Current Issues in Theoretical Psychology
Wm J. Baker, M.E. Hyland, H. Van Rappard, A.W. Staats (Editors)

"THERE IS NOTHING MORE PRACTICAL THAN A GOOD THEORY" (KURT LEWIN) - TRUE OR FALSE?

H. J. Eysenck

Institute of Psychiatry, University of London
London, England

SUMMARY: An attempt is made to consider what constitutes a 'good theory' in the light of recent discussions by philosophers of science, and in the light of the writer's suggestion that there is a historical development from weak to strong theories, and that the requirements may differ depending on the stage reached. The suggested criteria for a good theory are discussed in relation to various experimental paradigms, and it is concluded that 'good' theories are distinguished by being part of a progressive and advancing research programme, while 'bad' theories are associated with a regressive research program.

Lewin's statement about the usefulness of a good theory has become famous in psychology, and few would probably disagree with him. However, his statement is more honoured in the breach than the observance; there is little evidence in their actual work that most psychologists pay much attention to theory, and some explicitly disassociate themselves from the search for theory and adopt a Baconian or pre-Baconian, purely inductive approach. Indeed, the lack of a proper *paradigm* in the social sciences generally, and in psychology in particular, noted by Kuhn (1962) and Barnes (1982), can be identified with a lack of accepted theories in these fields. Disagreement with the request and need for theories is not often voiced explicitly, or defended philosophically; it forms one of the unverbalized axioms underlying much of psychological thinking.

Even where the need for a good theory is not only explicitly acknowledged, but also forms part of a given psychologist's research philosophy, there still remains the question of when a theory is 'good,' and when it is not. For many psychologists, to give but one example, Freudian theories are regarded as 'good' because they are believed to be full of 'insights', and to touch on vital questions of motivation, mental disorder, psychotherapy, memory, and other human concerns. To others, like Popper (1959, 1974a, 1974b), Eysenck (1985a), Grunbaum (1984), Rachman (1963), Zwang (1985), and many others psychoanalysis appears to be the prime example of a bad theory. Clearly we need criteria as to what constitutes a 'good' theory, and this in turn would seem to demand some kind of answer to the demarkation dispute about science and pseudo-science.

There has been much philosophical argument on this topic; Achinstein (1965), Bergmann (1951, 1954, 1957), Bergmann and Spence (1941), Carnap (1966), Feyerabend (1975), Hanson (1958), Hempel (1952, 1965, 1966), Krige (1980), Lakatos (1968), Lakatos and Musgrave (1970), Putnam (1962), Quine (1962), Suppe (1974), Tarski (1941, 1956), Toulmin (1953), and many others have attacked the problem without reaching agreement. Eysenck (1985) has given a detailed survey of the present position; the only point of agreement among modern philosophers appears to be that the theory of 'logical positivism' associated with the Viennese school, usually the only theory known to and considered by psychologists, is definitely out of favour, and regarded as obsolete.

Perhaps the most relevant solution to the problem, and the one most closely related to this specific question of what constitutes a 'good' theory, may be related to a suggestion made by Lakatos (1968; Lakatos & Musgrave, 1970). His view, which is an advance on Popper's well known 'falsification' criterion, is widely accepted amongst philosophers of science. It aligns a 'good' theory with an advancing and ***progressive*** research programme, while a 'bad' theory is associated with a ***regressive*** research programme, i.e., one which, instead of predicting and discovering new facts, is concerned with explaining away failures and anomalies. On this basis, clearly, psychoanalysis is a 'bad' theory, because it has failed to predict and discover new facts, and has rather been forced to explain away failures and anomalies. Examples are the failure of psychoanalysts to cure neurotic and psychotic illnesses more decisively than do other methods of psychotherapy and behaviour therapy (Rachman and Wilson, 1981; Smith, Glass and Miller, 1980; Strupp, Hadley, and Gomes-Schwartz, 1977); the fact that behaviour therapy successfully cures the 'symptoms' of neuroses, but does not lead to relapse or symptom substitution (Kazdin & Wilson, 1978; Schorr, 1984); and the fact that the experimental study of Freudian theories has been almost entirely negative (Eysenck & Wilson, 1973; Kline, 1981; Eysenck, 1985a). In spite of the strong evidence that psychoanalytic theory is in fact a 'bad' theory, which could be documented at much greater length were it considered necessary, it is still widely accepted by many clinical psychologists, an obvious proof of our assertion that theories may be accepted or rejected for reasons other than their 'goodness' or 'badness.'

If we are willing to accept Lakatos' proposal of aligning a 'good' theory with an advancing and progressive research programme, we are still left with certain problems. One of these is the distinction between 'strong' and 'weak' theories, first discussed in detail by Eysenck (1960), and later made a central part of his discussion of the place of theory in a world of facts (Eysenck, 1985b). It is characteristic of weak, as opposed to strong theories that "only few observations, and these of doubtful accuracy, are available. Few quantitative or even qualitative laws, universally established, are available in sub-fields. The nature of the phenomena in question is by no means clear-cut or well understood. Mathematical relations are often very complex, and predictions are neither straight-forward nor precise" (Eysenck, 1960, p. 304). This type of theory is clearly differentiated from a strong theory, like Newton's theory of gravitation, or Einstein's theory of relativity.

Eysenck (1976) has suggested that there is a historical development from weak to strong theories, related to the changing criteria for scientific acceptability. Figure 1 shows this relationship. At an early stage of development, we have hunches based on observation and induction. Slowly we graduate to hypothesis formation and the verification criterion of the Vienna school. Hypotheses develop into wide-ranging and more specific theories, and Popper's falsification criterion is now apposite. Finally, we reach a stage where general laws become possible, and these may only be challenged by the statement of alternative theories leading to Kuhnian revolutions. Clearly psychology is still at an early stage, where induction and verification are primarily relevant, and where weak rather than strong theories are the rule. This imposes important restraints on our experimentation, theory development, and theory testing. What, it may be said, is the use of weak theories? The answer, I believe, can be found in the words of the famous physicist, J. J. Thomson: "A theory in science is a policy rather than a creed." This statement indicates the heuristic nature of theories, particularly weak theories; the value of such a theory lies in the fact that it directs attention to those problems which repay study from a systematic point of view; in Thomson's words, it defines a policy of action and research. It is by giving rise to worthwhile research, rather than by necessarily being right, that a weak theory makes its greatest contribution to science.

A good example of the usefulness of weak theories in pointing in the right direction is the work of John Dalton (Greenaway, 1966) on the atom. All that Dalton said about atoms, apart from the effect of their existence, which was not novel, was wrong. They are not indivisible nor of unique weight, as he thought; they need not obey the laws of definite

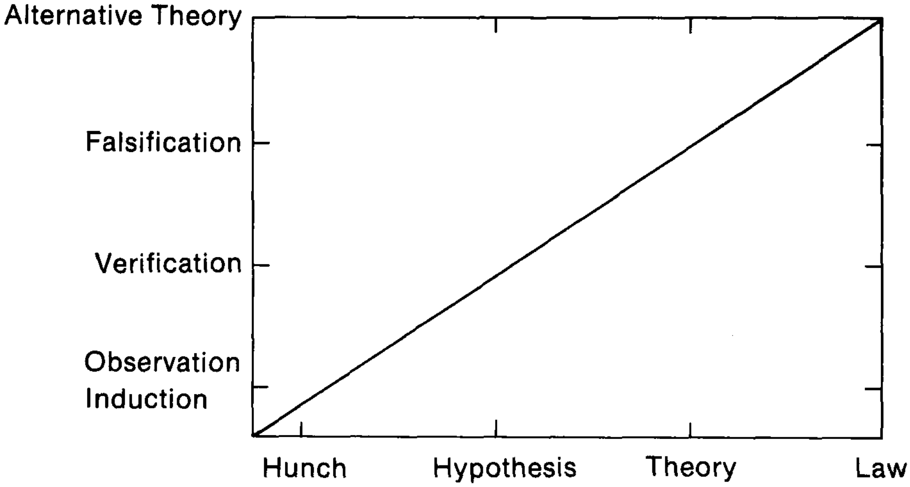

Fig. 1. Stages in the development of scientific theories from weak to strong.

or multiple proportions, as he believed; and in any case his values for relative atomic weights and molecular constitutions were for the most part incorrect! Yet, for all that, John Dalton, more than any other single individual, was the man who set modern chemistry on its feet. His theory is a good example of Lakatos' "advancing and progressive research programme;" it is a weak theory, because very little was in fact known about atoms at the time, but it led to the elaboration of experiments and theories which in fact created modern chemistry, and they were also fundamental for much in modern physics.

The distinction between strong and weak theories is an important one, because psychologists are often impatient with the failure of theories to predict accurately and to explain exhaustively all the known facts. This is a fundamental failure to understand the nature of science. All theories, even the stronger theories in the most advanced physical sciences, are full of anomalies when tested experimentally. Newton's theory of gravitation appeared full of exceptions and errors, and for 300 years scientists were busy trying to iron out these anomalies, attempting to explain the observed deviations from theory, and generally carrying out what Kuhn calls "the ordinary business of science," i.e., the solving of puzzles created by discrepancies between theory and fact. Weak theories, of course, suffer even more from these difficulties, and even more patience is required. Psychologists often reject theories impatiently when they fail to give 100% accurate predictions and explanation; this is unreasonable. Had physicists and chemists proceeded with the same abandon, it is safe to say that no physical or chemical theories would have been made in the hard sciences. We must take seriously the proposition that "a theory is a tender growth, naturally imperfect when first proposed, and only likely to become acceptable after many years and possibly centuries of experimental work to explicate its complexity, and theoretical efforts to improve it and make it more precise and accurate. It must be given time to do these things, otherwise it will die of neglect" (Roley, 1959).

From the point of view of metatheory, we have drawn attention to two features of scientific theories. In the first place, they are 'good' when they lead to an advancing and progressive research programme, and in the second place they are likely to be 'weak' in the early stages of a science, and to grow 'strong' only later, through a great deal of experimental research and theory development. It may be possible to make a little more precise the distinction between weak and strong theories. In testing a given theory (H), many assumptions (K) must be made in order to make the testing of the theory possible. Failure of the experiment may indicate not that H was erroneous, but that some of the assumptions under K were incorrect. In weak theories the part played by K, and general ignorance about K, is much larger than in strong theories; hence the testing of H is very much more difficult, and H should not readily be abandoned simply because H + K can be shown to be falsified by actual experiment (Cohen & Nagel; 1936). Let us put the logic of this argument in the following form: if H and K, then P. If our experiment shows P to be false, then either H is false or K (in part or complete) is false. If we have good grounds for believing K is not false, H is refuted by the experiment. Nevertheless, the experiment really tests both H *and* K. If, in the interest of the coherence of our knowledge, it is found necessary to revise the assumptions contained in K, the crucial experiment must be reinterpreted, and need not then decide against H. The difference between weak and strong theories lies in the reliance we may place on K; far more is known about K in the case of strong than in the case of weak theories.

This discussion leads quite logically to a consideration of the importance of parameter values. Consider Eysenck's (1967) prediction that introverts would show eye-blink conditioning more rapidly and more strongly than extraverts. This was indeed found to be so when weak intensity UCSs were used, but the reverse was found when more intense UCSs were used (Eysenck, 1981). In this case it was not H that was responsible, but K; the intensity parameter is clearly implicated, and proper predictions can be mediated by the use of Pavlov's law of transmarginal inhibition, which states that the relation between intensity of stimulus and strength of conditioning is not linear, but curvilinear. Many other examples of a similar kind are given in Eysenck (1981). Ideally, the general theory should always state the parameter values under which it can operate, but in weak theories this inevitably forms part of a research programme and is not normally included in the statement of the original weak theory. It is only by discovering the requisite parameter values that the weak theory can become a strong one.

It might be said that to explain a 'good' theory in terms of it being linked with a progressive research programme is tautological, unless a 'progressive research programme' can be defined more precisely. The literature on the philosophy and history of science suggests that there are five criteria which may be used in order to formulate such a judgement, and I will try to state these at the same time as giving examples taken from psychological research. The first criterion, and one which is universally admitted, is the power of the theory to *predict new facts*. It is, of course, desirable, and it may be essential, that these facts should not be predictable in terms of any other theory, and if the facts go counter to common sense expectation, so much the better. Such predictions are particularly impressive when the new facts contradict existing scientific and/or common sense views. As an example, consider enuresis. Here, as Morgan (1978) points out, there are two behaviouristic theories, one a classical conditioning paradigm, the other an avoidance learning paradigm. Both make similar predictions, however, and may therefore be considered together. In either case, the theory declares enuresis to be the result of a failure to learn the connection between the US (the enlargement of the bladder) and the response (waking up and urinating in the toilet). The bell-and-blanket method provides the missing link, the ringing of the bell being the CS needed to provide the required conditioning. This theory contrasts with the psychoanalytic theory, which regards enuresis as a consequence of anxiety and other emotional states of the individual, and declares that symptomatic treatment is useless because it will only exacerbate the underlying emotional conflicts.

Predictions made from the learning theory model are as follows: (1) enuresis should be abolished in more subjects (other than those suffering from certain physical disabilities, infections, etc.) by means of this technique than any other. (2) There should be a fair degree of relapse, because conditioning stops the moment the criterion is reached. (3) The *intensity* of the US should be related to the success of the treatment and the lack of relapse. (4) Intermittent reinforcement should prove superior to 100% reinforcement in preventing relapse. (5) Over-learning should be effective in preventing relapse. These are quite specific predictions, often counter-intuitive, but all following directly from well-known learning theory principles. As Morgan has demonstrated, the evidence is very much in line with these predictions. In addition it appears that a sixth prediction is also borne by the facts, namely (6) that the patients' anxieties and other mental symptoms would markedly decrease once enuresis had been eliminated. All this goes completely counter to the Freudian position.

This evidence suggests that learning theory is a *progressive,* and hence a 'good' theory, whereas psychoanalysis which has to rely on trying to argue away these very critical findings in some way, is *regressive,* and hence a 'bad' theory. A much more detailed discussion of behaviour therapy and learning theory, on the one hand, and psychoanalysis on the other, is given elsewhere (Eysenck, (1985a).

The second virtue of a good theory is its ability to *explain* what previously appeared to be anomalies. As an example, consider the anomalies which have arisen since Urbantschitsch (1883) put forward the hypothesis that the perception of visual, auditory, tactile, pressure, pain, olfactory, and gustatory stimuli can be facilitated by simultaneous heteromodal sensory stimulation. The hypothesis essentially states that the arousal produced by more intense heteromodal stimulation lowers sensory thresholds for other types of stimulation. The experimental evidence in many studies has been contradictory, some of it being positive, some of it negative, and some neutral (Eysenck, 1976). Shigehisa and Symons (1973; see also Shigehisa, Shigehisa, & Symons, 1973) put forward the theory that (a) the regression should be curvilinear, following Pavlov's law of transmarginal inhibition, and (b) that in line with Eysenck's (1967) theory relating extraverion-introversion to cortical arousal, the transition from decreasing to increasing thresholds should take place earlier in introverts than in extraverts. In a whole series of studies they managed to demonstrate that the regression was indeed curvilinear, and related to personality in the manner posited. In this way it was possible for Eysenck's personality theory to clear up anomalies in the experimental literature which were completely inexplicable before. This is an important function of a 'good' theory.

The third sign of a good theory is that it may act as a *criterion* between different interpretations which, in the absence of a good theory, could not be empirically separated. As an example, consider the long-lasting controversy concerning Spearman's general factor of intelligence, *g*. There have been many attempts to solve the problem of its existence by means of factor analytic studies, but it became obvious that the way in which variances were distributed by means of the rotations implicit in factor analysis could not be determined objectively. Thurstone's 'simple structure' seemed at first to present a useful criterion, but it soon became obvious that the two criteria suggested, namely 'simple structure' and 'orthogonality of axes,' were opposed to each other. Given the typical factor analysis of a matrix of intercorrelations between a large number of intelligence tests, it was possible *either* to reach simple structure and abandon orthogonality, or to preserve orthogonality, and abandon simple structure. Thurstone (1967) abandoned simple structure. Clearly there is no statistical answer to this problem, and both interpretations are still very much alive in the literature.

An answer to this problem has been suggested by Eysenck (1982), who advocated taking seriously Galton's original theory that intelligence had a biological foundation, and that measures of intelligence should directly address this physiological basis. Using the event-related potential on the EEG as a measure of this underlying physiological basis of individual differences in intelligence (Eysenck & Barrett, 1985), it was found that a very high correlation of .83 existed in a group of 219 15-year old school children between the AEP (Average Evoked Potential) and the Wechsler IQ. This clearly suggested that the advocates of *g* were right and those denying it were wrong. This is highly indicative, but not conclusive. A more conclusive deduction can be made possible by using Eysenck's method of criterion analysis (Eysenck, 1950).

The particular form this method takes in this connection might also be called the ***proportionality criterion***. If we take the 11 sub-tests of the WISC on this sample, intercorrelate them and extract the first or general factor, the factor loadings indicate to what extent each test measures that which is common to them all, i.e., *g* according to our theory. Again, if we correlate the AEP score with each test separately, then, if we assume that the AEP is a direct measure of *g*, then each such correlation indicates the extent to which each sub-test of the WISC correlates with *g*. If both these hypotheses are correct, then clearly the two sets of values should be ***proportional***, and consequently should show a very high correlation. In actual fact the correlation was found to be .95, which is not significantly different from unity. This result is impossible to interpret along the lines of a theory which denies the existence of *g*, and hence the theory helps us to decide between two interpretations of a large body of empirical data, namely that furnished by factor analytic investigations carried out over the past 75 years.

The theoretical considerations underlying this argument can be applied in many ways, and are fundamental in deciding between many interpretations of empirical data. Another example in which it has been used is the application of criterion analysis to the theory that there exists a continuum between normality and psychiatric abnormality, i.e., neurosis and psychosis, and that these two continua are essentially independent. The evidence indicates that the theory is essentially correct, and thus enables us to to make firm affirmations concerning large numbers of data and observations which previously had not been looked at from this particular theoretical point of view.

The fourth advantage of a 'good' theory is its ability to unify apparently separate disciplines. Experimental psychology and correlational psychology have in the past pursued rather divergent courses, and have had little to do with each other. Cronbach (1957), in his famous presidential address to the APA talked about the "two disciplines of scientific psychology," and argued that we would never succeed in establishing psychology as a basic science if we did not bring these two disciplines together. Eysenck (1967) has argued very much along similar lines. The question, of course, is how this can be done, and it may be suggested that a personality theory which explains individual differences, and the dimensions created by these individual differences, along the lines of concepts worked out by experimental psychologists, may be the answer to this question (Eysenck & Eysenck, 1985).

Such a theory, e.g., that of explaining differences in extraversion-introversion in terms of differences in cortical arousal, mediated by the ascending reticular formation, extends even beyond the limits of psychology proper. As an example, consider psychopharmocology, where Eysenck (1983a, b) has suggested that a classification of drugs which alter human behaviour can best be made in terms of the major dimensions of personality (psychoticism, extraversion, neuroticism). A great deal of evidence is cited in these studies to demonstrate the viability of this suggestion.

Eysenck (1983c, 1984) has suggested that by linking the two disciplines of scientific psychology, it is possible to create a paradigm in personality research, and that a theory which is thus capable of uniting a great number of divergent approaches has certain advantages which set it apart from competing theories. The ability to generate a paradigm of the Kuhnian type would certainly be an index of a good theory, and this ability to unify apparently separate disciplines characterizes a far-reaching paradigm.

In direct line with Lewin's statement is a fifth criterion, namely, the *practical application* of a given theory. Scientists who stress the differences between pure and applied science are often hostile to statements of this kind, and prefer the 'purity' of research unemcumbered by application. In philosophical discussion it is possible to argue the two sides to this question, particularly with reference to the physical sciences, but in psychology I would submit that such a debate would be quite irrelevant.

The reason for this lies in the particular character of psychology as a science. The hard sciences (with certain exceptions, like astronomy) rely on experiment rather than on observation, and the main characteristic of such experiment is the limitation and exclusion of variables which affect the outcome of the experiment. What is aimed at is a simple functional statement of the kind: $a = f(b)$, i.e., a is a function of b. Usually a mathematical statement of the type of function is looked for. Novel experimental techniques and analyses may be needed before it becomes possible to investigate directly any such functional relationships, but this analytic work is the essence of experimental physics.

The difficulty in psychology, of course, is that we are dealing with ***behaviour***, which implies an intact ***organism***. In the organism, functional relationships of the kind mentioned above are always complicated by the fact that other variables which also determine the occurrence of a cannot be excluded, so that the formula has to read: $a = f(b, o)$, where o denotes the organism, i.e., a multiplicity of influences which act upon the dependent variable, and make it impossible to study the independent variable in isolation.

It is one of the major errors of experimental psychology when it assumes that by isolating an individual in the laboratory, one can exclude unwanted influences residing in the organism. This is clearly impossible. It is well known that differences in intelligence, differences in attitude, differences in degrees of anxiety, and other such variables inevitably determine a person's reaction to experimental manipulations, and may be even more powerful than the independent variables selected by the experimenter. This is now so widely agreed that there is little need to substantiate these statements.

Although the experimental situation enables us to exclude or minimize certain independent variables, it also imposes upon us certain restrictions which are related to moral and ethical problems. If we wish, for instance, to study strong emotions as the independent variables in our design, we are unable to do so in the laboratory because we cannot arrange matters in such a way that our experimental subjects experience devastating anxieties, suicidal depressions, etc. That means that these variables can only be studied *outside* the laboratory, and again ethical considerations make it necessary that they should be so studied in the context of a therapeutic endeavour. This immediately means that we cannot study strong emotions other than in a practical setting, which alone enables us to test theories of the origin, preservation, and extinction of such emotions through some kind of therapeutic endeavours (Eysenck, 1985b). We can, of course, to some extent address the issue by working with animals, but convincing evidence regarding the place of emotion in human conduct cannot easily be supplied by *only* working with animals. Animal experiments have been important in suggesting to us theories and hypotheses in conditioning and learning which have been extremely valuable in creating modern behaviour therapy, but such theories always require application to human conduct before they can be accepted and that means inevitably that we must apply them in a therapeutic setting.

Much the same might be said about intelligence testing. As Jensen (1984) has pointed out, divergent notions about the generality or specificity of intelligence can with advantage be tested in the applied field. As he says: "If the specificity notions of test validity that prevail through federal enforcement agencies lead to unnecessary and costly validity studies, or to greatly relaxed selection standards, or to quotas, or to abandoning the use of tests altogether in an organization, it would seem important to estimate the actual monetary consequences of these alternatives. If tests are abandoned or replaced by less valid selection techniques, such as interviews, there is bound to be a decline in the overall quality of those hired, and lower-performing personnel mean a loss in productivity" (p. 107). According to *Time*, the rate of productivity growth in the U. S. A. has declined, since 1970, from 3.5% to 1%, and Schmidt (1979) has suggested that this decline is due in part to a reduced efficiency in allocating people to jobs because of governmental obstacles put in the way of using tests and optimal selection model.

Some figures may be of interest. It has been estimated that the total annual savings in training costs to the armed forces as a result of test selection classification of enlisted personnel was 442 million dollars (Jensen, 1984). Hunter, Schmidt, and Rauschenberg (1984) have improved on earlier statistical methods for cost estimates of the consequence of different selection strategies for various jobs, and have shown that employee differences in job proficiency correspond to considerable differences in the actual dollar value of their performance. In a study of budget analysts, Schmidt and Hunter (1980) have estimated, for instance, that the dollar value productivity of superior performers was 23,000 dollars per year *greater* than that of low performers; this value should, of course, be multiplied by the number of persons employed. In a study of the Philadelphia police department, with 5,000 employees, Hunter (1979) has estimated that abandonment of general ability tests for the selection of police officers would cost a total of 180 million dollars over a ten-year period. The estimated gain in productivity resulting from one year's use of a more valid selection procedure for computer programmers in the federal government ranges from 5.6 to 92.2 million dollars for different sets of estimation parameters, and for the whole federal government, with 4 million employees, Hunter and Schmidt (1980) conservatively estimated optimal selection procedures would save 16 billion dollars per year. Hunter and Schmidt (1982) have also estimated the cost effectiveness of using tests for job selection on a national scale. The differences between not using and using tests for the allocation of the work force to jobs they calculated at about 169 billion dollars! If general ability tests, to the extent that they are currently used in selection, were to be abandoned, the estimated loss in national productivity would be about 80 billion dollars per year. Many other similar figures are available, but these are reasonably sufficient to indicate the practical usefulness of intelligence tests as selection devices in actual practice.

In conclusion, let me state certain points which seem to arise from the discussion. (1) Theories are an essential part of science, including psychology. (2) It is possible to distinguish 'good' from 'bad' theories. (3) 'Good' theories are distinguished by being part of a progressive and advancing research programme, while 'bad' theories are associated with a regressive research programme. (4) There are five major criteria by which to judge a given theory, and to decide whether it is part of an advancing and progressive, or a regressive research programme. Psychologists should take theories and their assessments more seriously than they do at present, and might, with advantage, apply the standards and criteria suggested.

References

Achinstein, P. (1965). The problem of theoretical terms. *American Philosophical Quarterly, 2,* 193-203.

Barnes, B. (1982). *T. S. Kuhn and social science.* London: Macmillan.

Bergmann, G. (1951). The logic of psychological concepts. *Philosophy of Science, 18,* 93-110.

Bergmann, G. (1954). *The metaphysics of logical positivism.* New York: Longman, Green.

Bergmann, G. (1957). *Philosophy of science.* Madison: University of Wisconsin Press.

Bergmann, G., & Spence, K. (1941). Operationism and theory in psychology. *Psychological Review, 48,* 1-14.

Carnap, R. (1966). *Philosophical foundations of physics.* New York: Basic Books.

Cohen, M. R. , & Nagel, N. (1936). *An introduction to logic and scientific method.* New York: Harcourt, Brace & Co.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671-684.

Eysenck, H. J. (1950). Criterion analysis: An application of the hypothesis - deductive method to factor analysis. *Psychological Review, 57,* 38-53.

Eysenck, H. J. (1960). The place of theory in psychology. In H. J. Eysenck (Ed.), *Experiments in personality* (Vol. 2, pp 303-315). London: Routledge & Kegan Paul.

Eysenck, H. J. (1967). *The biological basis of personality.* Springfield, Ill.: C. C. Thomas.

Eysenck, H. J. (1976). *The measurement of personality.* Lancaster, Eng.: MTP.

Eysenck, H. J. (1981). *A model for personality.* New York: Springer.

Eysenck, H. J. (1982). *A model for intelligence.* New York: Springer.

Eysenck, H. J. (1983a). Drug as research tests in psychology: experiments with drugs in personality research. *Neuropsychobiology, 10,* 29-43.

Eysenck, H. J. (1983b). Psychopharmacology and personality. In W. Jahuke (Ed.), *Response variability to psychotropic drugs.* (pp. 127-154). New York: Pergamon Press.

Eysenck, H. J. (1983c). Is there a paradigm in personality research? *Journal of Research in Personality, 17,* 369-397.

Eysenck, H. J. (1984). The place of individual differences in a scientific psychology. *Annals of Theoretical Psychology. 1,* 237-285.

Eysenck, H. J. (1985a). *Decline and fall of the Freudian empire.* London: Viking Press.

Eysenck, H. J. (1985b). Psychotherapy to behaviour therapy: A paradigm shift. In D. B. Fishman, F. Rotgers, & C. M. Franks (Eds.), *Paradigms in behaviour therapy: Present and promise.* New York: Springer.

Eysenck, H. J. (1985c). The place of theory in a world of facts. In K. B. Madsen & L. P. Mos (Eds.), *Annals of theoretical psychology.* (Vol. 3, pp. 17-72). New York: Plenum Press.

Eysenck, H. J., & Barrett, P. (1985). Psychophysiology and the measurement of intelligence. In C. R. Reynolds & V. Wilson (Eds.), *Methodological and statistical advances in the study of individual differences.* New York: Plenum Press.

Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences.* New York: Plenum Press.

Eysenck, H. J., & Wilson, G. D. (1973). *The experimental study of Freudian theories.* London: Methuen.

Feyerabend, P. K. (1975). *Against methods.* London: New Left Books.

Greenaway, F. (1966). *John Dalton and the atom.* London: Heinemann.

Grünbaum, A. (1984). *Foundations of psychoanalysis.* Berkeley: University of California Press.

Hanson, N. R. (1958). *Patterns of discovery.* Cambridge: Cambridge University Press.

Hempel, C. G. (1952). *Fundamentals of concept formation in empirical science.* Chicago: University of Chicago Press.

Hempel, C. G. (1965). Aspects of scientific explanation. In C. G. Hempel (Ed.), *Aspects of scientific explanation and other essays in the philosophy of science.* New York: Free Press.

Hempel, C. G. (1966). *Philosophy of natural science.* Englewood Cliffs, NJ: Prentice-Hall.

Hunter, J. F. (1979). *An analysis of validity, differential validity, test fairness and utility for the Philadelphia Police Officers selection examination.* Philadelphia: Report to the Philadelphia Federal District Court, Alverez vs. City of Philadelphia.

Hunter, J. E., & Schmidt, F. L. (1980). Noncompensatory aspects of the utility of valid personality selection. Unpublished manuscript cited by Jensen, 1984.

Hunter, J. E., & Schmidt, F. L. (1982). Fitting people to jobs: The impact of personnel selection on national productivity. In M. P. Dimutte & E. A. Fleishman (Eds.), *Human performance and productivity: Human capability assessment* (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hunter, J. E., Schmidt, F. L., & Rauschenberger, J. (1984). Methodological, statistical and technical issues in the study of bias in psychological tests. In C. R. Reynolds & R. T. Brown (Eds.), *Per-*

spectives on bias in mental testing (pp. 41-100). New York: Plenum Press.

Jensen, A. R. (1984). Test validity: *g* versus the specificity doctrine. *Journal of Social and Biological Structures, 7,* 93-118.

Kazdin, A. E., & Wilson, G. T. (1978). *Evaluation of behavior therapy.* New York: Ballinger.

Kline, P. (1981). *Fact and fantasy in Freudian theory.* London: Methuen.

Krige, J. (1980). *Science, revolution and discontinuity.* Sussex: Harvester Press.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Lakatos, I. (Ed.).(1968). *The problem of inductive logic.* Amsterdam: North-Holland.

Lakatos, I., & Musgrave, A. (Eds.). (1970). Criticism and the growth of knowledge. Cambridge: Cambridge University Press.

Morgan, T. T. T. (1978). Relapse and therapeutic response in the conditioning treatment of enuresis: A review of recent findings on intermittent reinforcement, overlearning, and stimulus intensity. *Behaviour Research and Therapy, 16,* 273-279.

Popper, K. R. (1959). *The logic of scientific discovery.* London: Hutchinson.

Popper, K. R. (1974a). *Conjectures and refutations.* London: Routledge & Kegan Paul.

Popper, K. R. (1974b). Replies to my critics. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper.* La Salle: Open Court Publishing Co.

Putnam, H. (1962). What theories are not. In E. Nagel, P. Suppe, & A. Tarski (Eds.), *Logic, methodology and philosophy of science.* Stanford: Stanford Press.

Quine, W. V. O. (1962). *From a logical point of view.* Cambridge: Harvard University Press.

Rachman, S. (1963). *Critical essays in psychoanalysis.* London: Pergamon Press.

Rachman, S., & Wilson, T. (1981). *The effects of psychological therapy.* New York & London: Pergamon Press.

Roley, T. B. (1959). An opinion on the construction of behaviour therapy. *American Psychologist, 14,* 129-134.

Schmidt, F. L. (1979,Jan/March). Poor hiring decisions, lower productivity. *Civil Service Journal.*

Schmidt, F. L., & Hunter, J. E. (1980). *Personnel Psychology, 33,* 41-60. (Quoted by Jensen, 1984.)

Schorr, A. (1984). *Die Verhaltenstherapie.* Weinheim: Beltz.

Shigehisa, T., & Symons, J. R. (1973). Effect of intensity of visual stimulation and auditory sensitivity in relation to personality. *British Journal of Psychology, 64,* 205-213.
Shigehisa, P., Shigehisa, T., & Symons, J. (1973). Effects of intensity of auditory stimulation on photopic visual sensitivity in relation to personality. *Japanese Psychological Research, 15,* 164-172.
Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy.* Baltimore: The John Hopkins University Press.
Strupp, H. H., Hadley, S. W., & Gomes-Schwartz, B. (1977). *Psychotherapy for better or worse: The problems of negative effects.* New York: Jason Eronson.
Suppe, F. (1974). *The structure of scientific theories.* Chicago: University of Illinois Press.
Tarski, A. (1941). *Introduction to logic and to the methodology of deductive sciences.* New York: Oxford University Press.
Tarski, A. (1956). *Logic, semantics, metamathematics.* Oxford: Clarendon Press.
Toulmin, S. (1953). *The philosophy of science.* London: Hutchinson.
Thurstone, L. L. (1938). *Primary mental abilities.* Psychometric Monographs, No. 1. Chicago: University of Chicago Press.
Urbantschisch, V. (1883). Über den Einfluss von Trigeminus Reizen auf die Sinnesempfindungen insbesondere auf den Gesichtssinn. *Archiv für die gesamte Physiologie, 30,* 129-175.
Zwang, G. (1985). *La statue de Freud.* Paris: Robert Laffont.