Psychological Bulletin
1978, Vol. 85, No. 5, 1154-1176

Clinical Psychophysics: Applications of Ratio Scaling and Signal Detection Methods to Research on Pain, Fear, Drugs, and Medical Decision Making

John M. Grossberg and Bridget F. Grant
San Diego State University

Recent developments in psychophysics termed *statistical decision theory* (or *signal detection theory*; SDT) and *magnitude estimation* (ME) are being applied to an increasing variety of previously recalcitrant problems in clinical psychology and medical decision making. One major value of the SDT method lies in the separation of human decision making into a discrimination accuracy measure and a criterion or bias measure. This characteristic is helpful in analyzing confusing situations such as pain assessment, where any number of impressive procedures or manipulations can and do affect the person's willingness to report pain but leave unaffected the detectability of pain-producing stimuli. ME methods as applied to personality research do not limit an individual's assessment of his or her personal reactions (e.g., degree of anxiety or intoxication) to a narrow range of categories devised by the experimenter but allow each person to construct as sensitive and unique a scale as is required. These methods also show superior agreement with psychophysiological measures of such factors. ME and SDT represent a significant advance over current chaotic and unreliable practices in terms of their greater objectivity and precision, parsimonious use of a single language for the laboratory and the clinic, and potential for the quantification of subtle covert psychological behaviors. This review describes applications of SDT and ME to the assessment and understanding of pain, anxiety, psychoactive drugs, and medical decision making.

Although scientific progress has been slow in clinical psychology, it has not been for lack of important problems to study. Rather, weak investigative tools and conceptual confusions have hindered advancement. However, recent developments in psychophysics, termed *statistical decision theory* (SDT; sometimes referred to as *sensory decision theory* or *signal detection theory*) and *magnitude estimation* (ME), have shown great promise and wide applicability to vital human concerns far removed from simple sensory metrics. Like behavior modification techniques, these methods represent a kind of technological advance quite rare in psychology: the discovery and elaboration of general principles under rigorous laboratory conditions, followed by the gradual application of these principles to increasingly more complex life situations.

The SDT approach was developed for more refined measurement of sensory detection and recognition, an attempt to correct and surmount the well-known deficiencies of the sensory threshold construct (Price, 1966). For example, threshold values differed depending on the particular method of measurement used, which precluded comparisons across experiments. Further, the threshold measure with its major focus on accurate report neglected much useful and important information, such as the frequency and patterning of misses and false affirmatives, which were often regarded as errors to be balanced and neutralized. Finally, the threshold con-

Bridget F. Grant is now at Rutgers University.

The authors wish to thank Edward Alf, W. A. Hillix, and the reviewers of the manuscript for their helpful comments.

Requests for reprints should be sent to John M. Grossberg, Department of Psychology, San Diego State University, San Diego, California 92182.

cept did not account adequately for the many nonsensory influences on performance, for example, the psychiatric patient who only reports the strongest and most apparent stimulus values for fear that incorrect identifications (false affirmatives) might be interpreted as hallucinations.

In the SDT experiment the observer's behavior is separated into two subprocesses, a sensory detection or information process and a decision process, thus allowing the investigator to evaluate the observer's discriminative sensitivity separately from motivational components or response biases. The observer's response to a specified class of stimuli will vary from trial to trial, producing a discriminal dispersion assumed to be normal in form. Different stimulus classes will yield different discriminal dispersions, and the distance between midpoints of these distributions, in standard deviation units, defines the statistic d', which is an index of discriminability between stimulus alternatives. If the distance between the midpoints of two distributions of stimulus alternatives is small, the distributions will overlap considerably, and in the region of overlap the individual cannot reliably identify which of the two distributions was sampled. The degree of unreliability in his or her responses to the two stimulus classes is important in that it provides an empirical basis for estimating the distance between midpoints. It is further assumed that the observer, in deciding between alternatives, establishes a decision rule or criterion such that any stimulus that falls short of this value will elicit the alternative response. The location of the criterion in a discriminal distribution may be inferred from the observer's hit rate and false affirmative rate. With many sets of hit and false affirmative probabilities, a graph can be drawn in which the probability of saying *signal* given a signal (hit) is plotted as a function of the probability of saying *signal* given no signal (false affirmative). Such a graph is called a receiver operating characteristic (ROC) curve.

Thus, SDT purports to provide a clear and formal separation of performance into a measure of the observer's accuracy in discriminating two or more response classes and a criterion measure that reflects the influence of various nonsensory biasing factors, such as the observer's willingness to respond in a particular way when uncertain about the identity of some stimulus. It is evident that our cautious psychiatric patient would be displaying a conservative stance, adopting a very stringent criterion for reporting stimulation.

ME procedures are not as directly related to a specific theoretical foundation as is the SDT methodology but instead represent one of several ratio scaling techniques developed by S. S. Stevens and his colleagues for studying regularities in human sensory judgment. According to Stevens (1975), investigators in sensory psychophysics have known for decades that the Fechner equations representing the relationship between physical stimuli and psychological reactions were inadequate, but the prestige of Fechner and his modern supporters as well as the lack of a superior alternative retarded progress in this area. However, Stevens's development of ratio scales obtained through observers' direct estimates of stimulus and response characteristics led to the discovery of lawful psychophysical relationships superior to the Fechnerian formulation. Factors of major interest in clinical psychology such as emotional reactions differ in reported intensity and thus constitute prothetic continua in Stevens's terminology. For such variables, Stevens (1957, 1960, 1966) consistently found that reported (subjective) magnitude was a power function of stimulus magnitude. The underlying invariance then becomes the simple principle that equal stimulus ratios produce equal subjective ratios. The basic equation expressing this relationship is written as

$$\Psi = KS^n$$

and indicates that the perceived magnitude Ψ grows as the physical value of the stimulus S is raised to some power n. When this equation is converted to logarithmic form, it yields a linear equation that is represented by a straight line in log-log coordinates, with a slope corresponding to the exponent n.

The ME procedure is widely used to obtain judgments of stimulus intensity. The ob-

server is instructed to assign numbers to a series of stimuli, with the restriction that the numbers be made proportional to the perceived magnitude of the stimulation he or she experiences. The experimenter may furnish a standard or modulus by presenting a stimulus from the middle of the stimulus range and assigning some particular numerical value to it, or he or she may leave the observer free to select some personal standard. Using this procedure, Stevens found that the power law held for a wide variety of perceptual continua, including loudness, brightness, taste, and duration (Stevens, 1975). He also noted (1966) that although direct scaling methods and the power law were originally applied to problems of sensory scaling, they have shown equivalent precision and reliability in areas of broader social concern such as the measurement of seriousness of crimes, occupational preferences, aesthetic values, economic utility, and the quantitative relationship between frustration and aggression.

The present article reviews and discusses the many recent extensions of SDT and ME methodology to a variety of recalcitrant problems in personality and clinical psychology as well as in general clinical decision making. One major aim is to demonstrate the greater objectivity and precision afforded by these quantitative approaches. An additional benefit is the parsimonious use of the same language, methods, and principles for clinical and laboratory concerns, eliminating the necessity for dual formulations and special construction for personality variables. To the degree that these applications of modern psychophysics are successful, we can feel secure in extending the analysis of sensory interactions to emotional behavior and other such subtle covert psychological experiences. We now have the potential for measuring fear and anger as precisely as loudness and brightness. Statistical decision methods that maximized the detection of submarine echoes against a background of noise are now being used to amplify our accuracy in discriminating tubercular lesions from a background of X-ray shadows. As the following review demonstrates, such a versatile clinical psychophysics has already begun.

Areas of Application

Overview

The precision of psychophysical methods has recommended their application in areas where verbal responses in conjunction with stimulus manipulation represent the behavior of interest. ME methods have demonstrated generalizable functional relationships among such stimulus and response events and have provided a common mathematical language for describing disparate psychological activities. SDT methods have been especially valuable in formally separating attitudinal/bias factors from sensitivity to stimulation or skill in making discriminations among similar stimuli. Sophisticated research programs on pain, anxiety, and psychoactive drugs have made frequent use of these procedures. There have also been less frequent applications to miscellaneous clinical problems such as medical and psychological assessment and posttraumatic and geriatric memory defects.

Pain

It has long been known that reported pain is a complex response pattern under the control of a variety of factors that include tissue pathology, expectancies, prior pain experiences, and audience reactions. Quantification of pain responses was greatly advanced by Stevens's and, later, Ekman's attempts to determine whether the power law held for unfamiliar stimuli such as aversive electric shock. Stevens, Carton, and Shickman (1958), in a series of four experiments with colleagues and graduate students as subjects, verified that the power law did indeed describe the relationship between increasing shock to the fingers and observers' magnitude estimates of shock intensity. The slope of the reported intensity curve was the steepest noted for any sensory continuum, with an average exponent of 3.5. Ekman, Frankenhaeuser, Levander, and Mellis (1964) improved and expanded this work by using a range of shock intensities twice as wide as that of Stevens et al. Their data were also fitted by a power function, but the obtained exponent of 1.81 was half that obtained by Stevens et al. Ekman attributed this differ-

ence to the greater range he employed, citing other investigators to support the assertion that slope varies inversely as a function of range. However, Ekman's observers were instructed to rate the shocks according to degree of unpleasantness rather than intensity, with unknown effects on performance. In a second experiment, Ekman, Frankenhaeuser, Levander, and Mellis (1966) analyzed the relationship between shock duration and intensity and observers' estimates of duration and unpleasantness. When low-intensity stimulus values were used, it was necessary to add a negative constant to the absolute threshold in order to obtain a power function relationship between observers' unpleasantness and duration estimates and the experimenter-controlled shock parameters. The constant took into consideration the distance above the absolute threshold. Cross, Tursky, and Lodge (1975) used a sophisticated cross-modality matching procedure in an attempt to produce an electric shock scale that was free of regression biases, range effects, and threshold anomalies. Subjects judged the apparent magnitude of shocks and line lengths by squeezing a dynamometer, adjusting the loudness of a noise, and verbally estimating magnitude. All the resulting cross-modality matching functions were well described by power functions. However, the characteristic exponent describing the relationship between strength of electric shock and subjects' estimates was approximately 2.2.

Adair, Stevens, and Marks (1968) used radiant heat pain stimuli presented by a Hardy-Wolff-Goodell dolorimeter. This device consists of a powerful projection lamp beam focused on a circular aperture. The heat from the aperture is applied to small, previously blackened skin areas, with heat intensity controlled by varying the voltage to the lamp. Adair et al. corroborated the power function relationship between pain responses and stimulus intensities measured as distances above the absolute threshold. Hilgard (1967, 1969) used magnitude estimates of pain as a dependent variable in studying the analgesic effects of hypnosis in college students with varying hypnotic suggestibility scores. Using cold pressor stimulation (Hilgard, 1967), he found that a power function described the relationship between pain state reports and immersion time in cold water. In a second experiment (Hilgard, 1969) he reported similar linear relationships between numerical pain estimates and both water temperature in the cold pressor test and duration of arm tourniquet application in ischemic pain. These data were so orderly compared with those provided by any other method for measuring pain that Hilgard (1969) recommended ME as follows:

> I emphasize these findings as a reply to those who would degrade the subject's statements as being "merely" verbal reports, as though some sort of physiological response would be sounder. I wish to assert flatly that there is no physiological measure of pain which is either as discriminating of fine differences in stimulus conditions, as reliable upon repetition, or as lawfully related to changed conditions, as the subject's verbal report. (p. 107)

Hilgard thus found quantitative pain reports reliable and useful, but he stated that many aspects of hypnotic analgesia remained a puzzle. His perplexity in large measure reflects the classical confounding of pain detectability with biasing factors affecting the criterion for reporting pain, a situation that cannot be resolved by merely adopting a superior system, ME, for ordering responses. SDT methods have shown great promise for overcoming both this obstacle and the classical disagreement over how to quantify pain threshold, although the latter problem is quite complex (Chapman, 1975; Clark, 1974). Lloyd and Appel (1976) have presented a good survey of SDT pain research, an analysis of methodological adequacy, and procedures for avoiding or overcoming specific research problems such as unequal noise and signal distribution variances.[1] Nevertheless, a number of serious empirical problems remain; for example, some researchers fail to use a "zero pain" stimulus, that is, a stimulus value that evokes no pain response, equivalent to blank trials. Thus, when stimuli of varying intensities that evoke pain responses are presented, responses to all these nonzero distri-

[1] After the present article had been accepted for publication, we received a comprehensive negative assessment of the SDT model in pain research (Rollman, 1977) and a rejoinder (Chapman, 1977).

butions could shift significantly toward zero without shifting significantly relative to each other. The *d'* values would not change, and one could conclude erroneously that some analgesic manipulation was ineffectual (Hayes, Bennett, & Mayer, 1975).

More serious problems seem to be generated by confusion over definitions of pain and measurement procedures. According to McBurney (1975), the use of radiant heat stimulation in pain research precludes the application of SDT to the data, since it is not possible to present a stimulus that is either painful or not perceived at all (blank). In other words, weak intensities evoke warmth reports, not pain reports, so one cannot use SDT to study absolute sensitivity to pain per se, although SDT could be applied to differential discriminability. As McBurney (1975) stated, "If heat is used as the pain stimulus, an intensity too weak to evoke pain will produce the sensation of heat. So one cannot study the absolute sensitivity to pain, per se, using signal detection theory, because the independent variable is heat, not pain" (p. 66).

McBurney's objection may result in part from his use of the term *pain* for both a stimulus and a response. In psychophysical research on most sensory systems such as vision and hearing, we often use the same term for both the stimulus and the response. Thus, in simple situations we present various intensities of *light,* and the subject reports *light* or *no light*. Pain is a different case, however, since pain is the response term used by subjects and cannot be applied legitimately to the stimulus dimension, although McBurney and others have frequently done so. The term *pain* does not describe a stimulus continuum or a response continuum that is consistently correlated with only a single stimulus continuum. Clark, Yang, and Hall (1975, p. 67) pointed out that in their SDT research on pain, the independent variable was not pain but differing intensities of thermal radiation, to which subjects attached various labels such as "nothing," "warmth," "heat," or "pain." McBurney would apparently restrict the phrase *absolute sensitivity* to situations in which subjects employ only one class name for all the various degrees of stimulation within one particular stimulus dimension, from none through faintest to most intense. He explicitly accepted the SDT model for measuring the presence or absence of light, for example. He also accepted the SDT model as applied to the discriminability of various light intensities, with more intense light as signal and less intense light as noise. He objected to using SDT for discriminations in which subjects demonstrate a change in the response labels they employ, although the same stimulus continuum is used. Let us consider the weakest thermal stimulus, Intensity A, consistently called painful by subjects. The A intensity is that value just greater than B, the intensity consistently called heat. If SDT methods indicated that some analgesic agent now resulted in subjects consistently calling Intensity A *heat,* McBurney would presumably regard this as a change only in differential sensitivity. Such a narrow definition seems to restrict the term *absolute sensitivity* to the changes from zero stimulation to the subject's first pain response, ruling out changes from a prior heat response to the first pain response.

It should now be clear that it is incorrect to refer to absolute sensitivity to pain, since this phrase assumes pain to be a stimulus. The pain dimension may be thought of as a complex interactional field, and a preemptive definition of absolute sensitivity seems unnecessarily restrictive in studying it. As Clark et al. (1975) pointed out, legitimate applicability of the SDT model is independent of such constructs. ROC analysis can be used to evaluate detection or decision performance in any situation for which true states can be divided into two or more classes and for which one of two or more decisions must be selected by the decision maker. However, Clark et al. (1975) exaggerated in asserting that the values of *d'* they obtained "measure the capacity of the nervous system to transmit information about the amounts of thermal radiation reaching the skin" (p. 68). The *d'* measure is not a direct measure of sensory discriminability, since no direct physiological observations of skin or nervous system reactions enter the calculations. Instead, discriminability is inferred from the numbers assigned to subjects' responses in the pres-

ence of varying stimulus characteristics, thermal intensities in the present example. The *d'* statistic is a measure of how consistently separable or discriminable subjects *say* various states of the world are. This important issue is explained more fully in the Discussion section.

Although theoretical debates persist concerning the meaning of absolute pain sensitivity and the limitations of SDT procedures in pain measurement, researchers have amassed an impressive body of research. W. Crawford Clark was the first investigator to explore the unique advantages of SDT methods in attacking the puzzle of pain, and he subsequently contributed a large number of important studies. Clark's first study (1969) measured the effect of a quinine placebo on radiant heat pain produced by the dolorimeter. During each of two counterbalanced sessions, 22 pain subjects received 25 heat trials at each of five different intensities, using one of 12 categories ranging from "nothing" through "faint warmth" to "extremely painful" for reporting stimulation. During one of the sessions, each subject was given a quinine placebo and strong suggestions about its analgesic effectiveness. Results indicated that changes in heat intensity altered thermal discriminative sensitivity, *d'*, but the placebo did not. However, the placebo significantly increased criterion levels for reporting pain, C_w. Furthermore, when the data were converted to traditional threshold measures by the method of constant stimuli, the placebo manipulation produced a significantly higher "threshold." Clark concluded that decreased pain reports during placebo conditions probably resulted from the increased social cost of a pain response, a significant criterion shift, whereas pain detectability remained unchanged. These data call into question the utility of the threshold concept.

Clark and Goodman (1974) applied SDT methodology to the effect of verbal suggestion on two aspects of response to radiant heat stimulation, pain detection threshold and pain tolerance. The former was defined as the lowest thermal stimulus value termed pain, the latter as the time until withdrawal from the thermal stimulus. Using 40 subjects divided into four equal groups, two groups were given suggestions designed to raise or lower their criterion for reporting pain, and two groups were given suggestions designed to increase or decrease pain tolerance. During the first of two sessions, all subjects received 12 repetitions of six different heat intensities without suggestions. The second session was identical except that all received the appropriate biasing suggestion. Results indicated that calculated "thresholds" showed reliable changes in the direction of the suggestion, but that discriminability, *d'*, did not change. Thus subjects raised their criteria for reporting minimal pain and for withdrawal, but stimulus detection was not affected.

Feather, Chapman, and Fisher (1972) corroborated Clark's findings. Nine nursing students received 100 radiant heat trials in an initial session and an analgesic placebo prior to a second 100-trial session. There were no significant *d'* differences between sessions, but subjects were significantly more conservative in their criteria for reporting pain during the placebo session. Chapman and Feather (1973) next evaluated the role of emotional arousal on pain reports by comparing the effects of the tranquilizer diazepam with those of a placebo capsule. In the first of two sessions, 15 females received a placebo prior to ischemic pain stimulation, with pain tolerance duration and state-anxiety test scores as dependent variables. A second group of 15 females received diazepam in the first session, with the drug–placebo order reversed for both groups' second sessions. Both drug–placebo order and drug versus placebo significantly reduced anxiety reports and pain tolerance times. In a second experiment, comparing aspirin with diazepam, diazepam significantly increased pain tolerance time. The third experiment employed SDT methods to compare diazepam with a placebo, using radiant heat stimulation, which Chapman and Feather claimed did not produce as much emotional arousal as did ischemia. Following diazepam or placebo administration, 12 paid female subjects received 150 trials at each of five different stimulus intensities. There were no significant differences in *d'* or response bias measures between drug and placebo conditions, indicating that the various heat intensities were equally detectable and were re-

ported as readily with or without the tranquilizer. The investigators concluded that the differences obtained in the first two studies resulted from manipulation of the emotional or motivational aspects of pain stimulation rather than the sensory-discriminative aspects or willingness to report pain.

Clark and Marmor (Note 1) used SDT methods for evaluating emotional arousal during pain stimulation, utilizing autonomic nervous system responses. Eight subjects received eight electric shocks at each of eight different intensities plus 16 blank trials, with palmar skin potential responses above a criterion value and magnitude estimates of pain as dependent variables. The d' values were calculated for discrimination between adjacent stimulus intensities. Results indicated that verbal estimates of magnitude were significantly more sensitive to changes in shock intensities than were skin potential responses.

Clark and Mehl (1971) used radiant heat stimulation to investigate the role of age and sex in pain responses. Previous research employing the usual pain threshold construct had produced contradictory findings, probably attributable to the confounding of attitudinal and sensory factors. Clark and Mehl's "young" group of 16 males and 16 females had a mean age of 22 years, and the same number of "old" subjects had mean ages of 43 for males and 49 for females. Every subject received 16 radiant heat stimuli at each of six different intensities and rated each stimulus on a 10-point scale from "nothing" to "withdrawal." Results indicated that both young and old females had significantly higher criteria for reporting pain, but the old females also showed significantly poorer discrimination among stimuli termed painful than did any of the other groups. Harkins and Chapman (Note 2) extended the investigation of pain response and age by using direct electrical stimulation of tooth pulp rather than radiant heat dolorimetry to avoid the potential confounding of decreased cutaneous sensitivity with aging. Their 10 young and 10 old men (mean ages were 23.7 and 71.4, respectively) were given a state–trait anxiety scale prior to testing. Based on 100 practice trials with electric shock, low and high intensity levels were set for each subject. This was followed by 200 test trials during which subjects rated each shock by pressing one of six buttons representing a pain intensity scale. Results indicated that there were no significant young–old differences using the traditional pain threshold measure and no differences in reaction times. However, the old group was significantly more liberal in labeling stronger stimuli as pain, which perhaps was attributable to its reliably higher anxiety scale scores. Finally, the old group showed significantly less ability to discriminate (d') between low and high shocks, a finding similar to that of Clark and Mehl (1971) for old females.

Recent research has centered on the evaluation of clinical analgesia. Chapman, Murphy, and Butler (1973) used SDT to evaluate the analgesic properties of nitrous oxide, a commonly used anesthetic gas, and Clark and Yang (1974) used similar procedures to evaluate acupuncture analgesia. Both experiments used radiant heat stimulation, with category scale self-reports of pain as the dependent variable. Chapman et al. (1973) administered nitrous oxide to 14 male volunteers in one session and room air in another session, with order counterbalanced. During both sessions, four different thermal intensities were presented 50 times each in random order. The experimenters found that nitrous oxide significantly reduced discriminability of various pain levels (d') compared with room air and also reduced subjects' willingness to report pain (increased their pain report criterion). On the other hand, Clark and Yang (1974) reported that acupuncture significantly raised their subjects' criterion for reporting pain but did not affect pain discriminability. Their 12 subjects were given 24 radiant heat stimulations applied to both forearms before, during, and after acupuncture needle placements designed to anesthetize the target arm without affecting the control arm. There were no significant differences in d' between target arm and control arm or in comparisons before and after acupuncture in the target arm. However, during acupuncture, subjects showed a significantly higher criterion for reporting pain in the target arm.

Chapman, Gehrig, and Wilson (1975b)

criticized Clark and Yang's (1974) experiments as an unfair test of acupuncture on the following grounds: (a) Acupuncture is said to be most effective for the head and trunk region, not the limbs; (b) the needles were not left in place long enough; and (c) too few trials were employed for reliable SDT data analysis. Clark, Yang, and Hall (1975) replied that they had followed accepted Chinese practice for locus and duration. With respect to number of trials, their means were stable and consistent. With respect to number of stimulus trials, they had previously found dose-related analgesic changes in d', using Carbocaine-induced median nerve block with fewer than 12 stimuli per intensity.

Chapman et al. (1975a) tried to clarify some of these issues by using direct electrical stimulation of the tooth pulp as a pain stimulus to avoid the potential confounding effects of heat receptor cutaneous stimulation, as well as by directing analgesic manipulation to the head rather than the limbs. Subjects were 42 unpaid male volunteers, divided into three groups of 14, who received acupuncture, nitrous oxide, or no experimental manipulation. For each subject an approximate detection threshold value was determined in 20 preliminary trials, and four stimulus values were selected relative to this level, a zero value, a value below threshold, a medium value, and a high value. All subjects received 25 trials at each of these values for practice in rating the stimuli on a 7-point scale. They then received two experimental sessions of 300 trials each, 75 trials per stimulus level. Analgesic manipulations or control conditions were present during the second session, with the first session merely serving as a baseline. Results showed that nitrous oxide and acupuncture significantly diminished d', the discriminability of various pain levels, compared with no treatment. Both nitrous oxide and acupuncture groups also showed significant increases in bias against reporting pain, but the control group did not. Finally, nitrous oxide was effective primarily at low stimulus levels and had little effect at higher intensity levels. The investigators concluded that acupuncture had produced "weak hypalgesia" but not analgesia.

A later study by Chapman, Wilson, and Gehrig (1976) provided more detailed information about the acupuncture procedure itself. Subjects were four groups of paid male volunteers ($n = 15$ each) who received four different intensities of electrical tooth pulp stimulation, as in the previous study. After 100 practice trials, all subjects were tested in two sessions during each of which they were given 75 trials at each intensity, for a total of 300 stimuli per session. Subjects in the acupuncture group had a needle inserted at the prescribed point on each hand. A transcutaneous electrical stimulation (TES) group received electrical stimulation from a surface electrode at the prescribed acupuncture point on each hand. The sham acupuncture group had needles inserted in a nonprescribed locus on the hands, and the control group was given attention and hand palpation. SDT analysis indicated that both the TES and acupuncture groups' discrimination of pain stimuli (d') differed significantly from that of the control group, but the sham group's d' did not. Similarly, the acupuncture and TES groups but not the sham group demonstrated significantly less bias than did the control group. Further analyses indicated that decreased pain ratings at lower intensities were attributable to discriminability (d') changes. Based on these and previous findings, acupuncture would appear to be most effective with mild pain-producing stimulation.

Craig and his associates (Craig, Best, & Ward, 1975; Craig & Coren, 1975) investigated the effectiveness of more cognitive or social approaches to pain relief, in contrast with the physiological and pharmacological. To test the common traditional assumption that social factors such as suggestion merely affect pain reports rather than some central pain experience, Craig and his associates had subjects in the presence of either pain-tolerant or pain-intolerant models rate discomfort stimulated by electric shocks. Craig et al. (1975) had three groups of 10 women each use an ME procedure to rate their discomfort produced by a series of electric shocks increasing in intensity. One group was exposed to a shock-tolerant confederate model who always reported less discomfort to the same shock intensities. A second group was

exposed to an intolerant model who always indicated more discomfort than did the subject, and the third group served as a control. The tolerant-model group required a significantly larger increase in shock intensity before moving on to a higher reported discomfort level. The power functions relating reported discomfort to shock intensity reflected model influences, with the tolerant-model group showing the smallest exponent (flatter slope) and the intolerant-model group showing the largest exponent.

Craig and Coren (1975) applied SDT procedures in a similar experimental situation in order to determine whether changed pain reports reflected criterion shifts or pain discriminability changes. Three groups of 10 men each received two series of increasingly intense shocks that they rated from "undetectable" to "painful" in the presence of a pain-tolerant or pain-intolerant model or no model. Then all subjects were asked to rate 60 additional shocks, 12 at each of five different intensity levels, according to painfulness. The intolerant-model group showed a significantly larger discriminability index (based on d') for adjacent shock intensities, and the tolerant-model group's change in pain reports was attributed to their adopting a higher criterion for reporting pain. Craig and Coren concluded that social experiences can successfully alter the sensory qualities of the pain experience.

It is obvious that pain as a construct is very complex, and research in this area is unusually difficult. In addition to the common research problems posed by the choice of appropriate subjects and the selection of appropriate independent and dependent variables, there are also exceptional linguistic and conceptual problems, as exemplified by the debate over absolute pain sensitivity. The experiments reviewed in this section represent laboratory pain of unknown generalizability to clinical situations. Clinical human pain often involves fear of death, mutilation, or handicap, concern for expenses, reaction from significant others, and a host of other such factors.

In order to maintain maximum flexibility in mounting a scientific investigation of pain, it would be most useful to regard pain as the name for a global sociobehavioral interaction involving many components—neurophysiological, hormonal, anatomical, social, economic, and so forth. Sometimes pain responses are clearly the result of injury or tissue damage, but at other times the social consequences of pain responses are primary determinants. It would follow that depending on the specific pain situation, one or another of the various components would participate to a greater extent and would thus constitute the focus of research interest. In hypnotic analgesia, for example, subject expectancies or the costs and values associated with pain reports would usually be more appropriate research targets than would neurological gating mechanisms. The study of these social factors should not be construed as more superficial than endorphin assays or neurophysiological recordings. In other words, the choice of measures and procedures should be dictated by whatever aspect of the global pain complex is of primary interest. In our present state of knowledge it would be foolish to constrain pain research through narrow a priori decisions about "true" measures of pain. In this context, psychophysical procedures represent only one step forward on what will most certainly be a long but rewarding investigative journey.

Stress, Anxiety, and Clinical Fear

Stevens's use of direct sensory scaling methods in psychophysics lends itself to objectifying those subtle human behaviors in personality psychology that have been difficult to quantify. Traditional assessments of emotional reactions and other such personal behaviors of clinical interest have relied almost exclusively on category scales that require subjects to rank order their experiences. However, such ordinal scales are not only crude but fail to correlate well with physiological and behavioral measures (Lang, 1968) and distort judgments by forcing the observer to cram some perceptual continuum into whatever limits are specified by the experimenter's choice of range. However, ME procedures so successful in yielding consistent and orderly data in sensory psychophysics have demon-

strated similar utility when applied to stress, anxiety, and clinical fears.

Frankenhaeuser and her associates at the University of Stockholm were among the first to realize the potential of ME procedures for describing subtle covert psychological behaviors previously recalcitrant to reliable and precise measurement. They initiated a series of experiments to see if direct estimation procedures would succeed in revealing orderly relationships between various physiological indicants and self-reports during stress.

Frankenhaeuser, Sterky, and Järpe (1962) related self-reported stress to epinephrine secretion. Four male medical student volunteers were subjected to gravitational stress trials of up to 3 G in a human centrifuge. For each ride, heart rate was continuously monitored, and urinary catecholamines were sampled before and after sessions and during rest intervals separating the various G levels. For each ride, subjects used magnitude estimates to rate the stressfulness of the experience as a percentage of a standard that was each person's most stressful life experience. The investigators found the predicted power function relationship between self-reported stress ratings and epinephrine secretions, with self-reports almost directly proportional to this physiological variable.

Frankenhaeuser, Fröberg, Hagdahl, Rissler, Björkvall, and Wolff (1967) used a ratio estimation procedure to study the quantitative relationship among a more psychological stress stimulus, the Stroop Color-Word Interference Test, and several physiological indices of arousal. Subjects were 15 female psychology students who were given five sessions with the Stroop test. Dependent variables included percent accuracy on the test, magnitude estimates of the stressfulness of each session, skin conductance, and urinary epinephrine content. The outstanding characteristic of the data was close similarity in response patterns for the different variables, with a particularly closely fitting linear relationship between both physiological measures and self-reports of stress. Frankenhaeuser, Mellis, Rissler, Björkvall, and Patkai (1968) administered tests of level of aspiration, aggression, and anxiety to 25 male students and collected urine samples. On the Stroop test, subjects with high norepinephrine and high epinephrine levels performed better when they reported less stress, and those with low levels of these hormones performed better while reporting more stress. There were no reliable differences on any of the other tests. Patkai, Frankenhaeuser, Rissler, and Björkvall (1967) also used magnitude estimates of stress to investigate the differential effects of stress on norepinephrine as compared with epinephrine by utilizing longer stress periods. Results indicated linear increases in stress reports with increasing duration of stress sessions.

Peck (1966) suggested the term "thymometry" for the psychophysical measurement of feelings. In the first of two demonstrations, he injected 18 private psychiatric outpatients with varying doses of sodium amytal and instructed them to match loudness of a tone to the degree of anxiety reduction experienced. Palmar sweating and salivation were concurrently recorded. The raw data showed general agreement between the decline in the audiometer setting and the reduction in the physiological measures. These findings were replicated with an additional 26 patients, using carbon dioxide inhalation as the independent variable. Peck's results must be regarded as tentative because he presented no details about physiological recording methods and no statistical tests.

Franzini (1970) used magnitude estimates to study the degree of aversiveness of various time intervals between a signal and an electric shock. For 60 student subjects a light preceded a low, medium, or high level of shock at each of four different time intervals. Results indicated that magnitude estimates of aversiveness were an increasing linear function of the interval between signal and shock.

Neufeld (1975) used an SDT methodology to test the therapeutic potential of two kinds of stress-alleviating attention instructions. A group of 60 undergraduates were shown slides of mutilated homicide or skin disease victims. Half the subjects heard an "intellectualization–denial" tape describing how useful the slides were in crime detection or medical training, and the other half heard an irrelevant "study habits" lecture. Dependent

measures were skin conductance responses and category scale ratings of stressfulness. Although there were no group differences in skin conductance, the intellectualization instructions significantly reduced stress ratings as indexed by decreases in d', the discriminability of the stress produced by differing degrees of mutilation. In other words, subjects who heard the study habits tape continued to be sensitive to differences in stressfulness among the slides, but the intellectualization–denial tape produced an attenuated reaction. Wolpe (1963), in a pioneering application of psychophysical methodology to phobias, furnished evidence that the growth and decline of patients' fears were orderly and quantifiable. He noted that in systematic desensitization treatment, the number of imaginal presentations or visualizations of scenes required to make a given amount of progress seemed to vary systematically. In phobias where fear increased as the patient approached the feared object, for example, the number of scene repetitions needed to reduce self-reported fear to zero was low when the imagined object was distant but increased with proximity, the cumulative curve corresponding to a positively accelerating function. In agoraphobics the number of scene repetitions was initially high and fell progressively as distance grew, the cumulative curve corresponding to a negatively accelerating function. Wolpe then examined the course of desensitization for 20 phobias in 13 of his patients and found that the relationship between number of scene repetitions and attained percentage of recovery was well described by a power function relating self-reported fear magnitude and stimulus magnitude. Tryon (1977) recommended a magnitude estimation procedure for hierarchy construction in systematic desensitization therapy. He presented rating scale data from 24 subjects that showed ME to be superior to other scaling methods in reflecting orderly increases in arousal as scenes in an imagined hierarchy increased in fearfulness.

Sullivan (1969, 1970, 1971, 1973) presented a series of experiments designed to extend ME procedures to the assessment of examination anxiety and snake phobias. In the first study (Sullivan, 1969) he had 20 college students select an aversive noise level that he termed the "minimal aversion threshold" (MAT). Subjects were also given a 7-point anxiety scale varying from "perfectly relaxed and comfortable" to "intense, unpleasant, intolerable" and were asked to judge what level on the scale would be the equivalent of their MAT noise. Subjects then matched noise level to the degree of anxiety they experienced at five separate points in time before, during, and after a nonstress ordinary lecture and at corresponding times during a midterm examination. Results showed that subjects consistently and reliably matched changes in self-rated anxiety with changes in noise amplitude, suggesting that cross-modal matching could be used to measure other such covert or experiential continua. Sullivan (1970) used a similar procedure to compare 26 subjects' direct estimates of anxiety with their category scale ratings of anxiety during both a midterm exam and a final. When category scales were plotted on the ordinate and magnitude estimates were plotted on the abscissa, the data yielded a concave-downward curve characteristic of intensive or prothetic dimensions. This result supported the hypothesis that anxiety could be regarded as an intensive continuum amenable to direct psychophysical scaling. Lundberg and Ekman (1970) found that simple power functions described the relationship between number of days prior to an examination and students' magnitude estimates of anxiety over a forthcoming exam, as well as their subjective estimates of time until the exam.

Grossberg (Note 3) had 20 snake-fearful and 20 nonfearful subjects give magnitude estimates of fear while recording heart rate and skin conductance as a snake cage was presented at varying distances on a runway. A live snake was in the cage during half of the trials, and instructions to imagine the snake were administered on the remaining trials. Results showed that the predicted power function held only for the relationship between fearful subjects' magnitude estimates of fear and distance from the snake, not for the relationship between heart rate or skin conductance and distance from the snake.

Sullivan (1971) also used ME to assess

anxiety in 10 snake-fearful females presented with a stuffed and then a real snake at varying distances and found a power function relationship between the subjects' estimates of fear and the reciprocal distance from the snake. Although the live snake evoked substantially higher estimates, the slopes for the live- and stuffed-snake conditions were very similar. Finally, Sullivan tested the validity of the previous procedure by allowing subjects to choose between increases in aversive white noise level and touching the snake. Subjects with the steepest slopes endured extremely loud noise levels, although tests of statistical significance were not presented. Sullivan (1973) elaborated on the noise versus snake procedure by presenting 25 snake-fearful females with a caged live snake on a runway, the cage moving closer unless the subject submitted to increasing amounts of noise. Results showed once again a power function relationship between magnitude estimates of anxiety and reciprocal distance from the snake and also a systematic quantitative relationship between steepness of the anxiety–distance function slopes and amount of noise endured. Eberle, Rehm, and McBurney (1975) used snake-fearful subjects' magnitude estimates of anxiety to select snake slides evoking different degrees of aversion. Then an additional group of 42 fearful subjects reported fear intensity by squeezing a hand dynamometer at 10 separate time points during the exposure of the slides, in order to assess fear habituation. Eberle et al. found similar habituation functions regardless of the rated fear intensity of the slide and also found increasing habituation with longer exposure durations.

Latané and Harkins (1976) measured stage fright by instructing 16 subjects to imagine reciting a poem before audiences systematically varied in size and status; the audiences were represented by color slides. Fear was assessed by adjusting the brightness of the screen and the loudness of a tone proportional to the subject's nervousness as each audience slide was presented. Results showed that reported stage fright was related by a power function to imagined audience size and audience status, defined as teenage listeners or mature adult listeners. This study would benefit from replication with preselected neutral and fearful subjects, real audiences, and perhaps additional fear measures. Chapman and Feather (1971) employed SDT methodology to quantify anxiety reports. They selected 30 student subjects with stage fright and had them rate repeatedly the amount of threat they experienced while contemplating five speech situations of varying fearfulness, along with one neutral situation. Half of the subjects rated the scenes while relaxed. Responses to each fear scene and the neutral scene were converted into conditional probabilities so that hit rates and false affirmative rates could be determined. The investigators calculated d' values based on subjects' ratings of the neutral scene, designated as a "noise" distribution, and their ratings of the fear scenes, designated as "signal" distributions. They found increases in d' when more arousing scenes were presented. In addition, the relaxed subjects had consistently higher d' levels, indicating greater discrimination of differences between fearful and neutral scenes.

In summary, ME procedures have yielded power functions consistent with the measurement of an intensive dimension in a variety of stress- or anxiety-arousing tasks. In addition, they appear to relate in very orderly fashion to physiological measures of anxiety and seem to be extremely valuable for assessing fear and fear reduction in clinical cases. There have been too few applications of SDT to evaluate its potential contribution to this area.

Evaluation of Psychoactive Drugs

In drug studies with human subjects, subjective bias is a major problem, and double-blind procedures are not always feasible or completely successful. Therefore, the precision and consistency of ME and the ability of SDT to isolate bias factors have recommended them for the evaluation of psychoactive drugs.

Frankenhaeuser and Järpe (1962) evaluated nitrous oxide by having 12 male medical students give numerical estimates of their intoxication, with "being a little high" as the standard. Subjects received five different sub-

anesthetic concentrations of nitrous oxide while performing a variety of motor, cognitive, and personality tests. They were run in squads of three and were rotated as active participants, with one subject inhaling nitrous oxide for 10 min. and reporting his intoxication level while the other two subjects simultaneously rated his performance. Then the second subject would inhale nitrous oxide, and so on until all three had received each concentration. Results showed very good agreement between self-reports and observer reports, particularly for nonperformance tasks. Ekman, Frankenhaeuser, Goldberg, Bjerver, Järpe, and Myrsten (1963) used similar procedures to evaluate self-reports and observer reports of alcohol intoxication. Subjects were eight male medical students, four of whom drank sufficient alcohol in 20 min. to bring blood alcohol level up to .1%. Then, over a 5-hr. period the four subjects were given arithmetic problems and a verbal memory task. During this time these subjects gave self-estimates of their intoxication, and the other four subjects also rated their intoxication. Dependent measures were blood alcohol level sampled periodically and arithmetic and memory test performance. Results showed marked similarity among changes in blood alcohol over time, self-estimates, and observer estimates.

Ekman, Frankenhaeuser, Goldberg, Hagdahl, and Myrsten (1964) used a similar procedure to study intoxication as a function of larger concentrations of alcohol (.33, .44, .55, and .66 g/kg of body weight). They administered three different concentrations of alcohol (whiskey) in six sessions 1 week apart to eight male medical students. Within each dosage condition, blood alcohol level was sampled periodically as subjects performed memory and arithmetic tests, gave estimates of their own drunkenness, and were rated by the other observers. Ekman et al. once again found that subjective estimate curves showed "remarkable" agreement with alcohol concentration, and self-reports were almost identical with the estimates of external observers. These findings should prove useful in current behavior modification procedures that aim to train alcoholics to monitor their own blood alcohol levels in order to stop drinking before they become unable to do so.

Rappaport, Silverman, Hopkins, and Hall (1971) used SDT to evaluate phenothiazine effects on a tone detection task with paranoid ($n = 22$) and nonparanoid ($n = 24$) schizophrenics compared with normals ($n = 16$). Research had suggested that phenothiazine reduced sensory sensitivity and central nervous system hyperarousal. Rappaport et al. tested a complicated set of interdependent hypotheses that assumed that paranoid schizophrenics were hypersensitive to stress and unable to focus attention normally, whereas nonparanoid schizophrenics were not. All patients were given three different dosage levels of phenothiazine and an auditory signal recognition task with two levels of difficulty. Results confirmed the hypothesis that phenothiazine helped paranoids focus attention, with increased ability to discriminate (d') stimuli, whereas nonparanoids showed decreases in d' at higher dosages.

Abel (1971) used SDT to determine whether marijuana impaired memory acquisition, storage, or retrieval. Subjects were divided into three groups, consisting of marijuana, tobacco placebo, and control, with 13 subjects per group. Prior to the experimental manipulation, all subjects received a word-list learning task followed by free-recall and recognition tests. Then the controls rested, the marijuana group smoked, and the placebo group smoked what they were told was marijuana. All subjects were then given 5 min. to recall words freely, followed by a delayed-recognition test containing previously learned words and new words. There were no significant differences among the groups in free recall, indicating that marijuana did not interfere with retrieval of previously learned material. However, on delayed-recognition tests the marijuana group showed significantly poorer ability to discriminate old from new words and gave significantly more false affirmatives, indicating less caution in reporting words as having been presented previously.

Moskowitz and McGlothlin (1974) studied the effects of four dosage levels of marijuana on an auditory signal-from-noise discrimination task during concentrated or divided at-

tention. Subjects were 23 healthy male paid volunteers, with each subject as his own control. During training with a "concentrated attention" tape, subjects were told to attend to the signal–noise stimulus that was presented in one earphone and to ignore the digits that were simultaneously presented in the other earphone. The "divided attention" tape instructed them to listen to both channels and report signals as well as recall digits. After practice, the subjects smoked and then received 100 trials under each attention condition for each of the four dosage levels. There were significant dose-dependent decreases in discriminative accuracy (d') for signals in both attention conditions and in the percentage of signals correct, but there was no significant effect on digit recall accuracy. In addition there were reliable dose-dependent increases in misses and false affirmatives, the latter corroborating Abel's findings.

It is evident from the above that a person's response to drugs may be accurately assessed by ME methods, which have demonstrated lawful relationships among the person's self-evaluation, the evaluation of others, and the corresponding degree of performance impairment. SDT methods have shown promise for specifying the nature of drug-produced impairment by separating discriminability changes from motivational factors, which were confounded in previous drug research.

Psychological Assessment and Medical Diagnosis

The vital importance of clinical judgment combined with its notorious unreliability have stimulated the search for better measurement and decision methods, resulting in recent applications of psychophysical logic (Hunt & Jones, 1962; Stone, 1968a). In a series of experiments, Stone investigated the relationship between psychiatrists' judgments and prognostic favorability, using magnitude estimates. In the first study, Stone and Skurdal (1968) sought evidence of the applicability of the power law to prognostic judgments by having 13 psychiatrist judges assess prognosis by using a 7-point category scale. When these data were plotted against magnitude estimates, they yielded the typical downward concavity similar in form to that obtained for brightness, loudness, and other sensory functions in Stevens's psychophysical research. By this criterion, then, prognostic estimates represented a prothetic or intensity continuum. The experimenters concluded that psychiatrists' opinions scaled in this manner were lawful and that the obtained subjective magnitude scale had the properties of a ratio scale. Stone (1969a) then examined the validity of the directly scaled prognostic estimates. He had 29 senior psychiatrists use ME to assess prognosis for patients selected from 33 published drug evaluation studies. The patients represented each of the 15 functional psychoses described in the *Diagnostic and Statistical Manual, Mental Disorders* (DSM-I; American Psychiatric Association, 1952) and all degrees of improvement from "recovered" to "unimproved." Data were plotted for 10 psychoses, with a power function representing the best least-squares fit to these data. Stone (1968b) also evaluated the relationship between psychiatrists' judgments of severity of patients' impairment and common clinical indices of patients' degree of pathology. Stone had 36 psychiatrists assign numbers proportional to the judged impairment severity of large groups of patients with diagnoses representing each of the 15 functional psychoses. He again found an approximate power function relationship when judgments were plotted against such pathology measures as median length of hospital stay, mean IQ score, and a Psychoticism score. Stone (1969b) next evaluated the function describing patients' judged susceptibility to external stress. Judges were 43 psychiatrists who used ME to assess susceptibility, and 13 of them also gave 1–7 category ratings on this dimension. Stone found low positive correlations between the two scaling methods. This failure to find the predicted power function relationship was attributed to the lack of consensus among judges in defining the stress dimension or to confusing scaling instructions. In addition, the 15 DSM-I psychosis categories may have been too heterogeneous, since post hoc analyses indicated a much better approximation to a power function when only

data for the 9 DSM-I schizophrenic reactions were plotted.

Stone and Linscheid (1971) pursued this point further and found that when the four depressive reactions were excluded from judgments, susceptibility to external stress ratings plotted against prognostic favorability ratings were well fitted by a power function. Stone (1970b) subjected these results to a more stringent test by using cross-modality matching to evaluate the validity of the power law relationships in psychiatric impairment and prognosis judgments. If the values of two prothetic continua sensation units are equated at various intensity levels, then the slope of a log-log plot of the resulting equal-sensation function should be given by the ratio of the two exponents of the two judgmental divisions. Stone accordingly plotted psychiatrists' magnitude estimates of prognostic favorability, and then impairment severity, on the ordinate, against patients' duration of hospitalization on the abscissa. The ratio of the two separate exponents thus obtained closely approximated the actual observed exponent obtained from directly plotting prognostic favorability against impairment severity.

With the introduction of the DSM-II (American Psychiatric Association, 1968) psychiatrists were required to use an adjective rating scale consisting of "none," "mild," "moderate," and "severe" when describing the role of precipitating stresses, degree of impairment, and premorbid personality predisposition. Stone (1970a) transformed such adjectives into a numerical category scale and compared it with psychiatrists' magnitude estimates on these dimensions. He found that these data no longer represented an intensity or prothetic dimension. However, since the adjectival scale demonstrated desirable equal-interval properties, it would be feasible to thus quantify the medical records of psychiatric patients.

Korboot and Damiani (1976) used SDT to evaluate the competing claims of two perceptual-deficit theories of schizophrenia: Yates (1966) had postulated that schizophrenics showed an abnormally slow rate of processing perceptual stimuli, and Chapman and McGhie (1962) had postulated hypersusceptibility to irrelevant perceptual stimuli, a poorly functioning filter system. Korboot and Damiani used five patient groups of eight subjects each: acute and chronic nonparanoid schizophrenic groups, acute and chronic paranoid schizophrenic groups, and a mixed neurotic group. Following a complex dichotic shadowing task, all subjects were given 100 auditory stimulation trials consisting of 24 signals, 24 intrusions, and 52 nonsignal trials. There were no significant differences in d' or criterion measures among psychotic and neurotic groups. The investigators concluded that the schizophrenic deficit did not reflect a sensory defect but a longer perceptual processing time, thus supporting Yates's view. However, there could have been a number of other causes for failure to find differences. Furthermore, these conclusions are weakened by the exceptionally small number of SDT trials, which would make an SDT analysis inappropriate, as is explained in the Discussion section.

Stenson, Kleinmuntz, and Scott (1975) found that SDT methods could be applied successfully to the kind of global personality judgments involved in MMPI profile analysis. Subjects were 2 experienced clinicians, 8 clinical graduate students, and 12 undergraduates, who were instructed to differentiate between 46 known abnormal MMPI profiles and 80 known normal ones. One of the two experts was required to sort the profiles into 12 categories from least to most likely to seek psychological treatment. The second expert sorted the profiles into five categories ranging from "definitely normal" to "definitely abnormal." Calculation of hits and false affirmative probabilities supported the assumptions of normality of discriminative distributions and equal variances. The two groups of student judges sorted the profiles into "hospitalized" and "never hospitalized" categories after being told that either 30% or 70% had previously been hospitalized. In a second session with the same profiles, the experimenter reversed the percentages for the groups, but there were no reliable order effects on d', and d' values remained constant. In a third experiment, the judges sorted pro-

files according to whether or not the patients needed hospitalization. One group was told that hospitalization was costly and stigmatizing; the other group was told that the patients were dangerous to society. This manipulation of costs and values did not alter *d'*. These results should encourage further use of SDT logic for such common clinical diagnostic decision-making situations.

Since SDT procedures were first applied to auditory signal-and-noise detection, it would seem natural to extend SDT to clinical audiometry, but there are several obstacles. Campbell (1965) acknowledged the defects of the threshold concept, such as the criterion shift/bias confounding. However, clinical audiologists must test each ear over a 100 dB (SPL) loudness range for at least seven different frequencies from 125 Hz to 8,000 Hz. For each ear, dividing the loudness range into approximately seven zones 15 dB apart would require seven *d'* values, which must be calculated for the seven frequencies, with each of these 49 *d'* values based on about 200 trials. By Campbell's estimate, then, the time involved for such testing would be prohibitive.[2] Campbell and Moulin (1968) explored the use of a forced-choice procedure that produced both *d'* and a threshold measure said to be free of response bias. There were 12 students with normal hearing in a "masked" condition who were instructed to detect auditory signals against a noise background. An additional 7 students with normal hearing in an "absolute" condition merely reported the presence or absence of signal. Each subject received 100 trials with neutral instructions, lax instructions to guess if doubtful, or strict instructions to be very sure before responding. Results indicated that all subjects adopted an extremely conservative criterion regardless of instructions. Campbell and Moulin concluded that for clinical purposes the *d'* measure was less useful than their bias-free threshold measure.

Lusted (1968, 1971a, 1971b) has written extensively about applications of SDT to X-ray diagnosis of pulmonary tuberculosis and breast cancer. Impetus for radical improvement in diagnosis came from studies reporting that trained radiologists overlooked 30% of the positive tuberculosis films in mass screening programs, specialists disagreed with their own diagnosis 20% of the time, and instructional manipulations produced marked shifts in diagnostic criteria from lax to strict. Lusted (1971b) summarized studies using SDT comparisons of diagnoses based on chest X-rays through direct plate viewing, TV viewing, or enhanced TV image viewing. The direct viewing yielded significantly larger *d'* values, which measured the diagnosticians' ability to discriminate known positive from negative plates. In another study involving discrimination of cancerous from noncancerous mammograms, secretaries and X-ray technicians showed significantly poorer discriminative ability (*d'*) compared with experienced radiologists. A further application involved comparisons of the chest X-ray diagnostic accuracy of technicians, radiology residents, and staff radiologists. After 5 months of training, the technicians' *d'* values were no longer significantly smaller than the radiologists', thus recommending the SDT procedure for assessing paramedical training in this area.

More recently, Lusted (1975) and his colleagues at the Center for Radiologic Image Research, University of Chicago, have successfully extended the ROC methodology to remote TV viewing of nuclear imaging scans (scintigrams). Diagnosis based on remote viewing was as accurate under certain conditions as that based on direct viewing; thus satellite stations that could not justify hiring high-salaried full-time nuclear diagnosticians could transmit diagnostic scintigrams to a central locus for reliable evaluation. Other pioneering work extended conventional ROC analysis to recognition as well as detection tasks (Metz, Starr, & Lusted, 1976; Starr, Metz, Lusted, & Goodenough, 1975). Here, the decisions involved lesion localization in addition to mere detection, and diagnostic

[2] One journal reviewer stated that variations in criterion produce *d'* changes on the order of 2:1 or 3:1 and, given *d'* proportional to acoustical power, a "threshold" variation of 3–5 dB (SPL). In clinical situations this variability is quite small compared with hearing losses of interest. SDT procedures are thus not worth the effort, and forced-choice methods represent a satisfactory compromise.

decisions involving three alternatives, Lesion Type A, Lesion Type B, or no lesion.

Adams, Hall, Pennypacker, Goldstein, Hench, Madden, Stein, and Catania (1976) used a modified method of constant stimuli to determine psychophysically the detectability of various-sized lumps in models of the human female breast. Three breast models were constructed, one with three large steel balls simulating lumps, one with five small steel balls, and one with no steel balls. On test trials, eight female subjects received seven sessions, each consisting of 6 practice palpation trials with the blank model and 10 trials with the other two models, all trials being appropriately randomized and counterbalanced. Results indicated that lump detection was a systematic function of lump size, with a markedly steep practice function that produced stable accurate detection after only 1½ sessions. The investigators stated that data from this feasibility study would be used to improve mass instruction and screening programs.

Several recent studies have used SDT analysis of memory processes in clinical situations. Memory items were presented during learning trials and were then re-presented interspersed with novel items. New items incorrectly identified as previously seen represented false affirmatives, and correctly recognized items represented hits (Banks, 1970). Gordon and Clark (1974) hypothesized that the widely reported memory deficits in the aged might well be a function of a conservative criterion placement attributable more to fear of making errors than to organic debility. They selected 22 well-educated old people (mean age of 71) and 22 young people (mean age of 25). Both groups were given prose paragraphs and were tested for immediate and delayed recall and recognition. Contrary to their expectations, Gordon and Clark found no reliable criterion differences between age groups but significantly lower d' values in the aged, indicating failure to discriminate items previously presented. E. Miller and Lewis (1977) noted that both elderly depressed patients and those diagnosed as having senile dementia did poorly on memory tests and that this diagnostic confusion could lead to therapeutic errors. They hypothesized that memory problems in depressed patients could be attributable to a more conservative criterion than a detection deficit and tested this assumption by using an SDT procedure. Three matched groups of 20 elderly depressed patients, senile patients, and normal controls were presented with 20 different geometric designs followed by 140 trials containing some new designs and some previously seen designs. Correct identification of a design as "old" was scored as a hit, and incorrect identification of a novel design as "old" was scored as a false affirmative. Results showed that senile patients had significantly lower d' scores, and depressed patients adopted a significantly higher criterion. The investigators concluded that the depressed group did not have a true sensory deficit but were more cautious in their decisions, and they recommended the SDT procedure to improve diagnostic accuracy. Brooks (1974) used an SDT analysis of visual recognition memory in patients with brain damage following head trauma to evaluate the relative roles of the damage and the fear of making mistakes. With 34 patients having three different degrees of brain impairment and a control group of 34 matched patients with leg injuries, Brooks found that the controls had a significantly larger d' or ability to discriminate novel items from previously presented ones and also that the brain-damaged groups had significantly more conservative identification criteria. These results indicated that even if brain-injured persons show little sensory impairment, they may nevertheless demonstrate poor memory performance due to a stricter criterion. Squire, Slater, and Chace (1976) tested the hypothesis that memory impairment after electroconvulsive therapy (ECT) was state dependent by applying the old-item versus new-item memory discrimination task at appropriate times before and after ECT. In their major experiment, 16 psychiatric patients were given a learning task 40 min. after ECT, with retention tested both 40 min. and 3½ hr. after a second shock. Then an equivalent learning task was administered 3½ hr. after the third shock, and retention was tested 40 min. and 3½ hr. after a fourth shock. A d' analysis showed no significant differences in

patients' ability to discriminate old from new items, regardless of the learning–retention interval, and thus it did not support the state-dependency hypothesis.

In the diagnostic assessment area, the principal contribution of ME methods has been the demonstration of psychophysical lawfulness in psychiatric diagnostic judgments on a variety of dimensions. SDT methods have demonstrated more socially useful potential in the diagnosis of serious physical disease, where it is obvious that medical judges' criteria are greatly affected by potential dangers of risky surgery, prior deaths of patients, and so on. Other important applications involve the effectiveness of training programs on diagnostic discriminations. Perhaps the more dramatic and visible advantages of sophisticated computer systems have overshadowed the largely untapped potential of SDT methods for improving medical decision making.

Discussion

In the 10 years since Price (1966) called attention to the potential of SDT methods in personality research, this method, as well as Stevens's scaling procedures, has had remarkably broad applications. The pragmatic contributions of these psychophysical tools to clinical and personality psychology include greater objectivity and precision of measurement and utilization of more of the information obtained in the experimental situation. For example, in SDT the subjects' proportion and type of errors as well as accurate responses are tabulated and constitute important data. Another contribution is the integration afforded by the parsimonious use of a common language and procedures for the clinic and laboratory. Clinicians frequently use simple ordinal scales to assess psychological reactions, often tests invented for the specific project at hand. Goldman (1977) compiled a directory of 339 commercially unpublished mental measures gleaned from articles in the 1970 editions of 29 professional journals of psychology, education, and sociology. He stated that less than 25% of them reported reliability or validity data and less than 60% reported the test format or number of items. Widespread use of ME procedures would serve to simplify such a chaotic situation.

To furnish a more balanced picture, it is necessary to discuss some of the controversial aspects of Stevens's psychophysical theory as well as SDT assumptions and methods. According to one common modern view of psychophysics (Luce, 1972), the sensory equipment of the individual transduces all kinds of physical input into a common mode that interacts with the central nervous system. The task of psychophysics is to discover the nature of this encoding and the way the input is processed by the organism in order to lead to responses. Based on his extensive research program, Stevens (1970, 1971) believed that his ratio invariance equation, which shows perceived magnitude growing as the physical value of a stimulus is raised to some power, was the best candidate for the psychophysical law for encoding and processing input. There have been, of course, other competing candidates for this role (see, e.g., Anderson, 1970) and attempts to reconcile the differences among the various models (Marks, 1974). At issue are fundamental questions about the general nature of measurement and the definition of psychological events, which are beyond the scope of this review. From the pragmatic standpoint of clinical application, the adequacy of some general law is of less immediate importance than the construction of methods that will yield reliable and systematic results across situations. Ratio scaling methods seem to approximate this requirement, although they are affected by a number of procedural factors that influence considerably the numbers assigned by subjects (Poulton, 1968). These factors include the range of stimulus values presented, instructions and training procedures for subjects, the distance of the stimuli above threshold, whether or not a modulus is used, the size of the modulus, and its locus relative to the extremes of the continuum. In other words, although Stevens's number-assignment methods have demonstrated substantial advantages in applied work, their limitations must be recognized. To predict ratings in any given context, one should assess both the psychological function and the

potential role of context factors (Birnbaum, 1974). With due consideration of these limiting factors, then, we can proceed to improve our assessment practices and find new areas of application.

One potentially fruitful line of research would be the use of ME methods to predict behaviors of clinical importance. Converting a power function equation between two variables to logarithmic form yields a linear equation that is represented by a straight line when plotted on log-log graph paper. The determination of two or more data points should enable one to predict response values for stimuli not yet presented. Assuming that Wolpe's (1963) finding of a power function relationship between phobic patients' number of fear scene repetitions and attained percentage of recovery applied to individuals, this would lend itself to the prediction of the specific number of sessions necessary to produce 100% recovery in systematic desensitization therapy. In a similar fashion, pain researchers or behavior therapists using aversion procedures might be able to predict the magnitude of a person's response to higher intensities of stimulation, given a power function relationship between the individual's reported pain and pain stimulus intensity. However, one must remember that the particular power function in each case is determined by individual subject factors, method of administering shock or pain-producing stimulation, and so forth, so it would not be possible to generalize the power function value from person to person.

There are also serious problems involving the legitimate application and interpretation of the SDT model. Many of the clinical investigations reviewed in the present article seem to identify the d' statistic with sensory sensitivity or discriminability, but the decision criterion measure, Lx or β, is assumed to be responsive only to biasing factors such as payoffs, attitudinal manipulations, and suggestion. For example, according to Clark (1974, p. 272),

> Sensory decision theory identifies two determinants of threshold performance. The first, d', measures discriminability and provides an index of sensory functioning that remains unaltered when variables such as attitude and expectation are varied. The other determinant, Lx, indexes the observer's criterion for emitting a particular response, and is a function of psychological (attitudinal) variables.

Although these statements are qualified by calling d' an *index* of sensory functioning and Lx a *function* of attitudinal variables, these cautions need more elaboration.

The assumption that d' is independent of criterion effects is one of the major reasons for recommending the SDT model in clinical and perceptual research, since one can then control, isolate, or quantify demand characteristics, sham sensory impairment, and failure to cooperate, which supposedly affect the criterion measure but not the sensitivity. A number of experiments have demonstrated that the d' statistic and the β statistic may be independently manipulated, but the assumption that any change in subject strategies will be reflected solely in a change in criterion but will not affect the sensitivity measure is invalid (see R. J. Miller & Leibowitz, 1976, for a good discussion relative to their hypnosis research).

To clarify this point, consider the primitive simple measurement task in which signal detection data are collected. The experimenter provides instructions and then records what the subject reports when he or she is presented with experimenter-controlled physical stimuli such as tones varying in intensity superimposed on a noise background. The subject must respond "noise" or "signal" on each trial in this simple example. After a large number of trials, an alert and cooperative subject's distribution of responses should approximate two Gaussian distributions along a continuum, a noise distribution on a signal-on-noise distribution. If the signal distribution is well separated from the noise distribution on the continuum, that is, if the discrimination was easy, there will be little overlap. However, the distributions will be closer together and overlap more as the discrimination becomes progressively more difficult. The d' statistic is equal to the distance between the distribution means and is measured in Z score units of the noise distribution, so d' is a measure of how separable the subject *says* the distributions are. In other

words, d' is not a direct measure of sensory sensitivity or sensory discriminability. The physiological responses of sense organs are not directly observed by electrodes or chemical reactions. Rather, sensitivity is inferred from the subject's verbal responses in the presence of varying sounds. In a way, this measure assesses accuracy of reporting or the strength of the relationship between the experimenter's presentation of some stimulus and the subject's report that a particular stimulus was presented. As R. J. Miller and Leibowitz (1976, p. 453) pointed out, if the subject is uncooperative, lying, or responding haphazardly, then d' will be small and a plot of the subject's responses will not show much discriminability between two separate states of the world but will approach a single distribution of guesses. Another problem arises if one group responds in a haphazard manner but a comparison group does not, as could occur with hypnotized subjects and nonhypnotized controls. In this context, what kind of alert cooperation could have been expected from Korboot and Damiani's (1976) four groups of hospitalized schizophrenic patients reporting for $2\frac{1}{2}$ hrs. on subtle differences between faint sounds presented via earphones?

It is evident that d' is not a pure measure of sensory sensitivity unaffected by situational contaminants. Instead it represents a measure of performance consistency over many trials, dispersion around some mean value taken as a typical performance of that subject in that particular setting under certain specific conditions. We are dealing with a complex interactional field whose interdependent components include sense organ reactions to physical stimuli, the subject's attitudes, and other factors. Investigators should therefore exercise considerable caution and restraint as they consider adopting SDT procedures and interpreting the meaning of the SDT statistical measures.

There remain many important research problems that do not involve the sort of subject variables that call into question the applicability of SDT methods. For example, modern medicine is making increasing use of diagnostic ultrasound in echocardiography, echoencephalography, and prenatal diagnostic studies that preclude X-ray examination. The ultrasonic output is used to make critical medical decisions and these pulse–echo signals are ideal candidates for SDT analysis. Similarly, ROC analysis would be of great benefit in comparing the diagnostic efficiency of different viewing procedures, such as nuclear scans, X-rays, and ultrasonograms, in various medical disorders.

Assuming that SDT is deemed appropriate for a given project, there are still a number of specific conditions that must be met in order to satisfy the assumptions of the model. Some of these were mentioned in the previous section on pain research. Additional requirements include a large number of presentations of each stimulus value in order to estimate ROC curves adequately. With the smaller number of trials typical of much applied clinical research, investigators should use more conservative nonparametric estimates of sensitivity and bias. It is also important to examine the shape of the distributions of events sampled and the variances of the samples. These considerations have been described and explained by Lloyd and Appel (1976), McNicol (1972), and Pastore and Scheirer (1974), and some excellent detailed examples have been provided by Clark (1974).

In summary, ME and SDT methods have a number of serious limitations but have demonstrated a number of advantages over current subjective clinical practices. The present review was designed to call attention to these developments in psychophysical measurement, since they do show great promise for increasing our ability to evaluate subtle and complex human behaviors so critical for effective psychological functioning. As can be seen from the wide range of clinical problems studied, it is evident that such procedures have the potential for making a substantial contribution to the understanding of psychological functioning and the alleviation of psychological distress.

Reference Notes

1. Clark, W. C., & Marmor, E. *Comparisons of sensory sensitivities* (d') *determined from verbal and autonomic responses to electrical stimulation.* Paper presented at the meeting of the Eastern Psychological Association, Philadelphia, April 1969.

2. Harkins, S. W., & Chapman, C. R. Detection and decision factors in pain perception in young and elderly men. In L. Storms (Chair), *Clinical psychophysics*. Symposium presented at the meeting of the Western Psychological Association, Los Angeles, April 1976.
3. Grossberg, J. M. *Clinical psychophysics: The relationship between magnitude estimates and physiological measures of fear*. Paper presented at the meeting of the Western Psychological Association, Seattle, April 1977.

References

Abel, E. L. Marijuana and memory: Acquisition or retrieval? *Science*, 1971, *173*, 1038–1040.

Adair, E., Stevens, J. C., & Marks, L. E. Thermally induced pain, the Dol scale, and the psychophysical power law. *American Journal of Psychology*, 1968, *81*, 147–164.

Adams, C., Hall, D., Pennypacker, H., Goldstein, M., Hench, L., Madden, M., Stein, G., & Catania, A. Lump detection in simulated human breasts. *Perception & Psychophysics*, 1976, *20*, 163–167.

American Psychiatric Association. *Diagnostic and statistical manual, mental disorders* (1st ed.). Washington, D.C.: Author, 1952.

American Psychiatric Association. *Diagnostic and statistical manual of mental disorders* (2nd ed.). Washington, D.C.: Author, 1968.

Anderson, N. H. Functional measurement and psychophysical judgment. *Psychological Review*, 1970, *77*, 153–170.

Banks, W. P. Signal detection theory and human memory. *Psychological Bulletin*, 1970, *74*, 81–99.

Birnbaum, M. H. Using contextual effects to derive psychophysical scales. *Perception & Psychophysics*, 1974, *15*, 89–96.

Brooks, D. N. Recognition memory after head injury: A signal detection analysis. *Cortex*, 1974, *40*, 224–230.

Campbell, R. A. Thresholds and (or?) signal detection theory. *Journal of Speech and Hearing Research*, 1965, *7*, 97–98.

Campbell, R. A., & Moulin, L. K. Signal detection audiometry: An exploratory study. *Journal of Speech and Hearing Research*, 1968, *11*, 402–410.

Chapman, C. R. Psychophysical evaluation of acupunctural analgesia: Some issues and considerations. *Anesthesiology*, 1975, *43*, 501–505.

Chapman, C. R. Sensory decision theory methods in pain research: A reply to Rollman. *Pain*, 1977, *3*, 187–211.

Chapman, C. R., & Feather, B. W. Sensitivity to phobic imagery: A sensory decision theory analysis. *Behaviour Research and Therapy*, 1971, *9*, 161–168.

Chapman, C. R., & Feather, B. W. Effects of diazepam on human pain tolerance and pain sensitivity. *Psychosomatic Medicine*, 1973, *35*, 330–340.

Chapman, C. R., Gehrig, J. D., & Wilson, M. E. Acupuncture compared with 33% nitrous oxide for dental analgesia: A sensory decision theory evaluation. *Anesthesiology*, 1975, *42*, 532–537. (a)

Chapman, C. R., Gehrigh, J. D., & Wilson, M. E. Acupuncture, pain, and signal detection theory. *Science*, 1975, *189*, 65. (b)

Chapman, J., & McGhie, A. A comparative study of disordered attention in schizophrenia. *Journal of Mental Science*, 1962, *108*, 487–500.

Chapman, C. R., Murphy, T. M., & Butler, S. H. Analgesic strength of 33 percent nitrous oxide: A signal detection theory evaluation. *Science*, 1973, *179*, 1246–1248.

Chapman, C. R., Wilson, M., & Gehrig, J. Comparative effects of acupuncture and transcutaneous stimulation on the perception of painful dental stimuli. *Pain*, 1976, *2*, 265–283.

Clark, W. C. Sensory decision theory analysis of the placebo effect on the criterion for pain and thermal sensitivity (d'). *Journal ofAbnormal Psychology*, 1969, *74*, 363–371.

Clark, W. C. Pain sensitivity and the report of pain: An introduction to sensory decision theory. *Anesthesiology*, 1974, *40*, 272–287.

Clark, W. C., & Goodman, J. S. Effects of suggestion on d' and C_e for pain detection and pain tolerance. *Journal of Abnormal Psychology*, 1974, *83*, 364–372.

Clark, W. C., & Mehl, L. Thermal pain: A sensory decision theory analysis of the effect of age and sex on d', various response criteria, and 50% pain threshold. *Journal of Abnormal Psychology*, 1971, *78*, 202–212.

Clark, W. C., & Yang, J. C. Acupunctural analgesia? Evaluation by signal detection theory. *Science*, 1974, *184*, 1096–1098.

Clark, W. C., Yang, J. C., & Hall, W. Acupuncture, pain, and signal detection theory. *Science*, 1975, *189*, 66–68.

Craig, K. D., Best, H., & Ward, L. M. Social modeling influences on psychophysical judgments of electrical stimulation. *Journal of Abnormal Psychology*, 1975, *84*, 366–373.

Craig, K. D., & Coren, S. Signal detection analyses of social modeling influences on pain expressions. *Journal of Psychosomatic Research*, 1975, *19*, 105–112.

Cross, D., Tursky, B., & Lodge, M. The role of regression and range effects in determination of the power function for electric shock. *Perception & Psychophysics*, 1975, *18*, 9–14.

Eberle, T., Rehm, L., & McBurney, D. Fear decrement to anxiety hierarchy items—Effects of stimulus intensity. *Behaviour Research and Therapy*, 1975, *13*, 255–262.

Ekman, G., Frankenhaeuser, M., Goldberg, L., Bjerver, K., Järpe, G., & Myrsten, A. Effects of alcohol intake on subjective and objective variables over a five-hour period. *Psychopharmacologia*, 1963, *4*, 28–38.

Ekman, G., Frankenhaeuser, M., Goldberg, L., Hagdahl, R., & Myrsten, A. Subjective and objective effects of alcohol as functions of dosage and time. *Psychopharmacologia*, 1964, *6*, 399–409.

Ekman, G., Frankenhaeuser, M., Levander, S., & Mellis, I. Scales of unpleasantness of electrical stimulation. *Scandinavian Journal of Psychology,* 1964, *5,* 257–261.

Ekman, G., Frankenhaeuser, M., Levander, S., & Mellis, I. The influence of intensity and duration of electrical stimulation on subjective variables. *Scandinavian Journal of Psychology,* 1966, *7,* 58–64.

Feather, B., Chapman, C. R., & Fisher, S. The effect of a placebo on the perception of painful radiant heat stimuli. *Psychosomatic Medicine,* 1972, *34,* 290–294.

Frankenhaeuser, M., Fröberg, J., Hagdahl, R., Rissler, A., Björkvall, C., & Wolff, B. Physiological, behavioral, and subjective indices of habituation to psychological stress. *Physiology and Behavior,* 1967, *2,* 229–237.

Frankenhaeuser, M., & Järpe, G. Subjective intoxication induced by nitrous oxide in various concentrations. *Scandinavian Journal of Psychology,* 1962, *3,* 171–176.

Frankenhaeuser, M., Mellis, I., Rissler, A., Björkvall, C., & Patkai, P. Catecholamine excretion as related to cognitive and emotional reaction patterns. *Psychosomatic Medicine,* 1968, *30,* 109–119.

Frankenhaeuser, M., Sterky, K., & Järpe, G. Psychophysiological relations in habituation to gravitational stress. *Perceptual and Motor Skills,* 1962, *15,* 63–72.

Franzini, L. R. Magnitude estimations of the averseness of the interval preceding shock. *Journal of Experimental Psychology,* 1970, *84,* 526–528.

Goldman, B. The commercially unpublished test reported in published research. *American Psychologist,* 1977, *32,* 386–387.

Gordon, S. K., & Clark, W. C. Application of signal detection theory to prose recall and recognition in elderly and young adults. *Journal of Gerontology,* 1974, *29,* 64–72.

Hayes, R., Bennett, G., & Mayer, D. Acupuncture, pain, and signal detection theory. *Science,* 1975, *189,* 65–66.

Hilgard, E. R. A quantitative study of pain and its reduction through hypnotic suggestion. *Proceedings of the National Academy of Sciences,* 1967, *57,* 1581–1586.

Hilgard, E. R. Pain as a puzzle for psychology and physiology. *American Psychologist,* 1969, *24,* 103–113.

Hunt, W. A., & Jones, N. F. The experimental investigation of clinical judgment. In A. J. Bachrach (Ed.), *Experimental foundations of clinical psychology.* New York: Basic Books, 1962.

Korboot, P. J., & Damiani, N. Auditory processing speed and signal detection in schizophrenia. *Journal of Abnormal Psychology,* 1976, *85,* 287–295.

Lang, P. J. Fear reduction and fear behavior: Problems in treating a construct. In J. M. Shlien (Ed.), *Research in psychotherapy* (Vol. 3). Washington, D.C.: American Psychological Association, 1968.

Latané, B., & Harkins, S. Cross-modality matches suggest anticipated stage fright and multiplicative power function of audience size and status. *Perception & Psychophysics,* 1976, *20,* 482–488.

Lloyd, M. A., & Appel, J. B. Signal detection theory and the psychophysics of pain: An introduction and review. *Psychosomatic Medicine,* 1976, *38,* 79–93.

Luce, R. D. What sort of measurement is psychophysical measurement? *American Psychologist,* 1972, *27,* 96–106.

Lundberg, V., & Ekman, G. Emotional involvement while anticipating an examination: A psychophysical study. *Perceptual and Motor Skills,* 1970, *31,* 603–609.

Lusted, L. B. *Introduction to medical decision making.* Springfield, Ill.: Charles C Thomas, 1968.

Lusted, L. B. Decision-making studies in patient management. *New England Journal of Medicine,* 1971, *284,* 416–424. (a)

Lusted, L. B. Signal detectability and medical decision-making. *Science,* 1971, *171,* 1217–1219. (b)

Lusted, L. B. Receiver operating characteristic analysis and its significance in interpretation of radiologic images. In E. J. Potchen (Ed.), *Current concepts in radiology* (Vol. 2). St. Louis, Mo.: Mosby, 1975.

Marks, L. E. On scales of sensation: Prolegomena to any future psychophysics that will be able to come forth as science. *Perception & Psychophysics,* 1974, *16,* 358–376.

McBurney, D. Acupuncture, pain, and signal detection theory. *Science,* 1975, *189,* 66.

McNicol, D. *A primer of signal detection theory.* London: Allen & Unwin, 1972.

Metz, C., Starr, S., & Lusted, L. B. Quantitative evaluation of visual detection performance in medicine: ROC analysis and determination of diagnostic benefit. In G. A. Hay (Ed.), *Medical images: Formation, perception and measurement.* London: Wiley, 1977.

Miller, E., & Lewis, P. Recognition memory in elderly patients with depression and dementia: A signal detection analysis. *Journal of Abnormal Psychology,* 1977, *86,* 84–86.

Miller, R. J., & Leibowitz, H. A signal detection analysis of hypnotically induced narrowing of the peripheral visual field. *Journal of Abnormal Psychology,* 1976, *85,* 446–454.

Moskowitz, H., & McGlothlin, W. Effects of marijuana on auditory signal detection. *Psychopharmacologia,* 1974, *40,* 137–145.

Neufeld, R. W. J. Effect of cognitive appraisal on *d'* and response bias to experimental stress. *Journal of Personality and Social Psychology,* 1975, *31,* 735–743.

Pastore, R. E., & Scheirer, C. J. Signal detection theory: Considerations for general application. *Psychological Bulletin,* 1974, *81,* 945–958.

Patkai, P., Frankenhaeuser, M., Rissler, A., & Björkvall, C. Catecholamine excretion, performance and subjective stress. *Scandinavian Journal of Psychology,* 1967, *8,* 113–122.

Peck, R. E. The application of thymometry to the

measurement of anxiety. *International Journal of Neuropsychiatry,* 1966, *2,* 337–341.

Poulton, E. C. The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin,* 1968, *69,* 1–19.

Price, R. H. Signal-detection methods in personality and perception. *Psychological Bulletin,* 1966, *66,* 55–62.

Rappaport, M., Silverman, J., Hopkins, H., & Hall, K. Phenothiazine effects on auditory signal detection in paranoid and nonparanoid schizophrenics. *Science,* 1971, *174,* 723–725.

Rollman, G. Signal detection theory measurement of pain: A review and critique. *Pain,* 1977, *3,* 187–211.

Squire, L., Slater, P., & Chace, P. Anterograde amnesia following electroconvulsive therapy: No evidence for state-dependent learning. *Behavioral Biology,* 1976, *18,* 335–343.

Starr, S., Metz, C., Lusted, L., & Goodenough, D. Visual detection and localization of radiographic images. *Radiology,* 1975, *116,* 533–538.

Stenson, H., Kleinmuntz, B., & Scott, B. Personality assessment as a signal detection task. *Journal of Consulting and Clinical Psychology,* 1975, *43,* 794–799.

Stevens, S. S. On the psychophysical law. *Psychological Review,* 1957, *64,* 153–181.

Stevens, S. S. The psychophysics of sensory function. *American Scientist,* 1960, *48,* 226–253.

Stevens, S. S. A metric for the social consensus. *Science,* 1966, *151,* 530–541.

Stevens, S. S. Neural events and the psychophysical law. *Science,* 1970, *170,* 1043–1050.

Stevens, S. S. Issues in psychophysical measurement. *Psychological Review,* 1971, *78,* 426–450.

Stevens, S. S. *Psychophysics.* New York: Wiley, 1975.

Stevens, S. S., Carton, A. S., & Shickman, G. M. A scale of apparent intensity of electric shock. *Journal of Experimental Psychology,* 1958, *56,* 328–334.

Stone, L. A. Clinical psychophysics. *Studia Psychologica,* 1968, *10,* 161–173. (a)

Stone, L. A. Bases for psychiatric impairment severity judgments: Psychophysical power functions? *Studia Psychologica,* 1968, *10,* 194–199. (b)

Stone, L. A. Bases for psychiatric prognostic favorability judgments: Psychophysical power functions? *Behavioral Science,* 1969, *14,* 133–137. (a)

Stone, L. A. Psychiatrists' judgmental evaluations of susceptibility to external stress for selected disorder classification stimuli. *Journal of Clinical Psychology,* 1969, *25,* 21–26. (b)

Stone, L. A. Magnitude estimation and numerical category scale evaluations of category scale adjectival stimuli on three clinical judgmental continua. *Journal of Clinical Psychology,* 1970, *26,* 24–27. (a)

Stone, L. A. A law of clinical judgment: A psychological mechanism based on the logic of psychophysics. *Journal of Clinical Psychology,* 1970, *26,* 312–317. (b)

Stone, L. A., & Linscheid, T. R. Another law of clinical judgment: A psychological mechanism based on the logic of psycho-dynamics. *Psychological Reports,* 1971, *28,* 851–855.

Stone, L. A., & Skurdal, M. A. Judged prognosis for functional psychosis disorder classifications: A prothetic continuum. *Journal of Consulting and Clinical Psychology,* 1968, *32,* 469–472.

Sullivan, R. Subjective matching of anxiety to intensities of white noise. *Journal of Abnormal Psychology,* 1969, *74,* 646–650.

Sullivan, R. Magnitude estimation of anxiety. *Psychonomic Science,* 1970, *21,* 209–211.

Sullivan, R. Magnitude estimation and relative aversiveness of anxiety: Phobia. *Journal of Abnormal Psychology,* 1971, *78,* 266–271.

Sullivan, R. Anxiety: A method for scaling its relative magnitude and aversiveness. *Journal of Abnormal Psychology,* 1973, *82,* 483–490.

Tryon, W. W. Psychophysical scaling and hierarchy construction. *Journal of Behavior Therapy and Experimental Psychiatry,* 1977, *8,* 53–56.

Wolpe, J. Quantitative relationships in the systematic desensitization of phobias. *American Journal of Psychiatry,* 1963, *119,* 1062–1068.

Yates, A. J. Data-processing levels and thought disorder in schizophrenia. *Australian Journal of Psychology,* 1966, *18,* 103–117.

Received July 29, 1977 ■