home) was it correctly identified. The authors report that 36% of the women met the DSM-III criteria for post-traumatic stress disorder because of sexual abuse. However none was so diagnosed. Staff reported that sexual abuse was rare among patients. However, patients reported that staff had never asked them about abuse.

Substance abuse. This sometimes goes undetected. Alcoholism is a common undiagnosed condition (Smith, 1983). Kanwischer and Hundley (1990) report that while questionnaires revealed that 56% of their 144 psychiatric inpatients were substance abusers, only half of them received treatment for substance abuse.

Underdiagnosis of substance abuse is sometimes attributed to patients' alleged denial of it. Safer (1987, p. 514), concluded that "Since lying about substance abuse is so common among young adult chronic patients. . ." special procedures are needed to obtain accurate information about them. The "lying" hypothesis has not been supported by many studies that have found that self-report measures of alcohol and drug use are valid measures of substance abuse (Barnea, Rahav, and Teichman, 1987; Benson and Holmberg, 1985; Johnston and O'Malley, 1985) even among prisoners (Wish and O'Neil, 1989). This suggests an alternate hypothesis that clinicians may overlook substance abuse and thus do not inquire about it.

Theoretical Orientation, Values and Mood

Other background influences on clinical judgment are practitioners' theoretical orientations, values and moods. Theoretical orientation guides the discovery of material and as such is essential to clinical evaluation; yet, it can be limiting when only material that supports the preferred theory is discov-

ered (Snyder and Thomsen, 1987). Because values suggest preferred courses of action and outcome they too can bias judgment: a practitioner who holds beliefs about the "appropriate role for women" might be biased in recommending child custody (Ziskin and Faust, 1988). Mood states bias judgments because clinicians tend to find and recall information that is consistent with their mood state (Salovey and Turk, 1987). The same client may be viewed differently depending on differences in background.

Phase I: Input

Sub-phase 1: Initial Stimulus Reception: Presenting Information

The first phase of the schema of clinical judgment presented in Figure 1 is the input phase. The input phase is like recognizing a puzzle picture, with only certain pieces revealed. The puzzle pieces are the information presented to the practitioner about the client. Clinicians ask to see other pieces based on their perceptions of these pieces. They stop searching for pieces when they think they can recognize the picture. The input phase ends with a working hypothesis.

Beginning at the top of the schema in Figure 1, input begins when stimulus is received. Stimulus, information about the client, is from a case record or is presented by the client. Client information includes verbal cues, appearance, and voice. Based on these, clinicians form first impressions. Typically client interview behavior — a small sample of behavior — is considered to be representative of the patient's usual behavior. Small samples are mistaken as being equally representative as large samples (Tversky and Kahneman, 1971). The interview sample can thus lead to incorrect impressions because it may not be representative of the patient's behavior.

Further samples of behavior can lead to greater misunderstanding because most peoples' backgrounds are rich enough with information to both support a "healthy" view and a "pathological" view. As Renaud and Estess (1961) state, pathology can be supported even when normal individuals are asked to describe their lives. They can reveal information that is not consistent with their usual behavior and supports a diagnosis of mental illness.

Sub-phase 2: First impressions: Perceiving the Puzzle

First impressions are formed rapidly, sometimes as quickly as 30 seconds, and appear to change little even in the face of contradictory information (Gauron and Dickinson, 1969; Sandifer, Hordern, and Green, 1970). This may be because the first revealed puzzle pieces, those behaviors of the client noticed first, orient perception of the puzzle. Subsequent puzzle pieces revealed or observations made then have much less effect.

First impressions anchor judgments and are the basis of initial hypotheses (Hogarth, 1980). Some researchers find that first impressions are significant predictors of later judgments (Nisbet and Ross, 1980; Richards and Wierzbicki, 1990). Other researchers (Ellis, Robbins, Schult, Ladany, and Banker, 1990; Klayman and Ha, 1987) find no evidence to support the anchoring effects of first impressions. In those studies clinicians changed their view of clients as they received new information.

Tversky and Kahneman (1974) illustrate the effects of first impressions on intuitive numerical estimations. Two groups of subjects were given five seconds to estimate the product of different numerical expressions. The first group was presented with 8X7X6X5X4X3X2X1 and the second group was presented with 1X2X3X4X5X6X7X8. The product of these expressions is 40,320. The median estimate for the first group was four times larger (2,250) than for the second group (512). Similarly, Pain and Sharpley (1989) report that by varying the order in which positive and negative information is presented, clinicians' judgments of severity of condition varied, with "bad" information presented early overshadowing the effects of "good" information presented later. These studies demonstrate the strong influence of first impressions on judgment.

Sub-phase 3: Hypothesizing

First impressions are transformed, via theoretical orientation, into initial hypotheses. These hypotheses explain the client based on the overturned puzzle pieces, i.e., information cues. Clinicians appear to generate few hypotheses. For example, Elstein, Shulman, and Sparfka (1978) found that physicians and medical students, responding to simulated cases, generated no more than seven hypotheses (with a mode of five) per case, regardless of the amount of information provided. Selected findings were ignored and general hypotheses were generated. The clinicians in that study also added newly collected information to support existing hypotheses rather than generating new hypotheses.

Sub-phase 4: Information Search

Clinicians next search for additional information. Sometimes this is to support, *not* test, hypotheses. Counterevidence is often ignored (Gauron and Dickinson, 1969; Ross, Lepper, and Hubbard, 1975; Strohmer, Shivy, and Chiodo, 1990) giving clinicians a false sense of confidence (Dawes, 1986; Ingram, 1986). The rapport between client and clinician probably also affects information search. Clients who feel at ease are likely to reveal more information than those who feel uneasy or suspicious.

Phase II: Processing

After the clinician collects enough information to form a judgment, the processing phase begins (see Figure 1, Phase II). Information appears to be organized around a central theme, typically the clinicians' initial hypothesis, and synthesized to support that hypothesis by using confirmatory strategies, for example, selective overweighing of confirmatory evidence (Faust, 1986).

Another shortcoming in how clinicians process information is that they sometimes prefer anecdotal (e.g., case studies) over systematic (e.g., statistical) information (Anderson, 1983; Taylor and Thompson, 1982). Similarly, clinicians rely on diagnostic tests like the DAP, Draw-A-Person Test, that are conceptually appealing but not valid based on empirical studies. In these studies (Chapman and Chapman, 1967, 1969; Wanderer, 1969) clinicians using the DAP were unable to differentiate between groups of subjects who had great differences in levels of psychopathology.

Chapman and Chapman (1967, 1969) demonstrate inaccuracies resulting from practitioners' reliance on impressions. They studied the validity of clinical impression by randomly pairing diagnostic signs and test responses with patients descriptions. In one study [1967], DAP drawings were randomly paired with descriptions of patients. Clinicians and laypersons were asked to find connections between symptoms and drawings. Both groups reported the same mistaken connections. Because the drawings and patients were randomly matched, it is not likely that there was a valid connection between the two. In a later study [1969], practitioners and undergraduate students concluded homosexuality from the same invalid interpretations of the Rorschach. Both groups failed to notice valid Rorschach signs of homosexuality. The authors suggest that clinicians use signs that have strong associative connections but lack validity.

Clinicians appear to synthesize information because it is difficult to contend with incongruous data. Miller (1956) calls this "chunking": combining information into a congruent whole to avoid overload of stimuli. Some cues are thus ignored. The processing phase ends with a few pieces of synthesized information.

Phase III: Output-Action

In the next stage illustrated in Figure 1, the Output–Action phase, processed information is used to select a particular alternative, diagnosis, treatment, or recommendation. As noted previously, which alternative is chosen is a function of its popularity in the setting. Alternatives are weighed based on information discovered, which is influenced by theoretical orientations, values, and mood. Choice is later weighed against consequences. The antici-

pated effects of the decision (e.g., patient will never be given a chance in life if diagnosed as having schizophrenia) sometimes modifies the decision.

Sub-phase 1: Summing Cues

Diagnostic decisions appear to be simply made (Arkes and Hammond, 1986). Studies have been able to model decisions best by using simple linear models (Goldberg, 1968; Hammond and Summers, 1965; Hoffman, Slovic, and Rorer, 1968; Rorer, Hoffman, Dickman, and Slovic, 1967; Shapiro, 1989; Wiggins and Hoffman, 1968). Research has not found judges who were able to perform complex configural analysis (for a review see Slovic and Lichtenstein, 1971). This suggests that decision-making is a linear process.

Only a few cues appear to be used in making a decision. Studies have found that decision makers typically use fewer than seven cues (Bieri et al., 1966; Miller, 1956; Oskamp, 1962a). After decision makers reach their saturation point, which appears to be early, additional information has little effect. For example, Oskamp (1962a) gave psychologists 13 MMPI scales together with the patient's age and education. The psychologists used only five pieces of information and one piece dominated. These findings run counter to clinical wisdom that more information leads to greater understanding of the client.

Faulty Strategies

Clinicians sometimes use faulty strategies in drawing inferences and making decisions. These faults, discussed below, are: (a) reliance on representativeness; (b) reliance on availability; (c) conjunction fallacy; (d) inability to incorporate probability into judgments; and (e) mistaken judgments of sign disorder co-variation. The influence of experience and training on accuracy in judgment is also discussed.

Representativeness. Frequency estimates are sometimes based on representatives, not on probability (Dawes, 1986; Tversky and Kahneman, 1974, 1983). When this occurs the likelihood of an event happening is judged by recalling a schema. The current situation is compared to the schema, not to the probability of the phenomenon. An example of this was demonstrated by Dawes (1986) where a group of practitioners were asked to predict choice of graduate school of an aloof student with interpersonal problems. The practitioners were given a choice between graduate school in library science or education. Sixty percent answered library science despite the fact that most of the group knew that there are many more graduate students in education than library science. This shows that the practitioners were not able to use their knowledge of probabilities, which would have led them to choose education—

instead they relied on a stereotypical view of librarians as being more aloof and as having more interpersonal problems than teachers.

Similar to the concept of representativeness is the concept of availability (Tversky and Kahneman, 1973): the tendency to make decisions based on cases that are easiest to recall. An example is the therapist who attempts to predict the likelihood of a client committing suicide by retrieving the most easily remembered prototype client as the basis for prediction (see Lichtenstein, Slovic, Fischhoff, Layman, and Combs, 1978). Even seasoned clinicians seem not to predict in a consistent fashion; recent and easier-to-recall cases anchor judgment.

Conjunction fallacy. The conjunction fallacy occurs when a special case is judged to be more probable than an initial case (Tversky and Kahneman, 1974, 1983). For example, in one study clinicians were asked to judge whether a particular college graduate was a bank teller or a bank teller and active in the feminist movement. Clinicians were told that subject was outspoken, bright, a philosophy major, deeply concerned about social justice, and participated in anti-nuclear demonstrations. Eighty-five percent responded with bank teller in feminist movement. This was the case despite the decreased likelihood introduced by the additional condition of being a feminist.

Ignoring probability. Clinicians frequently make a diagnosis based on the presence of diagnostic signs or symptoms without considering base rates. Sometimes signs are less accurate than guessing (Faust, 1986). For example, a given psychiatric test may screen for military unsuitability accurately 60% of the time. However, if more than 60% of the population is suitable, the test will be less valuable than using base rates.

Sign disorder co-variation. Judging the relationship between two variables is another difficulty that clinicians have. Does sign "A" relate to disorder "B" or does treatment "A" relate to outcome "B"? Co-variation is commonly determined by looking for cases in which both the sign and the disorder are present. This does not consider cases in which the disorder is present and the sign is absent. It is thus misleading. If the disorder does not occur more frequently when the sign is present than not, the two are not related (Faust, 1986). Consider for example, the practitioner who concludes that because 80% of his or her patients raised in single parent homes are divorced, there is a relationship between divorce and upbringing in a single parent home. However, without examining the patients who come from intact families, 80% of whom may also be divorced, there is no evidence to support this relationship. Individuals from single parent homes, then, would be no more likely to be divorced than other people.

Effect of Training

Surprisingly, training and experience do not necessarily lead to judgmental accuracy (Faust, Guilmette, Hart, and Arkes, 1988 as quoted in Faust and Ziskin, 1988; Goldberg, 1959; Hiler and Nesvig, 1965; Johnston and McNeal, 1967; Leirer, Werner, Rose, and Yesavage, 1984 as cited in Faust and Ziskin, 1988; Levy and Ulman, 1967; Luft, 1950; Oskamp, 1962b, 1967; Schaeffer, 1964; Silverman, 1959; Stricker, 1967; Walters, White, and Greene, 1988; Werner, Rose, and Yesavage, 1983). In some studies, laypersons predicted behavior as well as clinicians (Walters, White, and Greene, 1988; Ziskin and Faust, 1988) or even better (Stricker, 1967). Secretaries, for example, distinguished between normal and brain damaged individuals' visual–motor productions on a common screening test as accurately as psychologists (Goldberg, 1959) and interpreted drawings as well in another report (Schaeffer, 1964). One study compared neuropsychologists who differed significantly in clinical experience. The study found that education and experience were not predictors of accuracy in interpreting test results (Faust et al., 1988 as quoted in Faust and Ziskin, 1988). High school students and clinicians, working from the same data, fared equally poorly in predicting violent behavior. When making predictions, they both accorded importance to the same data (Leirer, Werner, Rose, and Yesavage, 1984 as quoted in Faust and Ziskin, 1988). Werner, Rose, and Yesavage (1983) report similar low levels of accuracy in predicting imminent danger and find that clinical experience did not relate to increased accuracy.

Sub-phase 2: Weighing Decision Against Consequences

After a decision is made it is often weighed against the expected consequences. If the expected consequences are unacceptable to the clinician he or she may select another alternative which is then weighed. Judgment can be restrained by the perceived value of responses. Physicians, for example, favor a Type II error (finding a condition that does not exist) over Type I error (not finding a condition that exists). Jurists, on the other hand, prefer making a Type I error (Scheff, 1963). Scheff explains that expected consequences of error can account for these differences.

In mental health decision-making, as opposed to general medicine, the situation is complicated by beliefs that diagnoses can be destructive labels. Less severe catch-all diagnoses like adjustment disorder are used instead of more severe diagnoses (Kirk and Kutchins, 1988; Sharfstein, Towery, and Milowe, 1980). This spares the client the stigma attached to more severe diagnoses.

While clinicians seem conservative about making complex diagnoses (affixing a specific label), they seem liberal in making affirmative simple

diagnoses: in need of treatment. For example, in a study by Kaeser and Cooper (1971) 85% of the patients referred for psychiatric consultation by general practitioners were asked to return for further intervention. In that study the fear of labeling, and the tendency toward Type I errors, seems to pertain only to complex diagnoses.

When faced with uncertainty or the possibility of negative consequences to a decision, clinicians sometimes use very general judgments ("Aunt Fanny descriptions") [Tallent, 1958] or general flattering judgments ("Barnum Effect") [Meehl, 1956]. These are not clinically meaningful and are designed to be validated by the client. Both provide relief from uncertainty or conflict. Sometimes clinicians avoid making diagnoses. In Brown's (1987) study, supervisors told clinicians to diagnose all intake patients, except for pre-released prisoners. Prisoner diagnoses were to be "deferred" in order to protect the staff from the uses of, and challenges to, a formal diagnostic label in court. Financial remuneration is another consequence affecting diagnosis. Diagnoses that "don't pay" are sometimes replaced with others that do (Brown, 1987; Kirk and Kutchins, 1988; Kutchins and Kirk, 1988).

Sub-phase 3: Action

After reconsidering the judgment based on expected consequences, action is taken. The action is based on the previous stages in the decision-making process. Actions are the results of a decision, for example, recommending that a mother get custody, or that a person be hospitalized or be given a particular diagnosis.

Phase IV: Feedback

The schema in Figure 1 ends with feedback. Feedback should serve to correct performance at each phase. Practitioners do not always benefit from feedback. Feedback varies in different settings and it tends to be limited and misleading. In many diagnostic settings, clinicians receive no feedback about the correctness of their decisions. In treatment, feedback obtained from patients can be misleading. Patients may agree with "palm reader" type descriptions of their situation. This gives the clinician a false sense of accuracy (Faust, 1986).

Conclusion

This paper presents a framework in which to organize the literature on diagnosis and clinical decision-making bias. A four stage schema of clinical judgment is presented. The shortcomings of each stage are described. The

influence of background variables, including the effects of client biases, oversights, clinician mood, theoretical orientation, values and setting on diagnosis and clinical judgment are also presented. The paper points to the need to develop strategies to improve the accuracy of clinical judgment. Considerable knowledge and skill are required to make judgments about people; we would do better to consider that we may be wrong more often than we are right.

Practitioners appear to give little consideration to the decision-making process and subsequently decision-making is ignored in clinical textbooks (Turk and Salovey, 1985). To make more credible decisions about others, we need to develop means to overcome the shortcomings set forth in this paper. Improvement in accuracy of clinical judgments as a discipline is in its infancy (Arnoult and Anderson, 1987; Faust, 1986). The literature offers suggestions with little evidence of success in reducing judgmental errors. Fischhoff (1982) reviews debiasing techniques of two common biases, overconfidence and hindsight (i.e., tendency to exaggerate what one thought would happen in hindsight). He concludes that a few techniques can reduce hindsight bias, none eliminates it, and that overconfidence is even more resistant to change. Arkes (1981) and Dawes (1982) recommend teaching statistical principles to clinicians to improve their judgments. However their strategy has not been tested. It is clear that considerably more attention needs to be given to removing the many obstacles that may prevent clinicians from making accurate judgments.

References

American Psychiatric Association (1980). *Diagnostic and statistical manual of mental disorders* (third edition). Washington, DC: Author.

American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders* (third edition revised). Washington, DC: Author.

Anderson, C. A. (1983). Abstract and concrete data in the perseverance of social theories: When weak data lead to unshakable beliefs. *Journal of Experimental Social Psychology, 19*, 93–108.

Arkes, H.R. (1981). Impediments to accurate clinical judgment and possible ways to minimize their impact. *Journal of Consulting and Clinical Psychology, 49*, 323–330.

Arkes, H.R., and Hammond, K.R. (1986). General introduction. In H.R. Arkes and K.R. Hammond (Eds.), *Judgment and decision making: An interdisciplinary reader* (pp. 1–11). New York: Cambridge University Press.

Arnoult, L.H., and Anderson, C.A. (1987). Identifying and reducing causal reasoning biases in clinical practice. In D.C. Turk and P. Salovey (Eds.), *Reasoning, inference and judgment in clinical psychology* (pp. 107–123). New York: The Free Press.

Barnea, Z., Rahav, G., and Teichman, M. (1987). The reliability and consistency of self-reports on substance use in a longitudinal study. *British Journal of Addiction, 82*, 891–898.

Barocas, R., and Vance, F.L. (1974). Physical appearance and personal adjustment counseling, *Journal of Counseling Psychology, 21*, 96–100.

Bartsch, D.A., Shern, D.L., Feinberg, L.E., Fuller, B.B., and Willet, A.B. (1990). Screening CMHC outpatients for physical illness. *Hospital and Community Psychiatry, 41*, 786–790.

Baskin, D. (1984). Cross-cultural conceptions of mental illness. *Psychiatric Quarterly, 56*, 45–53.
Benson, G., and Holmberg, M.B. (1985). Validity of questionnaires in population studies on drug use. *Acta Psychiatrica Scandinavica, 71*, 9–18.
Bieri, J., Atkins, A.L., Briar, S., Leamen, R.L., Miller, H., and Tripodi,T. (1966). *Clinical and social judgment: The discrimination of behavioral information.* New York: Robert E. Krieger Publishing.
Boxer, P.A., and Garvey, J.T. (1985). Psychiatric diagnosis of cuban refugees in the United States: Findings of medical review boards. *American Journal of Psychiatry, 142*, 86–89.
Brink, P.J. (1972). Behavioral characteristics of heroin addicts on a short-term detoxification program. *Nursing Review, 21*, 38–45.
Brown, L.S. (1990). Taking account of gender in the clinical assessment interview. *Professional Psychology Research and Practice, 21*, 12–17.
Brown, P. (1987). Diagnostic conflict and contradiction in psychiatry. *Journal of Health and Social Behavior, 28* (March), 37–50.
Cash, T.F., Kehr, J., Polyson, J., and Freeman, V. (1977). Role of physical attractiveness in peer attribution of psychological disturbance. *Journal of Counseling and Clinical Psychology, 45*, 987–993.
Chapman, L.J., and Chapman, J.P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology, 72*, 193–204.
Chapman, L.J., and Chapman, J.P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74*, 271–280.
Craine, L.S., Henson, C.E., Colliver, J.A., and MacLean, D.G. (1988).Prevalence of a history of sexual abuse among female psychiatric patients in a state hospital system. *Hospital and Community Psychiatry, 39*, 300–304.
Dawes, R.M. (1982). The value of being explicit when making clinical decisions. In T.A. Wills (Ed.), *Basic process in helping relationships* (pp. 37–58). Orlando, Florida: Academic Press.
Dawes, R.M. (1986). Representative thinking in clinical judgment. *Clinical Psychology Review, 6*, 425–441.
Dickert, J. (1988). Examination of bias in mental health evaluations of deaf patients. *Social Work, 33*, 273–275.
Drake, R.E., and Vaillant, G.E. (1985). A validity study of Axis II of DSM-III. *American Journal of Psychiatry, 142*, 553–558.
Ellis, M.V., Robbins, E.S., Schult, D., Ladany, N., and Banker, J. (1990). Anchoring errors in clinical judgment: Type I error, adjustment or mitigation?. *Journal of Counseling Psychology, 37*, 343–351.
Elstein, A.S., Shulman, L.S., and Sparfka, S.A. (1978). *Medical problem solving: An analysis of clinical reasoning.* London and New York: Harvard University Press.
Faust, D. (1986). Research on human judgment and its application to clinical practice. *Professional Psychology Research and Practice, 17*, 420–429.
Faust, D., Guilmette, T.J., Hart, K.J., and Arkes, H.R. (1988). Neuropsychologists' training, experience and judgment accuracy. *Archives of Clinical Neuropsychology, 3*, 145–163.
Faust, D., and Ziskin, J. (1988). The expert witness in psychology and psychiatry. *Science, 241*, 31–35.
Fernando, T., Mellsop, G., Nelson, K., Peace, K., and Wilson, J. (1986). The reliability of axis V of DSM-III. *American Journal of Psychiatry, 143*, 752–755.
Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, and A.Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422–444). New York: Cambridge University Press.
Flaherty, E.G., and Flaherty, J.A. (1983). Medical students' performance in reporting alcohol related problems. *Journal of Studies on Alcoholism, 44*, 1083–1086.
Flaum, M., Arndt, S., and Andreasen, N.C. (1991). The reliability of "bizarre" delusions. *Comprehensive Psychiatry, 32*, 59–65.
Fuhrer, R. Rouillon, F., and Lellouch, J. (1986). Diagnostic reliability among French psychiatrists using DSM-III criteria. *Acta Psychiatrica Scandinavica, 73*, 12–16.
Gauron, E.F., and Dickinson, J.K. (1969). The influence of seeing the patient first on diagnostic decision-making in psychiatry. *American Journal of Psychiatry, 126*, 199–205.

Gelinas, D.J. (1983). The persisting negative effects of incest. *Psychiatry, 46,* 312–332.
Goldberg, L. (1959). The effectiveness of clinical judgments: The diagnosis of organic brain damage from Bender–Gestalt Test. *Journal of Consulting Psychology, 23,* 25–33.
Goldberg, L. (1968). Simple models or simple process? Some research on clinical judgments. *American Psychologist, 23,* 483–496.
Hammond, K.R., and Summers, D.A. (1965). Cognitive dependence on linear and nonlinear cues. *Psychological Review, 72,* 215–224.
Heumann, K.A., and Morey, L.C. (1990). Reliability of categorical and dimensional judgments of personality disorder. *American Journal of Psychiatry, 147,* 498–500.
Hiler, E.W., and Nesvig, D. (1965). An evaluation of criteria used by clinicians to infer pathology from figure drawings. *Journal of Consulting Psychology, 29,* 520–529.
Hjortso, S., Butler, L., Clemmesen, P.W., Jepsen, M., Kastrup, M.,Vilmar, T., and Bech, P. (1989). The use of case vignettes in studies of interrater reliability of psychiatric syndromes and diagnoses. *Acta Psychiatrica Scandinavica, 80,* 632–638.
Hobfoll, S.E., and Penner, L.A., (1978). Effect of physical attractiveness on therapists' initial judgments of a person's self-concept. *Journal of Consulting and Clinical Psychology, 46,* 200–201.
Hoffman, P.J., Slovic, P., and Rorer, L.G. (1968). An analysis-of-variance model for the assessment of configural cue utilization in clinical judgment. *Psychological Bulletin, 69,* 338–349.
Hogarth, R.M. (1980). *Judgment and choice: The psychology of decision making.* New York: John Wiley and Sons.
Ingram, R.E. (Ed.). (1986). *Information processing approaches to clinical psychology.* New York: Academic Press.
Jacobson, A., and Herald, C. (1990). The relevance of childhood sexual abuse to adult psychiatric inpatient care. *Hospital and Community Psychiatry, 41,* 154–158.
Johnston, L.D., and O'Malley, P.M. (1985). Issues in validity and population coverage in student surveys of drug use. In B. Rouse, N.J. Kizel, and L.G. Richards (Eds.), *Self-report methods of estimating drug use: Meeting current challenges to validity.* Washington, D.C.: NIDA.
Johnston, R., and McNeal, B.F. (1967). Statistical versus clinical prediction: Length of neuropsychiatric hospital stay. *Journal of Abnormal Psychology, 72,* 335–340.
Kaeser, A.C., and Cooper, B. (1971). The psychiatric outpatient, the general practitioner and the outpatient clinic: An operational study review. *Psychological Medicine, 1,* 312–325.
Kanwischer, R.W., and Hundley, J. (1990). Screening for substance abuse in hospitalizing psychiatric patients. *Hospital and Community Psychiatry, 41,* 795–797.
Kirk, S.A., and Kutchins, H. (1988). Deliberate misdiagnosis in mental health practice. *Social Service Review, 62,* 225–237.
Klayman, J., and Ha, Y.W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94,* 211–228.
Koranyi, E.K. (1979). Morbidity and rate of undiagnosed physical illnesses in a psychiatric population. *Archives of General Psychiatry, 36,* 295–300.
Kutchins, H., and Kirk, S.A. (1986). The reliability of DSM-III: A critical review. *Social Work Research and Abstracts, 22,* 3–12.
Kutchins, H., and Kirk, S.A. (1988). The business of misdiagnosis: DSM-III and clinical social work. *Social Work, 33,* 215–220.
Langer, E., and Abelson, R.P. (1974). A patient by any other name. *Journal of Consulting and Clinical Psychology, 42,* 4–9.
Leirer, Y.O., Werner, P.D., Rose, T.L., and Yesavage, J.A. (1984). *Predictions of violence by high-school students and clinicians.* Paper presented at the Annual Convention of the Oregon Psychological Association, Newport, Oregon.
Levy, B.L., and Ulman, E. (1967). Judging psychopathology from painting. *Journal of Abnormal Psychology, 72,* 182–187.
Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., and Combs, B. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 551–578.
Lieberman, P., and Baker, F. (1985). The reliability of psychiatric diagnosis in the emergency room. *Hospital and Community Psychiatry, 36,* 291–293.

L'opez, S.R. (1989). Patient variable biases in clinical judgment: Conceptual overview and methodological considerations. *Psychological Bulletin, 106*, 184–203.

Luft, J. (1950). Implicit hypotheses and clinical predictions. *Journal of Abnormal Psychology, 15*, 115–119.

MacKinnon, R.A., and Michels, R. (1971). *The psychiatric interview in clinical practice*. Philadelphia: W.B. Saunders.

Matarazzo, J.D. (1978). The interview: Its reliability and validity in psychiatric diagnosis. In B.B. Wolman (Ed.), *Clinical diagnosis of mental disorders: A handbook* (pp. 47–96). New York: Plenum.

Matarazzo, J.D. (1990). Psychological assessment versus psychological testing: Validation from Binet to the school, clinic and courtroom. *American Psychologist, 45*, 999–1017.

Meehl, P.E. (1956). Wanted a good cookbook. *American Psychologist, 11*, 263–272.

Mellsop, G., Varghese, F., Joshua, S., and Hicks, A. (1982). The reliability of Axis II of DSM-III. *American Journal of Psychiatry, 139*, 1360–1361.

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.

Newcorn, J.H., and Strain, J. (1992). Adjustment disorder in children and adolescents. *Journal of American Academy of Child and Adolescent Psychiatry, 31*, 318–326.

Nisbet, R., and Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, New Jersey: Prentice–Hall.

Nurius, P.S., and Gibson, J.W. (1990). Clinical observation, inference, reasoning, and judgment in social work: An update. *Social Work Research and Abstracts, 26*, 18–25.

Oskamp, S. (1962a). How clinicians make decisions from MMPI: An empirical study. *American Psychologist, 17* (Abstract), 316.

Oskamp, S. (1962b). The relationship of clinical experience and training methods to several criteria of clinical prediction. *Psychological Monographs, 76* (28, Whole No. 547).

Oskamp, S. (1967). Clinical judgment from the MMPI: Simple or complex? *Journal of Clinical Psychology, 23*, 411–415.

Pain, M.D., and Sharpley, C.F. (1989). Varying the order in which information is presented: Effects on counselors' judgments of clients' mental health. *Journal of Counseling Psychology, 36*, 3–7.

Perlin, M. (1989). *Mental disability law—Civil and criminal*. Charlotsville,Virginia: Michie.

Plag, J.A., and Arthur, R.J. (1965). Psychiatric re-examination of unsuitable naval recruits: A two year follow-up. *American Journal of Psychiatry, 122*, 534–541.

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

Richards, M.S., and Wierzbicki, M. (1990). Anchoring errors in clinical-like judgments. *Journal of Clinical Psychology, 46*, 358–365.

Rorer, L.G., Hoffman, H.D., Dickman, H.D., and Slovic, P. (1967). Configural judgments revealed (summary). *Proceedings of the 75th Annual Convention of the American Psychological Association, 2*, 195–196.

Rosenhan, D.L. (1975). The contextual nature of psychiatric diagnosis. *Journal of Abnormal Psychology, 84*, 462–474.

Ross, L., Lepper, M.R., and Hubbard, M. (1975). Perseverance in self-perception: Biased attribution process in the debriefing paradigm. *Journal of Personality and Social Psychology, 32*, 880–892.

Safer, D.J. (1987). Substance abuse by young adult chronic patients. *Hospital and Community Psychiatry, 38*, 511–514.

Salovey, P., and Turk, D.C. (1987). Some effects of mood on clinicians' memory. In D.C. Turk and P. Salovey (Eds.), *Reasoning, inference and judgment in clinical psychology* (pp. 107–123). New York: The Free Press.

Sandifer, M.G., Hordern, A., and Green, L.M. (1970). The psychiatric interview: The impact of the first three minutes. *American Journal of Psychiatry, 126*, 968–973.

Schaeffer, R. W. (1964). Clinical psychologists' ability to use the Draw-a-Person test as an indicator of personality adjustment. *Journal of Consulting Psychology, 28*, 383.

Scheff, T.J. (1963). Decision rules, types of error, and their consequences in medical diagnoses. *Behavioral Science, 8*, 97–107.

Schoefield, W. (1964). *Psychotherapy: The purchase of friendship*. Englewood Cliffs, New Jersey: Prentice–Hall.

Shapiro, M. (1989, November). *Computerized decision technology improves decision practice in social service: Decision support system in youth probation service*. Paper presented at the International Experts' Consultation on the Use of Computers in Social Work, Jerusalem, Israel.

Sharfstein, S., Towery, O., and Milowe, I. (1980). Accuracy of diagnostic information submitted to an insurance company. *American Journal of Psychiatry, 136*, 70–73.

Silverman, L.H. (1959). A Q-sort study of the validity of evaluations made from projective techniques. *Psychological Monographs, 73* (7, Whole No. 477).

Skodol, A.E. (1991). Axis IV: A reliable and valid measure of psychosocial stressors? *Comprehensive Psychiatry, 32*, 503–515.

Slovic, P., and Lichtenstein, S.C. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance, 6*, 649–744.

Smith, J.W. (1983). Diagnosing alcoholism. *Hospital and Community Psychiatry, 34*, 1017–1021.

Snyder, M., and Thomsen, C.J. (1987). Interactions between therapists and clients: Hypothesis testing and behavioral confirmation. In D.C. Turk and P. Salovey (Eds.), *Reasoning, inference and judgment in clinical psychology* (pp. 124–152). New York: The Free Press.

Stricker, G. (1967). Actuarial, naive clinical, and sophisticated clinical prediction of pathology from figure drawings. *Journal of Consulting and Clinical Psychology, 31*, 492–494.

Strohmer, D.C., Shivy, V.A., and Chiodo, A.L. (1990). Information processing strategies in counselor hypothesis testing: The role of selective memory and expectancy. *Journal of Counseling Psychology, 37*, 465–472.

Tallent, N. (1958). On individualizing the psychologists' clinical evaluation. *Journal of Clinical Psychology, 14*, 243–244.

Taylor, S.E., and Thompson, S.C. (1982). Stalking the elusive "vividness" effect. *Psychological Review, 89*, 155–181.

Temerlin, M.K. (1970). Diagnosis bias in community mental health. *Community Mental Health Journal, 6*, 110–117.

Temerlin, M.K., and Trousdale, W.W. (1969). The social psychology of clinical diagnosis. *Psychotherapy Theory Research Practice, 6*, 24–29.

Turk, D.C., and Salovey, P. (1985). Cognitive structures, cognitive process and cognitive behavior modification: II. Judgments and inferences of the clinician. *Cognitive Therapy Research, 9*, 19–33.

Turk, D.C., and Salovey, P. (Eds.). (1987). *Reasoning, inference and judgment in clinical psychology*. New York: The Free Press.

Tversky, A., and Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105–110.

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

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

Tversky, A., and Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 293–315.

Vitiello, B., Malone, R., Buschle, P.R., Delaney, M.A., and Behar, D. (1990). Reliability of DSM-III diagnoses of hospitalized children. *Hospital and Community Psychiatry, 41*, 63–67.

Walters, G.D., White, T.W., and Greene, R.L. (1988). Use of the MMPI to identify malingering and exaggeration of psychiatric symptomatology in male prison inmates. *Journal of Consulting and Clinical Psychology, 56*, 111–117.

Wanderer, Z.W. (1969). Validity of clinical judgments based on human figure drawings. *Journal of Consulting Psychology, 33*, 143–150.

Werner, P.D., Rose, T.L., and Yesavage, J.A. (1983). Reliability, accuracy, and decision-making strategy in clinical predictions of imminent dangerousness. *Journal of Consulting and Clinical Psychology, 51*, 815–825.

Westermeyer, J., and Zimmerman, R. (1981). Lao folk diagnosis for mental disorder: A comparison with psychiatric rating scales. *Medical Anthropology, 5*, 425–443.

Wiggins, N., and Hoffman, P.J. (1968). Three models of clinical judgment of abnormal psychology. *Journal of Abnormal Psychology, 73*, 70–77.

Wish, E.D., and O'Neil, J.A. (1989). *Drug use forecasting (OUF):Research update*. Washington, DC: US Department of Justice, NIJ.

Ziskin, J., and Faust, D. (1988). *Coping with psychiatric and psychological testimony* (Volumes 1–3). Venice, California: Law and Psychology Press.