The *Illusion* of Mental Health

Jonathan Shedler
Institute of Advanced Psychological Studies Adelphi University
Martin Mayman
Department of Psychology University of Michigan
Melvin Manis
Department of Psychology University of Michigan

ABSTRACT

It is argued that researchers' reliance on "objective" mental health scales and disregard for clinical judgment has led to many mistaken conclusions. Specifically, standard mental health scales appear unable to distinguish between genuine mental health and the facade or illusion of mental health created by psychological defenses. Evidence is presented indicating that (a) many people who look healthy on standard mental health scales are not psychologically health, and (b) illusory mental health (based on defensive denial of distress) has physiological costs and may be a risk factor for medical illness. Clinical judges could distinguish genuine from illusory mental health, whereas "objective" mental health scales could not. The findings call into question the conclusions of many previous studies that rest on standard mental health scales. They suggest new ways of understanding how psychological factors may influence health. Finally, they suggest that clinical methods (which researchers often malign) may have an important role to play in meaningful mental health research.

This article addresses two issues that are usually discussed in separate literatures. The first has to do with the assessment of mental health. We will argue that the most widely used and cited measures of mental health suffer from a serious limitation. The limitation is that they cannot distinguish between genuine mental health and the facade of mental health created by psychological defenses. The second issue has to

do with the relation between psychological factors and physical health. We will argue that psychological defenses have concrete physical costs and may be risk factors for medical illness.

Genuine and Illusory Mental Health

Countless scales exist to assess one or another facet of mental health. In general these scales are straightforward. Items tend to be transparent in intent, and investigators tend to accept scale scores at face value–assuming, for example, that high scores on depression scales signify depression, and low scores signify relative psychological health. [1] In contrast, psychoanalytic thinkers (and depth psychologists more generally) are often unwilling to accept self-report data at face value. They take seriously the notion of unconscious processes and unconscious defenses and asume that psychological distress is often covert, experienced and expressed only indirectly. From this perspective, many people who report psychological health may not be healthy at all.

To explore this possibility, we will investigate the following hypothesis: *Among people who "look good" on mental health scales, there are two subgroups. One subgroup is psychologically healthy. A second subgroup is made up of people who are psychologically distressed, who maintain an illusion of mental health through defensive denial of psychological distress.* [2]

People in the defensive group would be characterized by a need to see themselves as well adjusted, despite underlying vulnerability. Presumably, they preserve a belief in their "adjustment" by disavowing much of their emotional life, and so have little awareness of their needs, wishes, and feelings. We shall refer to these people as *defensive deniers* and, alternatively, as having *illusory mental health.*

To distinguish between genuine and illusory mental health, we rely on both traditional self-report measures and clinical judgment. Subjects complete standard mental health scales and, independently, are evaluated by an experienced clinician. Those reporting psychological health and judged healthy by the clinician are classified as *genuinely healthy.* Those reporting psychological health but judged distressed are classified as having *illusory mental health.* A third group is made up of *manifestly distressed* subjects, who report distress and are judged distressed. We have chosen these labels for the implications they convey; perhaps our findings will persuade the reader that these labels are, in fact, justified.

The clinical evaluations are based on subjects' accounts of their earliest memories (the Early Memory Test; see Mayman, 1963 , 1968 ; Mayman & Faris, 1960). We treat early memories like projective tests. That is, we do not assume that the memories are accurate accounts of past events. Rather, we see them as a source of information about how a person construes, organizes, and presents his experiences–that is, as a source of information about the "lenses" through which a person sees himself and his world.

Our reliance on qualitative clinical methods follows from the assumptions that human communication can convey multiple levels of meaning beyond its face value or manifest content, that these meanings emerge most readily when communication is not overly structured (cf. J. Weinberger & McClelland, 1990), and that another human being is the best (as of now, perhaps the only) "instrument" for registering these meanings. Given these assumptions, a person may report psychological health, while the covert meanings carried by his or her communications convey emotional distress.

Definition of Mental Health Scales

We use the term *mental health scale* to refer to a wide range of scales intended to assess one or another facet of mental health—distress. This includes popular depression, anxiety, and neuroticism scales as well as measures assessing self-esteem, optimism, self-efficacy, and related constructs. Such scales are often discussed in separate literatures as if they measured separate things. In fact they correlate highly, and they assess a common mental health—distress factor (see, e.g., Dohrenwend, Shrout, Egri, & Mendelsohn, 1980 ; Gotlib, 1984 ; Smith, Pope, Rhodewalt, & Poulton, 1989 ; Tanaka-Matsumi & Kameoka, 1986 ; Watson & Clark, 1984 ; Watson & Pennebaker, 1989).

The general mental health—distress factor has been recognized and re-recognized over the years. J. Block (1965) identified it as the first major factor in self-report inventories, referring to it as Ego-Resilience. In the "five-factor" personality taxonomy (e.g., McCrae & Costa, 1987 ; see Goldberg, 1993 , and John, 1990 , for excellent overviews) it is the Neuroticism or Emotional Stability factor. Most recently, Watson and Clark (1984) have labeled the factor Negative Affectivity. When we speak of "mental health" scales, we refer to this general factor and to all the various and sundry scales that assess it (some popular examples include the Beck Depression Inventory [BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961], Spielberger State-Trait Anxiety Inventory [Spielberger, Gorsuch, & Lushene, 1970], Rosenberg Self-Esteem Scale [Rosenberg, 1965], Taylor Manifest Anxiety Scale [J. A. Taylor, 1953], Eysenck Neuroticism scale [Eysenck & Eysenck, 1975], Life Orientation Test [Scheier & Carver, 1985], and all scales assessing the first factor of the Minnesota Multiphasic Personality Index [MMPI]; many other popular scales could also be cited). [3]

Reconciling Our Critique With Previous Findings

A huge literature documents the validity of mental health scales. These scales predict a wide range of relevant outcomes (see, e.g., J. Block, 1965 ; Watson & Clark, 1984). It appears, then, that our criticism of mental health scales contradicts the available evidence. This apparent contradiction can be explained.

First, we do not propose that mental health scales are simply invalid. Rather, we suggest that scores are valid when they fall near the "distressed" end of the healthy—distressed continuum and ambiguous when they fall near the "healthy" end. Statistically, this situation will attenuate correlations between mental health scales and relevant outcome measures, but it will not eliminate them (cf. Fisher, 1959 , on the "twisted pear"). Thus, *mental health scales may be flawed as we have described, but validation studies will still show positive results.*

Second, psychological defense does not manifest itself only in responses to self-report scale items, but it pervades every aspect of life. For this reason, the defensive processes that enable people to look healthy on mental health scales may also enable them to look healthy on the criterion measures used to validate these scales. Consider, for example, validation studies showing that mental health scores correlate with observer ratings (e.g., McCrae, 1982). Why wouldn't they correlate? The defensive processes that hide psychological pain even from the self may well hide it from the average observer. Thus, mental health scores and observer ratings may reflect (i.e., be confounded by) the very same defensive processes. The same argument can be made for many other criterion variables used in validation studies. (In rare

instances in which outcome variables have not been subject to psychological defense, intriguing findings have emerged. For example, Rose, 1956 , reported that the soldiers who looked healthies on mental health scales were the most likely to break down in combat. Derogatis, Abeloff, & Melisaratos, 1979 , reported that cancer patients who reported high levels of mental health sooner than patients who acknowledged greater distress.)

Psychological Defense and Physiology

Because the defensive processes that allow defensive deniers to look healthy on self-report scales may also allow them to look healthy in other domains, finding suitable criterion measures (on which defensive deniers and genuinely healthy people will differ) is problematic. Clinical wisdom holds that defensive denial has many costs, restricting one's capacity for love, work, and play. But operationalizing this kind of cost would be a formidable task, possibly requiring long-term, in-depth, interpretive study of individual lives. There is a need for criterion measures that are readily at hand. We propose that physiological measures may serve this purpose. Thus, our second hypothesis: *Psychological defense has physiological costs. It is associated with autonomic reactivity and may be a risk factor for medical illness.*

Support for this hypothesis comes from both the laboratory and the clinic. Research by Pennebaker and his associates indicates that the process of inhibiting thoughts and feelings entails physiological work, reflected in the short run in autonomic reactivity and in the long run in increased health problems (see Pennebaker, in press , and Pennebaker & Susman, 1988 , for overviews of this research). They argued that the physiological work is a cumulative stressor, increasing susceptibility to a variety of illnesses over time. The evidence is not only correlational; findings indicate that confronting painful thoughts and feelings can decrease autonomic arousal (Pennebaker & Beall, 1986), produce measurable changes in immune functioning, and reduce health care visits (Pennebaker, Kiecolt-Glaser, & Glaser, 1988). Research by other investigators also points in this direction (e.g., Hare, 1966 , 1973 ; Jensen, 1987 ; Kneier & Temoshok, 1984 ; Lacey, 1959 ; Lazarus & Alfert, 1964 ; Parsons, Fulgenzi, & Edelberg, 1969 ; Scarpetti, 1973 ; Temoshok, 1987 ; D. Weinberger, Schwartz, & Davidson, 1979).

Clinical observation and clinically derived theory complement these empirical findings. McDougall (1989) , writing from a psychoanalytic perspective, has presented extensive case history evidence linking psychological defense with physical illness. McDougall noted that somatic disturbance is the infant's earliest means of expressing distress and argued that with development comes the capacity to express distress in increasingly symbolic forms–that is, to express distress in the psychic rather than the somatic domain. When distress cannot find expression psychically, McDougall maintained, then it may find expression through more primitive somatic channels, sometimes with life-threatening consequences.

Clinical evidence of a different kind comes from a study in which Type A heart attack survivors received psychotherapy. The investigators (M. Friedman & Ulmer, 1984) noted that these patients "harbored insecurities and in most cases insufficient self-esteem," but this was "not immediately apparent either to the [therapists] or the participants themselves" (p. 167). They concluded that defensive efforts to compensate for insecurity are at the core of the coronary-prone personality. Moreover, the rate of second heart attacks was reduced by nearly 50% among patients who received psychotherapy, relative to a control group of patients who received normal medical care.

In the case of heart disease, the relation between psychological defense and illness may be mediated by autonomic reactivity. Flux in heart rate (HR) and blood pressure appears to create turbulence and sheer stress in the coronary arteries, damaging the inner lining of the arteries (the endothelium) and setting in motion a process leading to atherosclerosis and, ultimately, to heart disease proper (Krantz & Manuck, 1984 ; Manuck, Kaplan, & Clarkson, 1983).

In light of these empirical and theoretical recognitions, we use measures of heart rate and blood pressure reactivity as criterion variables to evaluate the success of our procedure for distinguishing genuine from illusory mental health. To the extent that a link between psychological defense and physiology is conceptually reasonable, positive findings provide support for both of the hypotheses we have proposed (that people who report psychological health can be divided into genuinely healthy and defensive subgroups, and that psychological defense has physiological consequences).

Verbal Evidence of Psychological Defense

To converge or "triangulate" on the concept of psychological defense, Studies 1 and 2 also include a second criterion variable, based on subjects' verbal responses to a phrase association test (after Mandler, Mandler, Kremen, & Sholiton, 1961 , and D. Weinberger et al., 1979). This test requires subjects to verbalize their first associations to a variety of stimulus phrases. The phrases have themes some subjects may perceive as threatening, having to do with sexuality, aggression, and dependency. We coded the subjects' verbal responses for indications of efforts, conscious or not, to avoid dealing with the content of the stimulus phrases. If our procedure for assessing defensive denial predicts verbal manifestations of defense as well as physiological reactivity, this is further evidence for its validity.

Overview of Studies 1, 2, and 3

Studies 1, 2, and 3 explore the linked hypotheses that (a) people characterized by illusory mental health can be distinguished from genuinely healthy people, and (b) illusory mental health has physiological costs. The research design is similar in all three studies. Each study has two phases:

- Subjects complete a standard self-report measure of mental health (the Eysenck Neuroticism scale in Studies 1 and 2 and the BDI in Study 3). They also are evaluated by clinical judges, who base their clinical inferences on the Early Memory Test. When self-report and clinical judgment converge in indicating health, subjects are classified as genuinely healthy; when self-report indicates health but clinical judgment indicates distress, they are classified as having illusory mental health; when both data sources indicate distress, subjects are classified as manifestly distressed.

- The subjects participate in a laboratory session in which they are exposed to psychological stressors, and changes in heart rate and blood pressure are monitored. The genuine mental health, illusory mental health, and manifestly distressed groups are then compared with respect to coronary reactivity. A measure of verbal defensiveness provides a second criterion variable in Studies 1 and 2.

Study 1

Method Subjects

Subjects were 58 members of the University of Michigan community (students or staff), 25 male and 33 female. They ranged in age from 17 to 47, with a mean age of 21.5 years. People with a known history of cardiovascular disease were excluded from participation.

Psychological Measures

The subjects completed a battery of self-report scales that included the Eysenck Neuroticism scale as well as measures relevant to other investigations. The Neuroticism scale assesses the general psychological health—distress factor. The subjects also completed a written version of the Early Memory Test (EMT), which provided a basis for subsequent clinical inferences. The EMT instructs subjects to relax, allow their thoughts to go back to early childhood, and recall their earliest memory. It then asks for a written account of that memory. Open-ended follow-up questions ask subjects for their impressions of themselves in the memory, their impressions of other people, and the mood or feeling tone associated with the memory. Following this format, the test goes on to ask for the earliest memory of mother, earliest memory of father, earliest happy memory, earliest unhappy memory, a memory "in which you feel most fully yourself," and "the most active early memory you can think of."

Physiological Measures

During the laboratory session, heart rate (HR), systolic blood pressure (SBP), and diastolic blood pressure (DBP) were recorded automatically by an Air-Shields Model BP203NA automatic blood pressure monitor. The physiological readings were not visible to the experimenter during the session.

Materials

The subjects engaged in three laboratory tasks intended to elicit stress: (a) solving mental arithmetic problems, (b) telling stories in response to Thematic Apperception Test (TAT) cards, and (c) responding to a phrase association test. The mental arithmetic task made use of seven problems form the Wechsler Adult Intelligence Scale. The TAT task made use of TAT Cards, 1, 7GF, 13MF, 18GF, and 13B from the standard Murray TAT set.

The phrase association test required subjects to say the first thing that came to mind in connection with various stimulus phrases. Thirty-five phrases from Mandler et al. (1961) served as stimuli. Fourteen phrases were neutral in content (e.g., "the horses worked well together," "the library purchased more books"), 7 had aggressive themes (e.g., "father convicted for torturing son," "student attacked by gang"), 7 had dependency themes (e.g., "father neglects his sick child," "mother bear deserts baby cubs"), and 7 had sexual themes (e.g., "prostitutes do anything men desire," "after the operation he was impotent").

Procedure

The session began with an eight-minute baseline period, during which subjects sat quietly and relaxed. HR, SBP, and DBP were recorded at 60-second intervals during the baseline period. Two additional physiological readings were taken after the baseline period, while subjects read aloud an affectively neutral passage from an undergraduate textbook. After this, the subjects engaged in several stressful tasks:

Mental arithmetic.

This task was introduced with instructions intended to elicit performance anxiety:

> The first test is a test of mental ability, an IQ test. It is important to try to do well, because we are going to compare your performance with the performance of others like you. I'll ask you some questions involving mental arithmetic, and you give the answers. To do well, you must give the correct answer as quickly as you can. I'll time you with a stopwatch.

The first two arithmetic problems served as warmup items. From the third problem on, the blood pressure monitor was triggered to record as each problem was presented. A two-minute rest period followed the mental arithmetic problems, and physiological recordings were taken at the end of this rest period.

TAT stories.

The TAT was introduced as follows:

> I am going to show you a series of pictures, and I want you to make up a story about each one. Tell me what is happening in the picture, what led up to it, and what the outcome will be. Also describe what the characters are thinking and feeling.

The first TAT card served as a warm-up. For the second through fifth cards, the blood pressure monitor was triggered to record as each card was presented. A two-minute rest period followed the last TAT card, and physiological recordings were taken at the end of this rest period.

Phrase association task.

The phrase association task was introduced as follows:

> For this next part I am going to read you some phrases. After each phrase, I want you to say the first thing that comes to mind, as quickly as possible. Give me a complete sentence or idea, not just a word. Anything you say as a response is fine, there are no right or wrong answers, but be sure to say the first thing that comes to mind.

The stimulus phrases were presented in blocks of seven phrases each, and the blood pressure monitor was triggered to record at the start of each block. The first and last block of phrases were neutral in content; the second, third, and fourth blocks contained phrases with aggressive themes, dependency themes, and sexual themes, respectively. Verbal responses were tape-recorded for subsequent analysis.

In all cases, intervals of 60 seconds or more separated consecutive physiological recordings. The sequence of tasks was the same for all subjects (i.e., the order of the tasks was not counterbalanced). The experimenter was blind to all other data.

Independent Variables Neuroticism scores.

Local norms for the Eysenck Neuroticism scale were available from a large University of Michigan sample ($n = 287$) drawn from the same population as the research subjects. For convenience, the Neuroticism scores were converted to T scores ($M = 50$, $SD = 10$), on the basis of norms established in the larger Michigan sample.

Clinical evaluations.

An experienced clinician (the second author) evaluated each subject with respect to mental health—distress, using the Early Memory Test as a basis for clinical inferences. The clinician was not concerned with whether the reported memories were "accurate," nor did he attempt to assess actual

childhood events. Rather, he assumed that the selection, organization, and style of recounting the memories revealed something about the subject's present-day psychological make-up (Mayman, 1968). In interpreting the EMT, the clinician attended to qualitative factors such as how the self was represented, how the interpersonal world was represented, the affective tone of the material, whether the memories were narratively coherent or contained inner contradictions (suggesting omissions and distortions), and so on. The clinician was blind to all other data. Prior researchers using the Early Memory Test have reported interrater correlations of .76, .94 (Jaffe, 1985), .74, and .88 (Diamond, 1983) for essentially similar judgments; researchers who have treated the EMT scores as dichotomous have reported interrater agreements of 80% (Bronson, 1982) and 72% (Greenwald, 1977). It should be noted that these levels of agreement were reached by skilled, psychoanalytically oriented clinicians after considerable training and practice. The question of interrater reliability will be taken up at greater length in Study 2.

The clinician studied the Early Memory Test responses and made dichotomous judgments, classifying subjects as relatively healthy or relatively distressed. Twenty-nine subjects were judged distressed, 12 were judged healthy, and 17 were left unclassified because the clinical "data" were inconclusive. In general, subjects were left unclassified because their written responses to the Early Memory Test were too sparse for analysis.

Dependent Measures Physiological reactivity.

Viewing the heart as a pump, the work it performs is a function of both the rate and the force of its strokes. The rate—pressure product (RPP), defined as (HR * SBP)/100, captures both of these elements. Rate is reflected by heart rate, and force is reflected by systolic blood pressure. The rate—pressure product is a commonly used measure in cardiovascular research and has important medical concomitants (e.g., Kitamura, Jorgensen, Gobel, & Wang, 1972 ; Robinson, 1967).

Baseline measures of heart rate, systolic blood pressure, diastolic blood pressure, and the rate—pressure product were obtained by averaging the last two observations recorded during the eight-minute baseline period. These baselines were then subtracted from subsequent observations to create physiological change scores (reactivity scores), indicating increases in autonomic activity over and above baseline in response to stress.

Verbal defensiveness.

Audiotape recordings of responses to the phrase association stimuli were transcribed, then scored for manifestations of psychological defense using a coding scheme adapted from Mandler et al. (1961) . Defense in this context refers to efforts, conscious or not, to avoid the content of the stimulus phrases. Thus, *verbal defensiveness* was coded if a subject commented on the wording or phrasing of a stimulus phrase by way of avoiding its content, misinterpreted the stimulus phrase, said he could not think of a response, began a response but failed to complete the thought, asked the experimenter to repeat the phrase, had a response latency greater than eight seconds, took back a response (e.g., said "wait, that's not what I meant"), and so on. The coding was performed by a research assistant blind to all other data.

For each phrase association response, up to 15 separate manifestations of psychological defense could be scored. The verbal defensiveness score was simply the total number of manifestations of defense across the 21 stimulus phrases containing threatening (i.e., sexual, aggressive, or dependency) content. Prior researchers have reported interrater correlations of .75, .77 (Mandler et al., 1961), and .88 (D.

) for essentially similar measures.

Results Relation Between Self-Report and Clinical Evaluation

We have suggested that scores on mental health scales are valid when they indicate distress but ambiguous when they indicate health. If this is true (and if our clinical evaluations are valid), then the relation between the self-report Neuroticism scores and the clinical evaluations should be clearly asymmetric. People who report distress should be judged distressed by the clinician, but people who report health might or might not be judged healthy.

This asymmetric pattern was in fact observed. Figure 1 shows the relation between clinical evaluations (dichotomous) and Neuroticism scores (T scores). (For ease of interpretation, the graph is divided into quadrants.) Subjects who reported distress (scoring above the mean on Neuroticism) were generally classified by the clinician as distressed (in 11 of 14 cases). In contrast, subjects who reported health (scoring below the mean on Neuroticism) were sometimes classified as distressed and sometimes classified as healthy.

Classifying Defensive and Genuinely Healthy Subjects

Nine subjects scoring below the mean on Neuroticism and judged healthy by the clinician were classified as genuinely healthy (lower right quadrant in Figure 1). Eighteen subjects scoring below the mean on Neuroticism and judged distressed by the clinician were classified as having illusory mental health (lower left quadrant). Eleven subjects scoring above the mean on Neuroticism and judged distressed by the clinician were classified as manifestly distressed (upper left quadrant).

Physiological Measures

Eighteen physiological readings were available for each subject. Two were associated with the reading task, 5 with mental arithmetic problems, 4 with TAT cards, and 5 with the phrase association test. Two additional readings were taken during rest periods separating the tasks. The 18 measurement points and their sequence are indicated on the horizontal axis in Figure 2 .

Figure 2 shows the mean change in RPP (from resting baseline) for subjects classified as having illusory mental heath and subjects classified as genuinely healthy, at each measurement point. Subjects with illusory mental health show greater coronary reactivity at *every* observation point. The magnitude of the difference is often large enough to be considered medically significant. (For readability, data points for manifestly distressed subjects are not plotted; if plotted, they would fall between the other two groups. See below.)

Statistical comparisons.

To formally investigate differences in physiological reactivity between subjects classified as having illusory mental health and those classified as genuinely healthy, RPP change scores were aggregated across the 16 observations associated with stressors (e.g., the arithmetic, TAT, and phrase association tasks). On this aggregate reactivity index, the illusory mental health group showed significantly greater coronary reactivity ($M = 19.96$) than the genuinely healthy group ($M = 10.38$), $t(25) = 2.76$, $p = .01$, two-tailed. Note that these scores differ in magnitude by a factor of nearly two. Interestingly, subjects with illusory mental health also showed significantly greater coronary reactivity than manifestly distressed subjects ($M = 12.11$), $t(27) = 2.29$, $p = .03$, two-tailed.

An alternative data analysis strategy, and one that may have greater medical relevance, is to compare

subjects in terms of maximal (rather than mean) physiological reactivity (cf. Lacey & Lacey, 1958). Thus we selected, for each subject, the single observation at which RPP change was greatest. On this index of maximum reactivity, subjects with illusory mental health again showed greater physiological reactivity ($M = 41.74$) than genuinely healthy subjects ($M = 26.30$), t (25) = 3.25, p = .003, two-tailed, and greater physiological reactivity than manifestly distressed subjects ($M = 27.87$), t (27) = 3.06, p = .005, two-tailed. [4]

Separate analyses revealed that both HR and SBP contributed to these effects, with group differences reaching or approaching significance for all statistical tests. Comparisons involving DBP did not attain statistical significance. Prior research indicates that significant associations between psychosocial variables and DBP are rare (Contrada & Krantz, 1988 ; Mathews & Haynes, 1986).

Verbal Defensiveness Measure

Subjects with illusory mental health were compared with genuinely healthy and manifestly distressed subjects on the index of verbal defensiveness. As expected, those with illusory mental health showed more verbal manifestations of defense ($M = 10.53$) than either genuinely healthy subjects ($M = 6.78$), t (24) = 2.24, p < .05, two-tailed, or manifestly distressed subjects ($M = 7.18$), t (26) = 2.08, p = .05, two-tailed.

A Stronger Test of the Illusory Mental Health Hypothesis

In principle, the more a person defends against awareness of distress, the greater the physiological consequences. This implies that for subjects judged distressed by the clinician, lower Neuroticism scores should be associated with *higher* levels of coronary reactivity (because lower Neuroticism scores imply more denial of distress). For other subjects this relation should not be observed.

To test this, we correlated the Neuroticism scores with the RPP maximum change index, separately for (a) subjects judged distressed by the clinician and (b) a comparison group of subjects judged healthy or left unclassified by the clinician. [5] The expected relations were observed. In the distressed group, the correlation between Neuroticism and physiological reactivity was r (27) = − .48, p < .005. In the comparison group the correlation was r (25) = +.04, *ns* . [6] These two correlation coefficients differ significantly from one another, z = 1.99, p < .05, two-tailed. In short, clinical judgment acts as a moderator variable that significantly changes the relation between Neuroticism scores and physiological reactivity.

Figure 3 illustrates the important role of clinical judgment in this study. The figure shows the relation (the least square regression line) between Neuroticism scores and physiological reactivity for subjects judged distressed by the clinician, and the very different relation for subjects not judged distressed. *The findings suggest that mental health scales may be measuring different things in different people: For some, they may be assessing mental health, but for a sizable group of others, they may instead be assessing degree of psychological defense.*

Study 2

Unlike self-report scales, the value of the Early Memory Test depends on the skill and sensitivity of a human judge. Study 1 demonstrates that one human judge, an experienced clinician, can make meaningful psychological judgments using the Early Memory Test as a basis for clinical inference. We

must acknowledge that we do not know what percentage of clinicians could reproduce these judgments, or whether nonclinicians could do so.

To further explore the properties of the human judge as an assessment "instrument," Study 2 reanalyzes the Study 1 data. This time, the clinician is replaced by panels of undergraduate students, who interpret the Early Memory Test and provide evaluations of mental health. The student judges had no prior experience with the EMT, received no practice in its use, and were given only minimal instructions. Thus, the study is an unusually harsh test of both the robustness of subjective human judgment and the utility of the EMT as an assessment tool.

Method Student Judges

Two panels of student judges provided "clinical" evaluations, and we performed statistical analyses separately for each panel. Thus, two replications of Study 1 were conducted. Panel 1 was made up of 37 University of Oregon students enrolled in an undergraduate personality course in the spring of 1989. Panel 2 was made up of 33 students enrolled in the same course in the fall of 1989.

Procedure

The student judges were informed of the procedures and results of Study 1 and were invited to participate in a contest to "beat" the clinician at predicting the outcome measures. They received the following instructions regarding interpretation of the Early Memory Test:

> The clinician [of Study 1] treated the memories as a projective test (like the Rorschach or TAT). He assumed that the memories do not simply represent factual accounts of real events. Rather, they are seen through the lenses of the subject's present psychological make-up. These "lenses" may influence the selection, content, and telling of the memories.

> The central issue in judging between psychological health or distress is how the person sees himself or herself in relation to the world, and whether the relations with the world are associated with good or bad feelings:

> Is the world seen as somehow threatening, dangerous, malevolent, or frustrating? Is it associated with injury, disaster, traumatic punishment, or frustration? Does the person represent himself as at the mercy of external forces? These kinds of representations of self in relation to the world may indicate distress.

> Alternatively, is the world seen as comfortable, safe, secure, benign, and gratifying? Do others, especially parents, come across as sources of gratification, comfort, or security? These kinds of representations of self in relation to the world may indicate psychological health.

> Another important aspect is narrative believability. Sometimes when a subject says he or she is happy in a memory, you will be convinced of this happiness. But sometimes when a subject says that he or she is happy, the actual details will not convey the feeling of happiness, or they may even seem to contradict the subject's explicit statement. Similarly, a person may describe a parent as warm and comforting, but the details needed to convey an impression of warmth and comfort are not there. Trust your subjective impressions rather than the subject's explicit statements.

The student judges rated each subject on a 7-point scale with 1 indicating a judgment of *extremely unhealthy,* 7 a judgment of *extremely healthy,* and 4 indicating *uncertain.*

Results

To enhance the reliability of the mental health ratings, we aggregated the student-judge ratings and created one composite mental health rating for the Panel 1 judges and one composite mental health rating for the Panel 2 judges. Composite or aggregate ratings are generally superior to the ratings of individual judges because the idiosyncrasies of individual judges tend to cancel one another out, and the composite rating comes to reflect the core consensual wisdom of the group (the principle of aggregation; see, e.g., Horowitz, Inouye, & Siegelman, 1979 ; Rushton, Brainerd, & Preisley, 1983). Alpha reliabilities for the composite mental health ratings were .92 and .94 for Panels 1 and 2, respectively.

The wisdom of aggregation was born out when we examined the agreement between the student judges and the clinician from Study 1. The correlations between the clinician and individual student judges ranged widely, with an average correlation of .25 for the Panel 1 judges and .37 for the Panel 2 judges. In contrast, the correlations between the clinician and the composite ratings were .50 and .62 for Panel 1 and Panel 2, respectively. In nearly every case, the correlation between the clinician and the composite equaled or exceeded the correlations between the clinician and individual student judges.

Classification of Subjects

To conduct analyses parallel to those reported in Study 1, the composite mental health ratings were dichotomized and subjects were classified as healthy or distressed. For the Panel 1 composite ratings, cut-points were selected to reproduce the distribution obtained by the clinician in Study 1 (i.e., 12 subjects were classified as healthy, 29 as distressed, and 17 were left unclassified). [7] The same cut-point values were then applied to the Panel 2 composite ratings without regard to the resulting distribution. This procedure permitted cross validation of the findings.

On the basis of the "clinical" evaluations and the self-report Neuroticism scores, subjects were classified into genuine mental health, illusory mental health, and manifestly distressed groups as in Study 1. The classification was made twice, once using the "clinical" evaluations provided by the Panel 1 judges, and once using the "clinical" evaluations provided by the Panel 2 judges.

Statistical Comparisons Panel 1.

The dependent variable considered here is the RPP maximum change index reported in Study 1; hypothesis tests are one-tailed. Using the Panel 1 "clinical" evaluations as a basis for subject classification, subjects with illusory mental health showed more physiological reactivity ($M = 37.4$) than genuinely healthy subjects, ($M = 28.8$), $t(24) = 1.77$, $p < .05$, and more physiological reactivity than manifestly distressed subjects ($M = 29.6$), $t(27) = 1.61$, $p = .06$.

Subjects with illusory mental health also scored higher on the verbal defensiveness measure ($M = 8.3$) than genuinely healthy subjects ($M = 5.25$), $t(23) = 2.13$, $p < .05$. The difference between illusory mental health subjects and manifestly distressed subjects ($M = 7.3$) was in the expected direction but not statistically significant.

Panel 2.

Using the Panel 2 "clinical" evaluations as a basis for subject classification, subjects with illusory mental health again showed more physiological reactivity ($M = 40.1$) than genuinely healthy subjects ($M =$

27.1), $t(22) = 2.53$, $p = .01$, and more physiological reactivity than did manifestly distressed subjects ($M = 31.9$), $t(27) = 1.75$, $p = .04$.

Also, they had higher scores on verbal defensiveness ($M = 9.1$) than either genuinely healthy subjects ($M = 6.1$), $t(20) = 1.71$, $p = .05$, or manifestly distressed subjects ($M = 7.0$), $t(24) = 1.27$, $p = .10$.

A Stronger Test of the Illusory Mental Health Hypothesis

In Study 1 we saw that among subjects judged distressed by the clinician, lower Neuroticism scores predicted *higher* levels of physiological reactivity (presumably because lower Neroticism scores indicated more defensive denial). Among the remaining subjects, Neuroticism and physiological reactivity were not significantly related. This pattern was also replicated. Among subjects classified as distressed by the Panel 1 student judges, Neuroticism scores correlated negatively with physiological reactivity, $r(27) = -.30$, $p = .06$. For the comparison subjects, the correlation was $r(27) = +.18$, *ns*. The difference between these correlations is statistically significant, $z = 1.76$, $p = .04$.

Among subjects classified as distressed by the Panel 2 student judges. Eysenck Neuroticism scores correlated negatively with physiological reactivity, $r(26) = -.30$, $p = .06$. For the remaining subjects, the correlation was $r(28) = +.13$, *ns.* Again, the difference between these correlations is statistically significant, $z = 1.59$, $p = .05$.

These results reproduce the Study 1 findings. Once again, "clinical" judgment acts as a moderator variable that changes–indeed reverses–the relation between self-report and physiological reactivity. The results suggest that the outcome of Study 1 was not a coincidence and that the findings do not rest solely on the idiosyncratic skills of one unique clinician. Used appropriately, subjective human judgment can be robust.

Study 3

Because Studies 1 and 2 are based on the same small subject sample, it seemed desirable to replicate the findings linking illusory mental health and autonomic reactivity. Study 3 attempts to do this in a larger subject sample, using somewhat different methods. The Study 3 data were collected as part of an ongoing longitudinal study of ego and cognitive development initiated by Jack and Jean H. Block (see J. H. Block & Block, 1980 , for an extended description of the study).

Method Subjects

Data were available for 74 subjects participating in the Age 23 assessment of the Block and Block longitudinal study. The subjects were first recruited into the study at age 3, while attending nursery school in the San Francisco Bay area, and they were assessed on wide-ranging psychological measures at periodic intervals.

Psychological Measures

As part of the psychological assessment conducted at age 23, subjects completed the Beck Depression Inventory (Beck et al., 1961), a marker of the general psychological health—distress factor.

The subjects also participated in a videotaped interview lasting approximately 1 to 1/2 hours, organized around the theme of personal memories. This interview was conducted by an experienced, psychoanalytically trained clinician who was unaware of all other data. The clinician asked subjects to

recount 10 memories that seemed personally meaningful; she then asked for specific memories following the Early Memory Test format.

On the basis of this interview, the clinician provided a psychological description of each subject using the California Adult Q-sort (CAQ; [Block, 1978](#)). The CAQ is a personality assessment tool that allows clinicians to express their observations and formulations in a quantifiable form by assigning scores to 100 standard personality-descriptive statements. A score of 9 indicates that a statement is highly descriptive of a given person, 1 that it is highly undescriptive.

In this study, classification of subjects as healthy or distressed was based on the score on the single Q-sort item, "Has a brittle ego-defense system; has a small reserve of integration; would be disorganized or maladaptive when under stress or trauma." This Q-sort item was chosen a priori. Subjects with scores of 6 or higher were deemed relatively distressed, and those with scores of 4 or lower were deemed relatively healthy. Using this procedure, 35 subjects were classified as relatively distressed and 26 as relatively healthy.

Physiological Measures

The subjects also participated in a laboratory session during which they were exposed to a variety of stressors. Systolic and diastolic blood pressure were recorded during the laboratory session using an Ohmeda Model 2350 Finapress blood pressure monitor. This blood pressure monitor makes use of a pressurized finger cuff and is capable of providing near-continuous blood-pressure readings, updated with each heart beat. Blood pressure data from the Ohmeda monitor were fed to a J&J Enterprises Model I-330-5648 physiological monitoring system, which recorded the data on magnetic disk. The J&J physiological monitoring system also provided heart rate data via a plethysmograph sensor positioned on the thumb of each subject's nondominant hand. HR, SBP, and DBP were sampled at three-second intervals throughout the laboratory session.

Materials

The laboratory session involved several stressful tasks similar to those described in Study 1. These included a mental arithmetic task, the TAT, and a sentence association test designed especially for this study. The TAT task made use of TAT Cards 1, 7GF, 13MF, 18GF and 13B from the standard Murray TAT set. The sentence association test required subjects to read aloud 15 stimulus sentences and give their first association to these sentences. The stimulus sentences were phrased in the first person and made statements designed to be psychologically threatening (e.g., "My mother never loved me," "I don't have what it takes to succeed in life," "I feel ugly," "There is something wrong with me sexually," "Inside I know I am an unhappy person").

Laboratory Procedure

The laboratory session began with a 10-minute baseline period, during which subjects were asked to sit quietly and relax.

Counting backward.

The first stressor was a counting task, requiring subjects to count backward by 13s from 609. The task was introduced as a test of mental ability. Subjects were told that speed and accuracy both count, and they were asked to make their maximum effort. The experimenter timed the task. After 30 seconds, and regardless of the subject's actual performance, he asked the subject, "Can you go faster?"

Sentence association.

Next, the experimenter introduced the sentence association test as follows: "I am going to show you some cards, and each card has a sentence printed on it. Please read each sentence aloud, loud and clear, then tell me the first thing that comes to mind after reading the sentence." The 15 stimulus sentences were presented in the same sequence for all subjects. A two minute rest period followed.

TAT test.

The TAT task was introduced with the same instructions used in Study 1. TAT responses were tape-recorded for subsequent analysis.

Unstructured inquiry.

The final phase of the laboratory session was a brief interview lasting approximately 15 minutes, beginning with an inquiry about the TAT stories just told. Subjects were asked to choose the TAT story that seemed most salient to them and were asked whether they thought the story might reflect their own feelings or experiences in any way. The ensuing discussion was open-ended.

Measuring Physiological Reactivity

Baseline measures of heart rate and diastolic blood pressure were computed by averaging the observations recorded during the last minute of the 10-minute baseline period; a baseline measure of systolic blood pressure was obtained by selecting the single highest value recorded during this minute. [8] Similarly, measures of HR and DBP under stress were computed by averaging all values recorded during the ensuing laboratory tasks, and a measure of SBP under stress was obtained by selecting the single highest value recorded during these laboratory tasks.

The rate-pressure product was computed separately for baseline and stress periods (RPP = HR*SBP/100). Baseline values were then subtracted from the corresponding stress measures to create reactivity scores (change scores).

Results Creation of Comparison Groups

Sixteen subjects scoring below 7 on the BDI and judged healthy by the clinician were classified as genuinely healthy. Twenty-one subjects scoring below 7 on the BDI and judged distressed by the clinician were classified as having illusory mental health. Thirteen subjects scoring above 7 on the BDI and judged distressed by the clinician were classified as manifestly distressed.

Physiological Measures

Figure 4 shows the mean changes in HR (from resting baseline) for subjects classified as having illusory mental health and subjects classified as genuinely healthy, in response to the various laboratory stressors. [9] As before, subjects with illusory mental health show greater reactivity during every one of the laboratory tasks.

With RPP as the measure of physiological activity, subjects classified as having illusory mental health showed greater coronary reactivity ($M = 59.1$) than genuinely healthy subjects ($M = 36.0$), $t(34) = 2.75$, $p = .01$, two-tailed. They also showed greater coronary reactivity than manifestly distressed subjects ($M = 33.7$), $t(31) = 2.97$, $p < .01$, two-tailed. Both HR and SBP contributed to these effects, with differences in HR reaching significance independently. Also, subjects with illusory mental health showed greater diastolic blood pressure reactivity ($M = 18.0$) than genuinely healthy subjects ($M = 13.5$), $t(35) = 2.14$,

p < .05, two-tailed. Findings involving diastolic blood pressure are unusual, and prior investigators have reported them only rarely.

A Stronger Version of the Illusory Mental Health Hypothesis

In Studies 1 and 2, clinical judgment acted as a moderator variable that changed the relation between self-reported mental health and physiological reactivity. For subjects judged distressed by the clinician, higher self-reported health was associated with greater physiological reactivity; for other subjects, this relation did not hold. These findings were replicated again in Study 3. For subjects judged distressed by the clinician, the correlation between BDI scores and physiological reactivity was $r(32) = -.37$, $p < .05$, and for other subjects the correlation was $r(37) = +.10$, *ns.* These correlation coefficients differ significantly from one another, $z = 2.0$, $p < .05$, two-tailed.

The importance of clinical judgment is illustrated in Figure 5 , which shows the least squares regression lines relating BDI scores and physiological reactivity, separately for distressed and comparison subjects. Once again, the findings suggest that self-report measures of mental health assess different things in different people. For some people, low scores on the Beck Depression Inventory may signify relative psychological health. For others, low BDI scores appear to reflect not health but psychological defense.

Summary of Major Findings

Subjects completed standard self-report measures of mental health and were also evaluated by clinical judges. Subjects reporting psychological health and judged healthy were classified as genuinely healthy. Subjects reporting psychological health but judged distressed were classified as having illlusory mental health (i.e., based on defensive denial of distress). Subjects reporting distress and judged distressed were classified as manifestly distressed.

In response to psychological stress, subjects with illusory mental health showed higher levels of coronary reactivity than either genuinely healthy or manifestly distressed subjects. This finding was obtained three times. The differences between illusory mental health subjects and genuine mental health subjects were not only statistically significant but were often large enough to be considered medically significant. In addition, subjects with illusory mental health showed significantly more evidence of psychological defensive processes in their verbal associations to threatening stimuli.

Subjective clinical judgment proved to be of critical importance in our analyses; it acted as a moderator variable that changed the relationship between self-report mental health scores and coronary reactivity. For subjects judged distressed, higher self-reported mental health was associated with *higher* levels of coronary reactivity (i.e., with more physiological evidence of distress). This is consistent with the hypothesis that the mental health scales were not assessing mental health in these subjects, but instead were assessing defensive denial. For subjects not judged distressed, self-reported mental health was not associated with physiological reactivity, or was associated with lower levels of physiological reactivity. This pattern was also obtained three times.

Discussion

Many psychological researchers rely on mental health scales as a matter of course. Nevertheless, the present findings suggest that the use of these scales is highly problematic. Among people who look healthy on standard mental health scales, it is possible to identify a subgroup of people who may not be

psychologically healthy at all. Three data sources converge in indicating that the "mental health" reported by many people is, in fact, illusory: (a) They are judged distressed by clinicans, (b) their verbal associations betray the operation of psychological defensive processes, and (c) they show high levels of physiological reactivity in response to stress. This last finding, which was obtained three times, suggests that illusory mental health may be a risk factor for physical illness.

Stated somewhat differently, the findings indicate that high scores on standard mental health scales (or conversely, low scores on standard measures of distress) may, in and of themselves, be uninterpretable. In fact, it appears that *mental health scales assess different things in different people.* For some people, mental health scales appear to be legitimate measures of mental health. For other people, these scales appear to measure defensive denial. There seems to be no way to know from the test score alone what is being measured in any given respondent.

The present investigation raises questions about thousands of published findings that rest on self-report measures of mental health. Because the use of self-report scales is so pervasive in psychological research, it is beyond the scope of this article to explore the implications of illusory mental health for all domains in which it may have relevance. The following brief comments are intended only to raise some relevant questions, not to provide exhaustive analyses.

Illusion and Well-Being

The counterintuitive view that self-serving distortions and biases go hand-in-hand with mental health has gained currency among academic researchers (e.g., S. E. Taylor & Brown, 1988). Some proponents of this view have gone so far as to suggest that an appropriate therapeutic intervention would be to "teach" depressed patients to deny and distort reality. Evidence for this "positive illusion" position comes from findings that depressives perceive the world relatively accurately, whereas "normal" subjects (i.e., people with low scores on self-report depression scales) display various self-serving biases and distortions. For example, the illusion of control phenomenon is seen in "normal" but not in depressive subjects (Alloy & Abramson, 1979 , 1988); the self-evaluations of "normal" subjects are distorted in self-serving ways, whereas those of depressives appear more accurate (Lewinsohn, Mischel, Chaplin, & Barton, 1980); and so on.

Our findings suggest a very different interpretation of the data. Positive illusion studies nearly always assess mental health using simple self-report scales. It is therefore likely that the "normal" groups in these studies contain a mix of genuinely healthy people (who may not distort very much) and defensive deniers (who may distort a great deal). The presence of the defensive deniers would account for the paradoxical finding that "normal" subjects distort more than depressive subjects. In short, positive illusion findings may be nothing but artifacts, due to researchers' inability to assess mental health in any meaningful way.

Indeed, one must have something approaching blind faith in self-report scales to conclude from existing evidence that illusory thinking is a component of mental health. Positive illusion studies show that some people perceive things in distorted, self-serving ways. They also show that these same people tend to look good on self-report scales (which ask transparent questions about how people perceive things). The most straightforward conclusion is that people who are prone to distort also give distorted responses to mental health scale items, and their scores simply cannot be taken at face value. For reasons that escape

us, many academic psychologists seem not to have considered this possibility.

Psychotherapy Outcome

Studies of psychotherapy outcome often yield perplexing and disappointing results. For example, anticipated differences between treatment modalities (e.g., cognitive therapy, interpersonal therapy, drug therapy) have not emerged reliably; differences between treatments conducted by experienced and inexperienced therapists, or professional therapists and laymen, have not emerged reliably; multiple outcome measures often fail to converge; and so on (see Lambert, Shapiro, & Bergin, 1986 , for a review).

The problems inherent in designing and interpreting psychotherapy outcome studies are complex and may be compounded by the use of self-report data to assess outcome (cf. Loevinger & Ossorio, 1959). For example, the failure of self-report measures to converge reliably with psychotherapists' ratings may be due, at least in part, to the distorting influence of psychological defenses on self-report scales.

More problematic is the possibility that mental health scales assess different things in different people. If, for some subjects, self-report measures of mental health assess not health but defensiveness, then an effect of successful psychotherapy might be to lower scores on these measures (or conversely, to raise scores on measures of distress). For less defended subjects, successful therapy should raise scores on measures of mental health. Failure to distinguish between these groups could therefore cloud the results of psychotherapy outcome studies, obscuring treatment effects that may actually be present. This reasoning also implies that treatment groups in psychotherapy outcome studies should, in general, show larger variances on outcome measures than control groups (because treatment would push scores in different directions for different subjects). This is, in fact, a common finding.

Links Between Psychology and Physiology

Interest in somatic correlates of psychological variables has waxed and waned over the years. The pattern has been one of bursts of enthusiasm for a program of research, followed by a decline of interest in the face of confusing and contradictory data. Although many laymen and medical professionals are convinced that psychological states are linked to physiological processes, effects have been difficult to demonstrate empirically. The failure of psychological measures of anxiety to converge reliably with physiological measures is but one example of this.

There is doubtless a confluence of factors working to obscure relationships between psychology and physiology. The present findings suggest that one source of difficulty may be the inadequacies of self-report scales. Note that in all three studies reported here, clinical judgment acted as a moderator variable that *reversed* the relation between self-reported mental health and physiological reactivity. For subjects judged distressed by the clinicians, the Neuroticism and Beck Depression scales correlated negatively with physiological reactivity, whereas for other subjects the correlation was positive. Had we not differentiated between these subjects, there would have been no statistically significant findings. Perhaps when researchers begin to take the concept of psychological defense seriously, and design research accordingly, effects will emerge more reliably.

Related Research

We are unaware of any prior research using clinical judgment as we have used it here, to identify defensive deniers among people who look healthy on mental health scales. There have, however, been efforts to identify defensive deniers using self-report scales alone. Research by D. Weinberger and his associates on the "repressive coping style" is among the most important of these efforts. D. Weinberger et al. (1979) divided subjects who reported psychological health into two subgroups on the basis of their scores on the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1964). Subjects reporting psychological health who had low Marlowe-Crowne scores were classified as low anxious, and subjects reporting psychological health who had high Marlowe-Crowne scores were classified as repressors. Consistent with the present findings, repressors were more physiologically reactive than low-anxious subjects, as assessed by heart rate and electrodermal response. Subsequent investigations have linked the repressive coping style with other somatic outcomes (e.g., Jensen, 1987 ; see D. Weinberger, 1990 , for an overview).

The Weinberger approach represents an advance over prior efforts to assess psychological defense using self-report scales. The limitation of the approach is that it is able to identify only a subset of defensive individuals. The problem is that the Marlowe-Crowne scale does not assess defensiveness per se but instead assesses a personality style characterized by oversocialization and overcontrol (inhibition) of impulse and affect (D. Weinberger, 1990). [1] Although these traits, taken to extreme, surely reflect a form of psychological defense, many defensive deniers will not exhibit these traits at all. On the contrary, defensive denial may just as easily express itself through undersocialization and undercontrol of impulse and affect, or it may express itself in domains unrelated to socialization and impulse expression.

For example, adolescents with conduct disorders typically show high levels of defensive denial yet would score low on the Marlowe-Crowne Social Desirability scale. Likewise, the Weinberger approach would not identify individuals with histrionic or narcissistic personality disorders, both of which are characterized by high levels of psychological defense but not by oversocialization or overcontrol of impulse and affect.

Our approach attempts to assess psychological defense per se, unconfounded by a specific personality style. Analyses to date suggest it is successful in doing so, identifying defensive deniers across a broad spectrum of character styles and psychopathology. Perhaps for this reason, the present approach yields effect sizes larger than those typically reported in studies linking psychological measures with physiological outcomes (cf. Friedman & Booth-Kewley, 1987).

On the Role of Clinical Judgment

The findings reported here rest largely on clinical judgment. The clinical judgments provided information about mental health that was, apparently, not available from "objective" mental health scales. Qualitative clinical methods have long ago fallen into disfavor among many psychological researchers, and much has been published about the inadequacies of clinical judgment. Indeed, a culture has developed among many academic psychologists in which it is considered acceptable and even laudable to disparage clinical methods. For example, one mainstream social psychology text (Myers, 1983) groups clinical personality assessment with fortune telling and astrology, in a chapter devoted to debunking irrational

beliefs. Perhaps it is time for researchers to attempt to understand not just the weaknesses but also the strengths of clinical judgment.

If human communication can carry meanings beyond its face value or manifest content, it is probably also true that humans have been uniquely endowed, over the course of evolution, with the capacity to understand these meanings. Given our present state of knowledge, the clinical judge may still be the only "measurement instrument" capable of registering some of the phenomena of greatest interest to psychology and psychiatry (cf. Sawyer, 1966).

In our enthusiasm for measures that appear "objective," we must be careful that we do not lose the ability to study what is psychologically important.

References

Alloy, L. B. & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General, 108*, 441-485.

Alloy, L. B. & Abramson, L. Y. (1988). Depressive realism: Four theoretical perspectives.(In L. Alloy (Ed.), *Cognitive processes in depression* (pp. 223—265). New York: Guilford.)

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561-571.

Block, J. (1965). *The challenge of response sets.* (New York: Appleton-Century-Crofts)

Block, J. (1978). *The Q-sort method in personality assessment and psychiatric research.* (Palo Alto, CA: Consulting Psychologists Press. (Original work published 1961))

Block, J. H. & Block, J. (1980). The role of ego-control and ego-resiliency in the organization of behavior.(In W. A. Collins (Ed.), *Minnesota symposia on child psychology* (Vol. 13, pp. 39—109). Hillsdale, NJ: Erlbaum.)

Bronson, S. (1982). A psychoanalytic interpretation of fear of success in males and females.(*Dissertation Abstracts International, 43,* 06B. (University Microfilms No. DDJ82—24916).)

Contrada, R. & Krantz, D. (1988). Stress, reactivity, and Type A behavior: Current status and future directions. *Annals of Behavioral Medicine, 10*, (2)-64.

Crowne, D. P. & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence.* (New York: Wiley)

Derogatis, L. R., Abeloff, M. D. & Melisaratos, N. (1979). Psychological coping mechanisms and survival time in metastic breast cancer. *Journal of the American Medical Association, 242*, 1504-1508.

Diamond, M. O. (1983). The transition to adolescence in girls: Conscious and unconcious experiences of puberty.(*Dissertation Abstracts International, 44,* 10B. (University Microfilms No. DEQ84—02270))

Dohrenwend, B. S., Shrout, P. E., Egri, G. & Mendelsohn, F. S. (1980). Nonspecific psychological distress and other dimensions of psychopathology. *Archives of General Psychiatry, 37*, 1229-1236.

Eysenck, H. J. & Eysenck, S. B. G. (1975). *Eysenck Personality Questionnaire manual.* (San Diego, CA: Educational & Industrial Testing Service)

Fisher, J. (1959). The twisted pear and the prediction of behavior. *Journal of Consulting Psychology, 23*, 400-405.

Friedman, H. S. & Booth-Kewley, S. (1987). The "disease-prone personality": A meta-analytic view of the construct. *American Psychologist, 42*, 439-555.

Friedman, M. & Ulmer, D. (1984). *Treating Type A behavior and your heart.* (New York: Knopf)

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*, 26-34.

Gotlib, I. (1984). Depression and general psychopathology in university students. *Journal of Abnormal Psychology, 93*, 19-30.

Greenwald, D. F. (1977). Environmental engagement and disengagement in men with gastrointestinal and dermatological disorders.(*Dissertation Abstracts International, 38,* 11B. (University Microfilms No. DDJ78—04713))

Hare, R. D. (1966). Denial of threat and emotional response to impending painful stimulation. *Journal of Consulting Psychology, 30*, 359-361.

Hare, R. D. (1973). Orienting and defensive responses to visual stimuli. *Psychophysiology, 10*, 453-464.

Horowitz, L., Inouye, D. & Siegelman, E. (1979). On averaging judges' ratings to increase their correlation with an external criterion. *Journal of Consulting and Clinical Psychology, 47*, 453-458.

Jaffe, M. R. (1985). Remembrances of things past: Psychosexual themes and quality of self-other differentiation in early memories (Doctoral dissertation, University of Michigan, 1984). *Dissertation Abstracts International, 46*, -07B.

Jensen, M. R. (1987). Psychological factors predicting the course of breast cancer. *Journal of Personality,, 55*, 317-342.

John, O. (1990). The big-five factor taxonomy: Dimensions of personality in the natural language and in questionnaires.(In L. Pervin (Ed.), *Handbook of personality theory and research* (pp. 66—100). New York: Guilford.)

Kneier, A. W. & Temoshok, L. (1984). Repressive coping reactions in patients with malignant melanoma as compared to cardiovascular disease patients. *Journal of Psychosomatic Research, 28*, 145-155.

Kinsman, R. & Staudenmayer, H. (1978). Baseline levels in muscle relaxation training. *Biofeedback and Self-Regulation, 3*, 97-104.

Kitamura, K., Jorgensen, C. R., Gobel, F. L. & Wang, Y. (1972). He modynamic correlates of myocardial oxygen consumption during upright exercise. *Journal of Applied Physiology, 32*, 516-522.

Krantz, D. & Manuck, S. (1984). Acute psychophysiologic reactivity and risk of cardiovascular disease: A review and methodological critique. *Psychological Bulletin, 96*, 435-464.

Lacey, J. & Lacey, B. (1958). Verification and extension of the principal of autonomic response stereotypy. *American Journal of Psychology, 71*, 50-73.

Lacey, J. (1959). Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome.(In E. A. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy* (pp. 14—37). Washington DC: American Psychological Association.)

Lambert, M., Shapiro, D. & Bergin, A. (1986). The effectiveness of psychotherapy.(In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (pp. 137—211). New York: Wiley.)

Lazarus, R. & Alfert, E. (1964). Short-circuiting of threat by experimentally altering cognitive appraisal. *Journal of Abnormal and Social Psychology, 69*, 195-205.

Lewinsohn, P., Mischel, W., Chaplin, W. & Barton, R. (1980). Social competence and depression: The role of illusory self-perceptions. *Journal of Abnormal Psychology, 89*, 203-212.

Loevinger, J. & Ossorio, A. (1959). Evaluation of therapy by self report: A paradox. *Journal of Abnormal and Social Psychology, 58*, 392-394.

Mandler, G. M., Mandler, J. M., Kremen, I. & Sholiton, R. D. (1961). The response to threat: Relations among verbal and physiological indices. *Psychological Monographs, 75(9, Whole No. 513)*,

Manuck, S. B., Kaplan, J. R. & Clarkson, T. B. (1983). Behaviorally-induced heart rate reactivity and atherosclerosis in cynomolgus monkeys. *Psychosomatic Medicine, 45*, 95-108.

Mathews, K. A. & Haynes, S. G. (1986). Type A behavior and coronary disease risk: Update and critical

evaluation. *American Journal of Epidemiology, 123*, 923-960.

Mayman, M. (1963). Psychoanalytic study of the self-organization with psychological tests.(In *Proceedings of the Academic Assembly on Clinical Psychology* (pp. 97—117). Montreal, Quebec, Canada: McGill University Press.)

Mayman, M. (1968). Early memories and character structure. *Journal of Projective Techniques and Personality Assessment, 32*, 303-316.

Mayman, M. & Faris, M. (1960). Early memories and relationship paradigms. *American Journal of Orthopsychiatry, 30*, 507-520.

McCrae, R. R. (1982). Consensual validation of personality traits: Evidence from self-reports and ratings. *Journal of Personality and Social Psychology, 43*, 293-303.

McCrae, R. R. & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observes. *Journal of Personality and Social Psychology, 52*, 81-90.

McDougall, J. (1989). *Theaters of the body.* (New York: Norton)

Myers, D. G. (1983). *Social psychology.* (New York: McGraw-Hill)

Parsons, O., Fulgenzi, L. & Edelberg, R. (1969). Aggressiveness and psychophysiological responsiveness in repressors and sensitizers. *Journal of Personality and Social Psychology, 12*, 235-244.

Pennebaker, J. (in press). Putting stress into words: Health, linguistics, and therapeutic implications.(*Behavior Research and Therapy.*)

Pennebaker, J. & Beall, S. K. (1986). Confronting a traumatic event: Toward an understanding of inhibition and disease. *Journal of Abnormal Psychology, 95*, 274-281.

Pennebaker, J., Kiecolt-Glaser, J. & Glaser, R. (1988). Disclosure of trauma and immune function: Health implications for psychotherapy. *Journal of Consulting and Clinical Psychology, 56*, 239-245.

Pennebaker, J. & Susman, J. (1988). Disclosure of traumas and psychosomatic process. *Social Science and Medicine, 26*, 327-332.

Robinson, B. F. (1967). Relation of heart rate and systolic blood pressure to the onset of pain in angina pectoris. *Circulation, 35*, 1073-1083.

Rose, A. M. (1956). Conscious reactions associated with neuropsychiatric breakdown in combat. *Psychiatry, 19*, 87-94.

Rosenberg, M. (1965). *Society and the adolescent self-image.* (Princeton, NJ: Princeton University Press)

Rushton, J. P., Brainerd, C. J. & Preisley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin, 94*, 18-38.

Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin, 66*, 178-200.

Scarpetti, W. (1973). The repression—sensitization dimension in relation to impending painful stimulation. *Journal of Consulting and Clinical Psychology, 40*, 377-382.

Scheier, M. F. & Carver, C. S. (1985). Optimism, coping, and health. *Health Psychology, 4*, 219-247.

Smith, T. W., Pope, M. K., Rhodewalt, F. & Poulton, J. (1989). Optimism, neuroticism, coping, and symptom reports: An alternate interpretation of the life orientation test. *Journal of Personality and Social Psychology, 56*, 640-648.

Spielberger, C. D., Gorsuch, R. L. & Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory.* (Palo Alto, CA: Consulting Psychologists Press)

Tanaka-Matsumi, J. & Kameoka, V. (1986). Reliabilities and concurrent validities of popular self-report measures of depression, anxiety, and social desirabilty. *Journal of Consulting and Clinical Psychology, 54*, 328-333.

Taylor, J. A. (1953). A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology, 48*, 285-290.

Taylor, S. E. & Brown, J. (1988). Illusion and well-being: A social psychological perspective on mental health. *Journal of Personality and Social Psychology, 46*, 621-635.

Temoshak, L. (1987). Personality, coping style, emotion, and cancer: Toward an integrative model. *Cancer Surveys, 6*, 545-567.

Watson, D. & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin, 96*, 465-490.

Watson, D. & Pennebaker, J. (1989). Health complaints, stress, and distress: Exploring the central role of negative affectivity. *Psychological Review, 96*, 234-254.

Weinberger, D. (1990). The construct validity of the repressive coping style.(In J. L. Singer (Ed.), *Repression and dissociation: Implications for personality theory, psychopathology, and health* (pp. 337—385). Chicago: University of Chicago Press.)

Weinberger, D., Schwartz, G. & Davidson, R. (1979). Low anxious, high-anxious, and repressive coping style: Psychometric patterns and behavioral and physiological responses to stress. *Journal of Abnormal Psychology, 88*, -1979.

Weinberger, J. & McClelland, D. C. (1990). Cognitive versus traditional motivational models.(In R. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition* (pp. 562—597). New York: Guilford Press.)

Wilder, J. F. (1968). *Stimulus and response: The law of initial values.* (Baltimore: Williams & Wilkins)

1

Lest the reader think we are unfair in characterizing mental health scale items as "transparent," consider some representative items: "I feel I am a complete failure as a person" and "I hate myself" (Beck Depression Inventory); "Would you call yourself a nervous person?" and "Have you ever wished that you were dead?" (Eysenck Neuroticism scale); "I am certainly lacking in self-confidence" and "I feel anxiety about something or someone almost all the time" (Taylor Manifest Anxiety Scale); "All in all, I am inclined to feel that I am a failure" and "I wish I could have more respect for myself" (Rosenberg Self-Esteem Scale). No doubt people who endorse such items are distressed. The question is, are people who deny them necessarily healthy?

2

People who deliberately dissemble or "fake good" would make up a third subgroup, but we are not directly concerned with them here.

3

At first glance, our concerns about mental health scales appear not to apply to the Minnesota Multiphasic Personality Inventory (MMPI) and other empirically derived scales. We believe they do apply and for two reasons. First, the MMPI scales were derived from an item pool containing a high percentage of transparent items–consequently the individual scales also contain a high percentage of transparent items. Second, the "healthy" criterion group in the original MMPI studies (to which psychiatric groups were compared) almost certainly contained defensive deniers. It did not take much to get included in this original healthy group: The main criterion was that one was not an institutionalized psychiatric patient.

4

Some researchers recommend that change scores be adjusted for initial baseline levels (e.g., Kinsman & Staudenmayer, 1978 ; Wilder, 1968). Consequently, the data reported here were reanalyzed by analysis of covariance, with effects of initial baseline controlled. The result of this covariate control was to *increase* the statistical significance of the reported comparisons. For ease of presentation, the simpler analysis strategy is reported in the text.

5

It would have been preferable to include only subjects judged "healthy" in the comparison group. Because relatively few subjects were classified as healthy ($n = 12$), we elected to increase the sample size at the expense of having a less pure comparison group. This is a conservative procedure, lessening the likelihood that the predicted difference will emerge.

6

Two outliers were excluded before computing this correlation. Had they been included, this correlation would have been larger, and the reported z score (for the difference between correlations) would have attained a higher significance level.

7

We forced the students' distribution to match the clinician's distribution rather than using the student's absolute ratings as a basis for classifying subjects. Had we relied on the students' absolute ratings, the reported effects would be smaller. The reason is that the students were relatively reluctant to classify people as distressed. In effect, we distinguished between the students' ability to make discriminations and their ability to make absolute judgments. We relied on the former but not the latter.

8

It became apparent that for some subjects a high percentage of the recorded systolic blood pressure (SBP) values recorded were inaccurate, falling below the actual values displayed by the Ohmeda blood pressure monitor. The problem was due to a technical failure of the J&J Enterprises physiological monitoring system, and it rendered meaningless a simple average of recorded SBP values. Despite this, occasional peaks in the data record were accurate (corresponding to the values displayed by the Ohmeda blood pressure monitor). Therefore it was decided that for all subjects, the single highest SBP value recorded during the base period would be used as the SBP baseline measure, and the single highest SBP value recorded during the subsequent tasks would be used as the measure of SBP under stress.

9

The plotted data points are averages of the heart rate (HR) readings recorded during the indicated segments of the laboratory session (the reader may recall that HR readings were recorded nearly continuously, at three-second intervals).

Some representative Marlowe-Crowne items are, "Before voting, I thoroughly investigate the qualifications of all the candidates," "I am always careful about my dress," "I am always courteous, even to people who are disagreeable," and "I have almost never felt the urge to tell someone off." Evidence indicates that people who endorse these items are not dissembling but believe they are responding truthfully (D. Weinberger, 1990).

Figure 1. Relation Between Eysenck Neuroticism Score and Clinical Evaluation

Plotted values indicate the actual number of data points at a given location
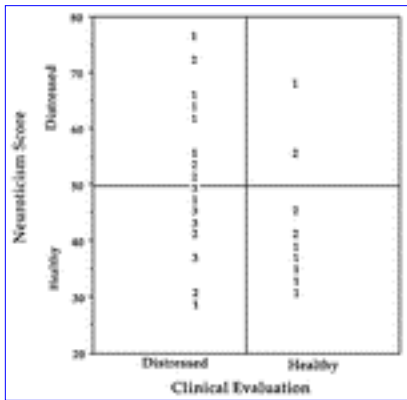
Figure 2. Mean Coronary Reactivity During Various Laboratory Tasks for Subjects With Genuine Mental Health and Subjects With Illusory Mental Health
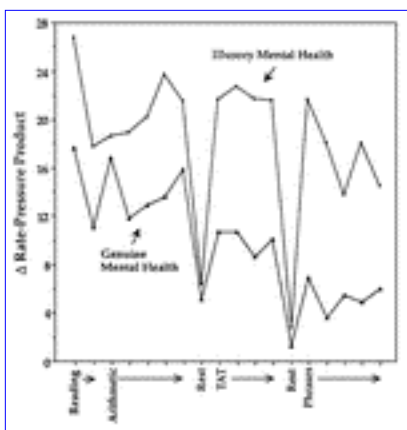

Figure 3. Relation Between Eysenck Neuroticism Score and Coronary Reactivity for Subjects Judged Distressed by the Clinician and for Comparison Subjects
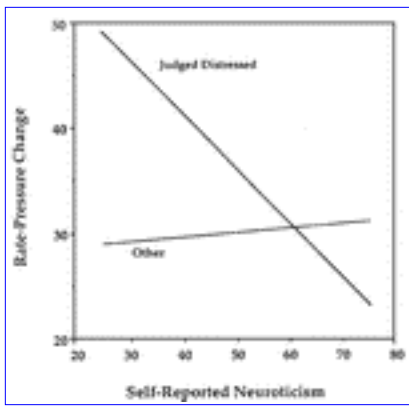
Figure 4. Mean Heart Rate Reactivity During Various Laboratory Tasks for Subjects With Genuine Mental Health and Subjects With illusory Mental Health
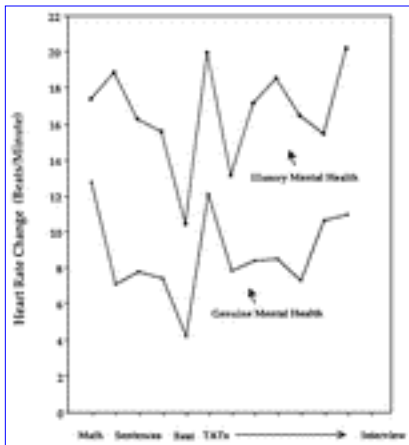

Figure 5. Relation Between Beck Depression Inventory (BDI) and Coronary Reactivity for Subjects Judged Distressed by the Clinician and for Comparison Subjects