2

The Place of Theory in a World of Facts

H. J. Eysenck

Abstract. When it is frequently said, following Kuhn, that social science in general, and psychology in particular, is in a preparadigmatic phase, this may be interpreted to mean that there are no widely accepted general theories covering important areas. Psychology and other social sciences appear to suffer from the added disadvantage that not only are such theories and paradigms lacking, but professional members of these groups often regard this lack of theory as a virtue and proclaim a lack of interest in theories in general, adopting a low-level sort of empirical pragmatism.

This paper argues for the vital importance of theory in psychology and gives a number of examples to demonstrate the empirical value of such theories in gaining new and better knowledge. To quote Lewin's famous saying: "There is nothing more practical than a good theory." This is extended to empirical research as well as to practical application, and it is suggested that psychology should be more interested in theories, generate theories more readily, and come to grips with the general importance of theories in scientific work. Only in this way, it is suggested, can psychology join the ranks of the properly accredited sciences and take its rightful place.

1. The Developmental Concept of Scientific Theories

Psychology as a whole has not been very hospitable to theory, and relatively little thought has gone into the problem of just what the function of theory might be in science generally and in psychology in particular, how best such theories might serve the purposes of the psychologist, and whether it is desirable or even possible to dispense with theory (Bergman, 1951; Spence, 1944). Lewin (Marrow, 1969) is indeed credited with the statement, "There is nothing as practical as a good theory," but even he did not discuss in any detail just what made

H. J. Eysenck • Institute of Psychiatry, University of London, De Crespigny Park, London SE5 8AF, England.

K. B. Madsen et al. (eds.), *Annals of Theoretical Psychology*

a theory "good," or indeed what kind of theories might be particularly useful in psychology.[1]

Philosophers, of course, have debated such issues as these endlessly although seldom with psychology in mind (Suppe, 1974). As the contributors to this volume make clear, logical positivism and the "received view" are now decisively rejected by philosophers of science, and it is curious to think that it is precisely these rejected views which are perhaps the only ones widely known and accepted amongst psychologists (Bergman & Spence, 1941). What is this received view? Shortly after the First World War, philosophers of science began to construe scientific theories as axiomatic calculae which were given a partial observation interpretation by means of correspondence rules. For the most sophisticated presentation of this view, see Carnap (1962, 1966) and Hempel (1952, 1965, 1966). See also Tarski (1941, 1956). It was Putnam (1962) who first referred to this type of analysis as the received view on theories, and, as Suppe points, out virtually every significant result obtained in the philosophy of science between the 1920s and 1950s either employed or tacitly assumed the received view (Bergman, 1954, 1957; Harré, 1975; Quine, 1962). Then we have the period during which a number of attacks were mounted challenging the very conception of theories in scientific knowledge, beginning perhaps with Toulmin (1953).

Some of these attacks were directed at specific features of the received view, such as the notion of partial interpretation, and the observational-theoretical distinction (Achinstein, 1965; and Putnam, 1962). Other critics advanced alternative philosophies of science which rejected the received view out of hand, and proceeded to argue for some other conception of theories in scientific knowledge (Hanson, 1958; Toulmin, 1953). As Suppe (1974) points out:

> These attacks were so successful that by the late 1960s the general consensus had been reached among philosophers of science that the Received View was inadequate as an analysis of scientific theories; derivatively, the analyses of other aspects of the scientific enterprise (for example, explanation) erected upon the Received View became suspect and today are subject to much criticism. At the same time the various proposed alternatives to the Received View have been subjected to strong critical attack, and none of them has gained general acceptance among philosophers of science. (p. 4)

The received view, of course, was the product of logical positivism and it is odd that it survived long after logical positivism itself had been rejected. An explanation might be that positivism had tried to force all

[1] I am indebted to Imre Lakatos for enlightening discussions on this topic; but for his untimely death I would have benefitted even more from his incisive comments.

empirical knowledge into a scientific mold, and it was possible to reject positivism as a general epistemology on the grounds that not all empirical knowledge was like scientific knowledge. Rejecting logical positivism as the general epistemology was compatible with the willingness to concede that positivism was adequate as an analysis of scientific knowledge, and logical positivism thus became a philosophy of science and continued to be acceptable as the philosophy dealing with a restricted range of empirical knowledge, namely, scientific knowledge. However that may be, there is no doubt that at the moment philosophy of science is in a state of turmoil; having rejected both logical positivism and the received view, it is desperately searching for a new unifying theory of scientific theories (Feyerabend, 1975).

This search is only of marginal interest to psychologists, and instead of following Suppe and others in scrutinizing the ramifications of alternative theories, we might better examine not theories of science but theories within science. However, it is difficult to do this without having some general point of view which coordinates one's thinking, and this is difficult to achieve without at the same time having some idea of just what constitutes a science, as opposed to nonscience, pseudoscience, and so forth. Thus it is difficult not to devote a few words to what Popper (1959, 1974a, 1974b) calls the "demarcation" dispute, that is, the problem of what is the distinguishing mark of science, as opposed to the various disciplines which claim to be scientific (Popper singles out astrology, psychoanalysis, and Marxism) but are not. Popper's views are too well known to repeat them here; so are those of Lakatos (1968), Kuhn (1962), and many others. Here I will rather depart from present-day controversies (Lakatos & Musgrave, 1970). Most of the controversies deal with wholly developed sciences, particularly physics and astronomy; I think from the point of view of the psychologist it is more useful to think of the different metatheories of science as taking a developmental course, rather than as being opposed to each other in a fundamental sort of way. Figure 1 illustrates the view I have put forward before (Eysenck, 1976b).

The view there expressed runs counter to the usual assumption that scientific theories, and theories about the nature of scientific theories, have a universal application, but this is almost certainly not so. Scientific concepts develop in the course of history, and different methods of investigation may be appropriate at different stages. As Figure 1 illustrates, usually development starts with ordinary observation and induction; on the basis of these, the investigator develops a hunch that certain features of the observations may be invariant, that is, the sun might rise again tomorrow because in the past it has always risen again after setting.

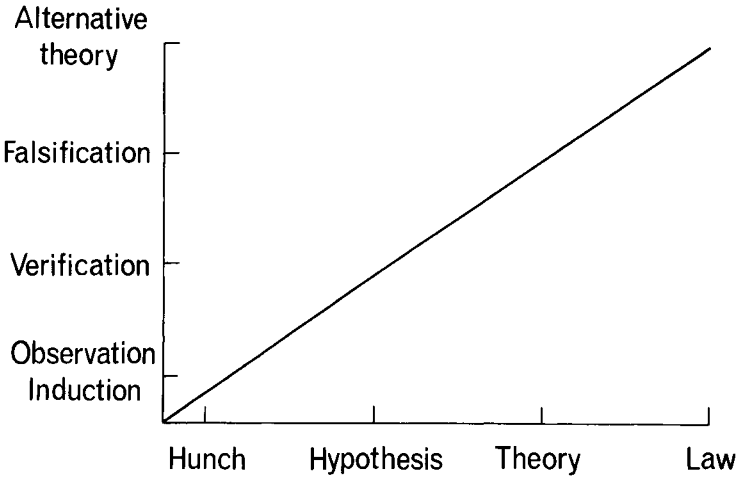

Figure 1. Changes in the nature of scientific theories corresponding to the level of development of a particular science. From Eysenck, 1976b.

Gradually limited hypotheses are formed, for instance, that the sun is moving around the earth, or vice versa. At this early stage, verification is sought of such hypotheses, and falsification is not very important, there are so many areas of ignorance that apparent falsification may not be as destructive to the hypothesis as it might be at a later stage. (A failure to observe stellar parallax did not render Copernicus's heliocentric hypothesis nugatory.) Gradually hypotheses become more firmly established, and related ones are seen to have certain features in common; out of these related hypotheses a theory is formed, such as Newton's theory of gravitation. Such a theory is highly specific in its predictions, and consequently falsification becomes important, although even at this stage simple falsification is not enough to overthrow a theory, as Lakatos has shown. Gradually theory develops into law; we tend to refer to theories which have become well established as natural or scientific *laws*. Falsification of laws is almost anathema; the anomalies in the precession of the perihelion of Mercury were known for centuries, but they were not admitted as disproof of Newton's laws. What is required is a Kuhnian revolution, in the form of an alternative theory; it needed Einstein's theory of relativity to overthrow Newton's theory. Falsification in the simple factual sense was not enough.

We can see that this developmental concept of scientific theory, and the nature of scientific thought and conceptualization, embraces all the various attempts to demarcate science as opposed to nonscience. The earliest observation–induction phase clearly corresponds very largely to the stress Bacon laid on these features of the scientific approach, and this stage may thus justifiably be called Baconian. The second stage, of hypothesis formation and verification with the stress on the latter, might be labeled the logical positivism stage, applicable to a higher level of organization and observation at which specific hypotheses are put forward and the attempt is to verify them. Once we get to the stage of wide-ranging theories, Popperian methods of falsification assume greater importance; it is taken for granted at this stage that many deductions have been and can be verified, and then it becomes more important to seek for falsification.[2] Finally, once these wide-ranging theories have become laws they become a paradigm in the Kuhnian sense, and replacement of one paradigm by another implies a revolution and a marked discontinuity in the theoretical concepts used (Krige, 1980).

Much of the debate between philosophers of science has centered on which of these different approaches is the *correct* one; from the point of view of the practicing scientist, however, there cannot be a *correct* answer to his quest for guidance in this choice of theories. This choice, and the conceptualization of science as a developing enterprise, requires rather that attention be paid to the stage of development of a given science; this will determine the kind of hunches, hypotheses, and theories which are appropriate to that science at that time. All sciences begin at a level which would be rejected by logical positivists and Popperians alike; there is a simple scrabbling for facts of the most elementary kind, uncoordinated and ill-defined. Gradually science pulls itself up by its bootstraps, a process which may not be pretty but which has received rather less attention from philosophers of science than have the later, rather more coordinated stages. Nevertheless, from the point of view of a very young science (like psychology) it is precisely the earlier stages that are of much interest, and it is doubtful whether the psychologist can receive much help from the philosophical conceptualizations of the more advanced sciences. There are, of course, links between all these stages, but there are also marked differences, and it behooves us to pay attention to these differences as well as to the similarities.

It might be thought that the stages in these developments illustrated in Figure 1 relate to an increase in rigor, in the sense that theories and confirmations which might pass muster at an early stage of theory development might not do so at a later stage. This certainly is the view of Popper, who regarded psychoanalysis as outside the scientific pale, not

[2]Actually, as Grünbaum (1976) has pointed out, Popper has misinterpreted Bacon, who was as much a falsificationist as Popper himself. I have here used the names of Bacon and others simply to denote in a rather simplified manner positions in the scheme of Figure 1, regardless of the fact that actually their thoughts were more complex than would fit easily into such a simple scheme.

because it failed inductively but rather because it did not produce theories which could be falsified and hence failed to accord with his demarcation principle. As Grünbaum (1976, 1977, 1979, 1981) has pointed out, the boot is on the other foot. Freud's theory would pass by Popper's criterion but fail by Bacon's. Popper suggested that psychoanalytic theory was a prime illustration of his thesis that inductively countenanced confirmations can easily be found for nearly every theory, as we look for them. But as Grünbaum (1981) points out, Popper ignores that the inductivist legacy of Bacon and Mill gives no methodological sanction to the ubiquitous "confirmation" claimed by some of these Freudians and Adlerians whom he had encountered in his early years. Popper regarded inductivism as probatively promiscuous because he believed the ubiquitous confirmations claimed for psychoanalysis to be sanctioned by the Bacon/Mill tenets. Grünbaum (1981) comments:

> It is ironic that Popper should have pointed to psychoanalytic theory as a prime illustration of his thesis that inductively countenanced confirmations can easily be found for nearly every theory, if we look for them. Being replete with a host of etiological and other causal hyotheses, Freud's theory is challenged by neo-Baconian inductivism to furnish a collation of positive instances from *both* experimental and control groups, if they are to be inductively *supported* instances. But . . . if such instances do exist, the psychoanalytic method is quite unable to furnish them. Moreover, to this day, analysts have not furnished the kinds of instances from controlled enquiries that are *inductively* required to lend genuine support to Freud's specific etiologies of the neuroses. Hence it is precisely Freud's theory which furnishes poignant evidence that Popper has caricatured the inductivist tradition by his thesis of easy inductive confirmability of nearly every theory! (p. 103)

Grünbaum is undoubtedly correct in his view that inductivism may pose a more serious threat to theories such as the Freudian than does the Popperian view. Indeed, Popper seems to be torn between two views which are antithetical to each other, namely, that the theories of Freud and Marx are unscientific because no testable deductions can be derived from them and the contrary view that both theories have been disproved![3] Eysenck and Wilson (1973), in their discussion of the experimental study of Freudian theories suggest that certain quite definite deductions can be made from the Freudian hypothesis and have indeed been made by Freud himself; it follows that the theory is scientific by Popper's standards (as being falsifiable), but it is not scientific by neo-Baconian standards (because these predictions are in fact either unsupported or falsified). Thus, to take a very simple case, the purely

[3]Popper (1974a) claims that Marxism was falsifiable and was indeed falsified; it was then rendered unfalsifiable. On the other hand, Freud's theory was unfalsifiable from the start.

symptomatic treatments of neurotic disorders, and even simple spontaneous remission, result in cures which do not give rise to symptom substitution or relapse; this should not happen on Freudian principles and consequently constitutes a falsification of the Freudian theory. As the deduction from Freudian principles was made by Freud himself and accepted by all his disciples, the disconfirmation is especially noteworthy.

The differences in these various forms of conceptions of scientific theories and the role of induction, verification, and falsification is not, as a consequence, a simple question of increasing rigor; it is more a question of increasing complexity, interrelation of different hypotheses, and extension of the factual realm over which the theory extends. But above all, it is a question of the general strength or weakness of the theory which is involved, and this dictates to a large extent the kind of research problem that is most fruitful at any particular stage of development. We must next turn to a consideration of the nature of strength and weakness of scientific theories.

2. "Strong" and "Weak" Theories in Science

Psychologists hold quite widely differing views about the usefulness and the general role of theory in psychology. Some stress simple inductive methods exclusively, maintaining that the time is not yet—or perhaps may never be!—when more ambitious generalizations of a theoretical kind would be appropriate. Others emphasize the similarities between all sciences and conclude that what is good for physics must be good for psychology. In all the discussion that has taken place, sight is often lost of a very simple fact which, to my mind, is crucial. Scientific theories differ among themselves to such an extent that any discussion about the place of theory in science should really be restyled: a discussion about the place of *theories* in science. In particular, there is a continuum ranging all the way from *weak* to *strong* theories, and failure to pay attention to the distinction between these two kinds of theories renders much discussion, and many criticisms, quite meaningless. This continuum, of course, is very similar in nature to that shown in Figure 1.

The theory which is always quoted as the perfect example of the application of the hypothetico-deductive method is of course Newton's theory of gravitation. This is a good example of a strong theory in science, and it may be worthwhile to look at those features of it which make it so. In the first place, it is based on a very large number of accurate observations, made over many years by large groups of people. In the second place, it brings together a number of subfields in which quantitative laws—such as those of Kepler and Galileo—had already been discovered and verified. In the third place, the phenomena in question were relatively clear-cut and unambiguous; in particular, they were not embedded in or entangled with groups of other phenomena. In the fourth place, the mathematical relations in question were not of a very complex order compared, say, with modern atomic theory. In the fifth place, and largely as a consequence of the preceding points, predictions were uncommonly straightforward and precise; verification and confirmation of deductions did not give rise to special problems.

Although for all these reasons Newton's theory is a good example for the beginner, it is seriously misleading for the practical scientist. Very few scientific theories are in fact strong theories of this kind; most of them lie toward the opposite end of a continuum going from strong to weak. The typical weak theory in science shows all the opposite characteristics to those mentioned in connection with Newton's. Only few observations, and these of doubtful accuracy, are available. Few quantitative or even qualitative laws, universally established, are available in subfields. 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 straightforward nor precise. What, it may be asked, is the use of such a weak theory?

The answer may be given in the words of the famous physicist J. J. Thomson: "A theory in science is a policy rather than a creed." I quoted these words on the front page of my book *Dynamics of Anxiety and Hysteria* (Eysenck, 1957) to indicate the heuristic nature of the theory there developed, and I believe that they contain the key to a proper understanding of the function of theory in psychology—where nearly all theories are weak theories almost by definition. The value of a weak theory, to put it briefly, lies in the fact that *it directs attention to those problems which most 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.

An example or two may make clearer what I have in mind. On the basis of the theory developed in *Dynamics of Anxiety and Hysteria,* I predicted that extraverted people would have greater reminiscence effects after massed practice than would introverted people. Using the pursuit rotor, several investigators have tested this deduction. All those using a practice period of five minutes have verified the deduction; none of those using a practice period of 90 seconds have done so. This illustrates two points. In the first place, our knowledge of the subfield of reminiscence was too restricted to allow of a more precise specification of differences between extraverts and introverts; clearly, length of practice should have been specified as an important parameter but could not be so specified because of lack of knowledge in this respect. In the second place, the partial verification of the deduction is of considerable interest; it suggests relationships between personality and learning theory which are worthy of closer study. At the same time, the direction which such study should take is emphasized by the findings; we clearly must concentrate on time relations in the practice period and presumably also in the rest period. In other words, the theory leads us to a more precise study of the growth and decay of reactive inhibition, which is supposed to underlie reminiscence; the fact that positive relations with extraversion have been firmly established *for certain time intervals* suggests that a policy of research concentrating on this aspect will not be a waste of time (Eysenck & Frith, 1977).

These considerations suggest that it is relatively absurd, particularly at an early stage of development of a scientific theory, to construct a score board, with each success and each failure of prediction written in, to give a kind of batting average. There are usually many reasons for the (apparent) failure of a prediction, but far fewer for its success; consequently, failures are of much less interest than successes in evaluating a theory at an early stage. In particular, failures may arise not because the theory is in error but because the deduction made in a particular subfield makes use of a theoretical model in that subfield which is incorrect; this does not in any way invalidate the general theory. As an example, we may take the prediction that extraverts would show a more pronounced bowing in the serial position curve effect in nonsense syllable learning. According to the Hull–Lepley theory, this effect is due to inhibition of delay, and the hypothesis that extraverts are more prone than introverts to generate inhibitory potential mediates a clear-cut prediction. This prediction could not be verified, however, and neither could another one linking depressant drugs with an increased bowing effect. A special experiment was therefore carried out to test the Hull–Lepley theory, by comparing bowing effects with and without intervals between successive presentations of the nonsense syllable series; the differences which should have appeared according to the Hull–Lepley theory failed to materialize, and consequently it was concluded that the theory was itself in error (Eysenck, 1959). Failure to verify the deduction from the general personality theory was due, therefore, not to an error in the theory itself, but to an error in that part of learning theory used to mediate a particular prediction.

Even when there is no such error, predictions may not be verifiable for a variety of reasons although the theory is in fact correct. Two well-known examples are the failure to observe parallax in stellar positions, which was one of the most direct predictions made from Copernicus's heliocentric theory of the planetary system, and the failure to discover the capillaries which according to Harvey's theory of the circulation of the blood should intervene between arteries and veins. The (correct) explanations given by Copernicus and Harvey, namely, that the stars were too far away, and the capillaries too small, to make observation of the predicted effect possible with available instruments, were not at the time susceptible to proof. Scientists tended to accept the theories in question because they unified a large number of facts, although these apparently crucial deductions remained unverified for a long time.

Altogether, the notion of an *experimentum crucis* to decide the correctness of a theory, or to decide between alternative theories, is one which appears more frequently in the pages of popular expositions of scientific method than in actual practice. It will be remembered that two members of the Thomson family, father and son, were both awarded the Nobel price in physics, one for showing conclusively that light was of the nature of a *particle*, the other that it was of the nature of a *wave*! Having performed crucial experiments to prove both of these alternative theories regarding the nature of light, physicists are still left with something which sometimes behaves like a wave, sometimes like a particle; they have also learned rather painfully that crucial experiments are seldom as crucial as they are supposed to be—even in connection with a strong theory!

We may put the whole matter slightly differently, by following a discussion given by Cohen and Nagel (1936). They take as their example Foucault's famous experiment in which he showed that light travels faster in air than in water. This was considered a crucial experiment to decide between two hypotheses: H_1, the hypothesis that light consists of very small particles travelling with enormous speeds, and H_2, the hypothesis that light is a form of wave motion. H_1 implies the proposition P_1 that the velocity of light in water is *greater* than in air, whereas H_2 implies the proposition P_2 that the velocity of light in water is *less* than in air. "According to the doctrine of crucial experiments, the corpuscular hypothesis of light should have been banished to limbo once and for all." However, as is well known, contemporary physics has revived the corpuscular theory in order to explain certain optical effects which cannot be explained by the wave theory. What went wrong?

As Cohen and Nagel (1936) point out,

> In order to deduce the proposition P_1 from H_1, and in order that we may be able to perform the experiment of Foucault, many *other* assumptions, K, must be made about the nature of light and the instruments we employ in measuring its velocity. Consequently, it is not the hypothesis H_1 alone which is being put to the test by the experiment—it is H_1 and K. (p. 115)

The logic of the crucial experiment may therefore be put in this fashion: If H_1 and K, then P_1; if now experiment shows P_1 to be false, then either H_1 is false or K (in part or complete) is false.

> If we have good grounds for believing that K is not false, H_1 is refuted by the experiment. Nevertheless the experiment really tests both H_1 *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 it need not then decide against H_1. (p. 115)

We may now indicate the relevance of this discussion to our distinction between weak and strong theories. Strong theories are elaborated on the basis of a large, well-founded, and experimentally based set of assumptions, K, so that the results of new experiments are interpreted almost exclusively in terms of the light they throw on H_1, H_2 . . . H_n. Weak theories lack such a basis, and results of new experiments may be interpreted with almost equal ease as disproving H as disproving K (see the example given above of serial position learning effects). The relative importance of K can of course vary continuously, giving rise to a continuum; the use of the terms *strong* and *weak* is merely intended to refer to the extremes of this continuum, not to suggest the existence of two quite separate types of theories. In psychology, K is infinitely less strong than it is in physics, and consequently theories in psychology inevitably lie towards the weaker pole.

Weak theories in science, then, generate research the main function of which is to investigate certain problems which, but for the theory in question, would not have arisen in that particular form; their main purpose is not to generate predictions the main use of which is the direct verification or infirmation of the theory. This is not to say that such theories are not weakened if the majority of predictions made are infirmed; obviously there comes a point when investigators turn to more promising theories after consistent failure with a given hypothesis, however interesting it may be. My intention is merely to draw attention to the fact, which will surely be obvious to most scientifically trained people, that both proof and failure of deductions from a scientific hypothesis are more complex than may appear at first sight and that the simple-minded application of precepts derived from strong theories to a field like psychology may be extremely misleading. Ultimately, as Conant has emphasized, scientific theories of any kind are not discarded because of failures of prediction,but only because a better theory has been advanced; this may account for the longevity of Hull's system in spite of the many onslaughts on it in recent years.

There is a further characteristic of weak theories, as contrasted with strong, which deserves mention. In strong theories the different postulates are *interdependent;* it is not possible to change one without changing the rest, indeed without throwing overboard the whole theory. In weak theories, such interdependence is much less marked, and changes in one part of the theory are quite permissible without the necessity of altering other parts as well. Thus Hull's theory of inhibition is peripheral and "work"-oriented; I have preferred (for various experimental reasons) to work with a central theory of inhibition rather more akin to Pavlov's. This substitution, as well as many others, can be made without extending the framework of Hull's theory unduly; nothing of this kind would have been possible with Newton's theory of gravitation. Weak theories are very flexible; that is why they are such good guides for research; strong theories have an air of "take it or leave it" which makes them superior as guides to action but also less likely to lead to important new discoveries. In a similar manner, and also dealing with reminiscence, is the substitution of a consolidation theory for an inhibition one (Eysenck & Frith, 1977). This is a totally different concept, yet it could explain all the known phenomena just as well as does the inhibition hypothesis, with the crucial addition that it also explains certain phenomena which previously could not be explained.

There is, of course, one point which all scientific theories have in common and which decisively sets them off from nonscientific theories. However weak a theory may be, it must generate predictions which admit of experimental or observational investigation. In other words, the theory must be *reality-oriented* and the manipulations of reality implicit in its testing must be capable of being made explicit without ambiguity. The hypothesis that planetary motions can be explained in terms of angels pushing the planets around on their courses, preordained by God, is not a scientific theory because it suggests no direct experimental investigation. It is an interesting question whether such concepts as the Oedipus complex, or the superego, or the archetype, have characteristics which make them suitable for use in scientific theories, however weak, or whether they are outside science altogether; it would take us too far to discuss this point in detail (Eysenck & Wilson, 1973).

One of the main consequences of having to deal with weak rather than with strong theories is that attention should shift from consideration of right and wrong to considerations of the fruitful or the useless. It is difficult enough to disprove a strong theory; it is almost impossible to disprove a weak one. The point has been well made by Roley (1959):

> Ultimately, theory must answer to the facts, but this is not the only requirement placed on a theoretical system. Logical consistency, economy of assumptions, and even a degree of elegance are by no means secondary factors in determining the overall staying power of a theory. These patrician qualities are quite unlikely to mature, however, if the demand for direct descriptive capability is too insistent. It is not . . . contended that theory construction should be totally unresponsible to the general body of knowledge about behaviour. Rather it is held that point-by-point testing of isolated facets of a theory against specific behavioural phenomena or experimental findings is at odds with the whole purpose of theoretical abstraction. Suggestive hypotheses should not be put directly to drudgery but should be entertained for a while, as rare and welcome guests. It might be thought that all this is as much applicable to a strong as to a weak theory, as indeed implied in Lakatos' model of "hard core" plus protective belt as the structure of a theoretical research programme. Perhaps the main difference would lie in the relative prominence of the "hard core" and the "protective belt" respectively! (p. 130)

Weak theories usually imply the absence of precise, trustworthy data, and it is interesting that in the history of science most strong theories started out, in fact, as weak theories; their very existence stimulated the accumulation of precise data which later transformed these theories into strong ones. The heliocentric theory of Copernicus was based on wretchedly poor and inaccurate data; indeed, available observations were so erroneous and unreliable that they could not be used to arbitrate between the Copernican and the Ptolemaic theories. It was a weak theory in every respect, but by its very existence it encouraged astronomers to seek for ever more accurate observations, until Tycho Brahe finally gathered data reliable enough for Kepler to verify his three laws of planetary motion; these in turn mediated Newton's great synthesis. This would appear to be the answer to those who feel that the publication of weak theories is premature and should await more precise measurement and other advances in knowledge; such precise measurement, and such other advances, are usually the consequence of the interest aroused by a challenging new theory, however weak. Without the existence of the theory, there would be little motivation for the more precise measurements to be made. Every strong theory started out as a weak one (but of course not every weak theory necessarily becomes a strong one); premature demands on such a theory for accuracy and rigor appropriate to strong theories are less likely to lead to advances than a realistic insight into the limitations of weak theories.

Roley (1959) goes on to point out that absolutist notions of validity in scientific theory find no counterpart in physics. He distinguishes three major types of validity:

> First, a theory may be valid in a *subjunctive* sense. That is, it may describe the behaviour of entities under conditions that never obtain in fact. The Galilean law of falling bodies, and the ideal gas laws are examples. Second, a theory may be *locally* true, that is, may hold over certain ranges of the relevant variables. Hooke's law of elasticity is a clear-cut example, and the Newtonian laws of motion (holding for "middle-sized" phenomena) are now accepted in this sense. Finally, a law may hold *statistically*, being supported by large numbers of observations although there are local exceptions. The standard example is the law of increasing entropy. In each case, the postulated "law" can be defined as the potential limit of observational approximations as certain stipulations are satisfied. (p. 131)

Roley goes on to point out:

> It may well be that behavioural laws must invoke all three of these forms of licence in relating predictions to the world of observation. In other words, we may be reduced to predicting relationships among parameters of statistical distributions which can be empirically sampled within certain limits only, and which are subject to extraneous disturbing factors. (p. 131)

In summary of this discussion, it might perhaps be fair to repeat that when we are dealing with a strong theory, successful predictions are commonplace and do not do much to enhance the validity of the theory, whereas failures to predict correctly are very serious and may be disastrous for the theory in question.[4] When we are dealing with a weak theory, however, the reverse is true. Successful predictions, particularly if unlikely on commonsense grounds, are important and valuable and do much to support the right of the theory to be considered seriously. Failures, however, are not unequivocal enough to be taken too seriously; they certainly do not damage the theory in the same way as do failures in the case of a strong theory.

We may put this point of view in terms of information theory. If we may take K as certain (or near certain), then the confirmation or information of H gives us one bit of information (provided both outcomes

[4] Successful predictions from strong theory may of course enhance the validity of the theory if they are novel and not of the same sort as previous predictions. Eddington's (1920) observations of the gravitational bending of light predicted by the general relativity theory very much enhanced the credibility of that theory.

are equiprobable.) If we take *K* as quite uncertain, then the failure of *H* to be verified may have many causes; its success, however, is unlikely to have a cause other than the one specified. Consequently, success gives us more information than failure. It gives us information not only about *H* but also about *K*. Success implies that both *H* and *K* are essentially correct; failure implies that either *H* or any part of *K* may be in error, and this is too vague to be very useful. Failure may of course be used to suggest more complex relationships, improvements in technique, or limitations in choice of parameters, but it is not in itself very informative. Information gained by successful tests of a strong theory is largely redundant, but failure is highly informative. The reverse is true of weak theories.

An example from physics may take this statement clearer. There are certain stars, the so-called "white dwarfs," to which Sir Arthur Eddington's (1920) mass–luminosity principle does not apply; they are supposed to be much more massive than might be expected from their luminosity (this is because in them matter reaches an extraordinary high degree of density at which the ordinary gas laws cease to apply). Use has been made of this condition to furnish another proof for relativity theory. This theory predicts that the apparent frequency of a periodic phenomenon, such as atomic vibration, is changed when the source of the vibration is situated in a strong gravitational field. The slowing down of the vibration shows itself in a slight displacement of spectral lines toward the red; it was observed in 1925 by Adams in the spectrum of the Companion of Sirius. Vaucouleurs (1957) points out: "As the observed effect was in good agreement with the theoretical prediction, both the theory and the existence of extremely dense matter in white dwarfs were confirmed." Here the relativity theory (H_1) is being tested in relation to a body of knowledge (*K*) not itself too firmly established; confirmation of the hypothesis is likely only if both H_1 and *K* are correct, and consequently contributes much to our knowledge of both relativity theory and of the nature of white dwarfs. Failure of the hypothesis could have been due to errors in the relativity theory or to lack of true knowledge of the compositon of white dwarfs; it would therefore have been ambiguous at best and not very informative.

What has been said in this section should not, of course, be taken to mean that the psychologist has unbridled licence to theorize to his heart's content, regardless of negative results. For a theory to be regarded seriously it must show some evidence, both internally and externally, that it can mediate experimental predictions. But once this evidence is available, early failures of *some* predictions should not be taken too seriously; they should be regarded as a challenge to discover the causes of the failure, rather than as necessarily disproving the theory. This, of course, implies a tolerance of ambiguity, an ability to suspend judgment, which is often difficult to maintain. Nevertheless, the whole history of science argues powerfully against rash decisions on the basis of inadequate data; time and time again we have seen buried theories rise afresh from the ashes to which they had been prematurely consigned.

Nor should expectations be too high of close relationships between experimental variables. Much refinement will have to go into the raw scores obtained from psychological tests before we can begin to claim to measure one single variable rather than a mixture of the most diversified traits, abilities, and attitudes. The best that can be expected is a set of low correlations usually in the expected direction, but occasionally directly opposed to prediction; on such a foundation we can then begin to erect the infinitely complex set of laws and functional relationships, concepts, and definitions which will ultimately, shorn of ambivalence and ambiguity, constitute that proper science of behavior and personality which so obviously does not exist at the present time, except possibly as a foundation for a palimpsest.

3. The Two Disciplines of Scientific Psychology

Before turning to more detailed theories within psychology, let me discuss at some length certain overall views which masquerade as theories, which are seldom verbalized, and which have determined very strongly (and adversely) the whole structure of scientific psychology. Attention was drawn to these "meta theories" by Cronbach (1957) in his presidential address to the APA, which had the same title as this section. Broadly speaking, he pointed out that there are two different traditions within psychology

> two historic streams of method, thought and affiliation which run through the last century of our science. One stream is *experimental psychology;* the other, *correlational psychology.* Dashiell (1938) optimistically forecast a confluence of these two streams, but that confluence is still in the making. Psychology continues to this day to be limited by the dedication of its investigators to one or the other method of enquiry rather than to scientific psychology as a whole. (p. 671)

The experimental method is characterized by the simple fact that the experimentalist studies the effects of certain manipulations of the environment, which may take place mostly in the laboratory but may also take place in everyday life situations. He is concerned with general laws, the effects of his manipulation on the average person, and the elaboration of laws which would adequately represent the observed changes.

By contrast, the correlational psychologist is concerned with individual differences; he observes the different reactions of people under identical conditions and searches for regularities by looking for correlations between people in different situations. In a manner of speaking the experimentalist is concerned with the mean, the correlationist with the standard deviation; one looks for generality, the other specificity. Both clearly are concerned with important aspects of human behavior, and one would have expected them to collaborate in the solution of the very complex and profound problem presented by human behavior. As Cronbach points out, however, this is not so. The two groups are hardly on speaking terms; they know little of each other's facts and theories; they read different journals. This is clearly a tragedy for psychology, and it is the result of a rather unthinking acceptance of traditions which are based on unacceptable theories which are seldom formulated but whose hold is very tenacious indeed.

Let us first of all look upon the sins of the experimentalist and the implicit theories (metatheories?) governing his conduct. He is following a general theoretical formulation of the problem of psychology which may be called functional; in this formulation, he looks upon the dependent variable as a function of the independent variable, and by manipulating the latter and studying the variation in the former he expects to be able to delineate general laws which will become the basis of a true science. His general formulation: $a = f(b)$, that is, a is the function of b, is pervasive, and, within limits, of course perfectly acceptable. It is clearly advantageous to see to what extent varying b causes changes in a, and insofar as that is the program of the experimentalist it can hardly be faulted. In exactly the same way can we say that physics or chemistry or astronomy studies functional relationships of a similar kind. This notion of functional relationships has of course given rise to the old-fashioned behaviorist stimulus–response method of analysis; the stimulus is the independent variable, the response the dependent variable, and what is to be studied is the functional relationship between the two.

The simple comparison with physics, however, leaves out of account a very fundamental aspect of the functional relationships recognized by physicists. Let us consider Hooke's law of elasticity: Stress $= k \times$ strain, where k is a constant (the modulus of elasticity) that depends on the nature of material and type of stress used to produce the strain. This constant k, that is, the stress–strain ratio, is called Young's modulus and is illustrated (with certain simplifications) in Figure 2a. A and B are two metals differing in elasticity; they are stressed by increasing loads, and

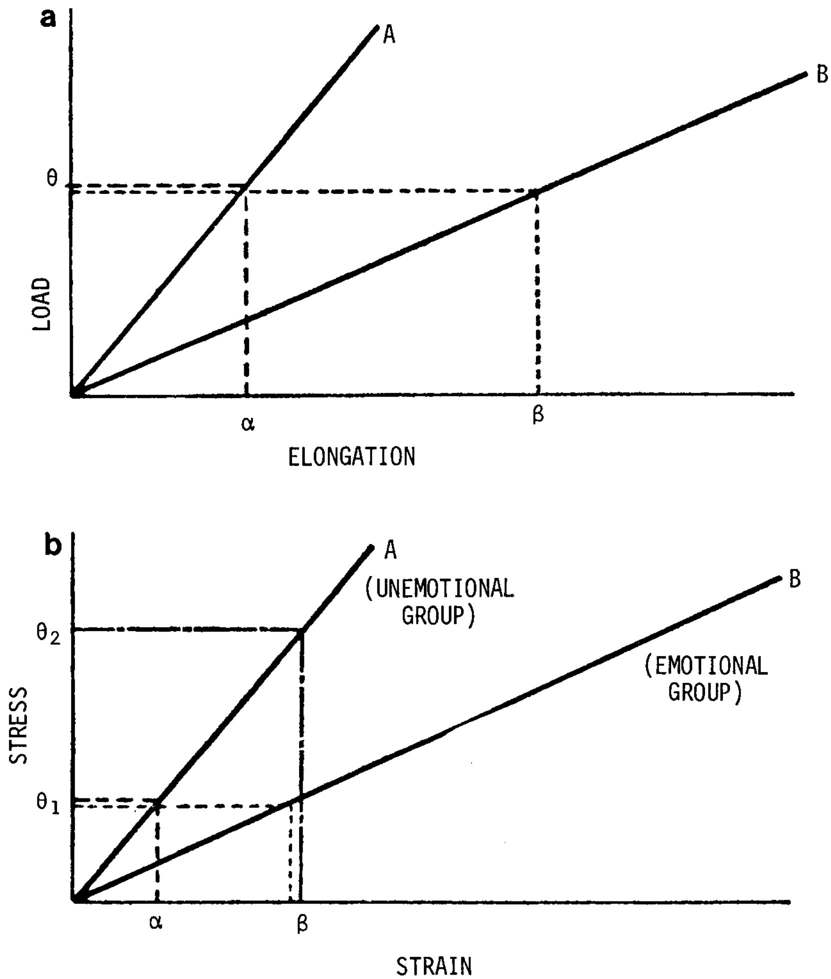

Figure 2. The concepts of stress and strain in physics (top diagram) and psychology (bottom diagram). From Eysenck, 1975b.

the elongation corresponding to each load is plotted on the abscissa. It will be seen that identical loads give rise to quite divergent elongations, α and β. Thus the physicist does not have the simple stimulus (independent variable)—response (dependent variable) relationship; it also incorporates the term *k*, which corresponds to the notion of *organism* in psychology. The more recent formulation: S—O—R attempts to do justice to this situation, but such a formulation certainly carries with it the implication that we ought to change our functional law to read: $a = f(b,0)$, in other words, *a* is a function not only of *b* but also of the particular nature of the organism which is being stimulated, just as the elongation of the wire in the experiment detailed in Figure 2 is a function partly of the nature of the wire, partly of the weight used to stretch it. As far as physics is concerned, strain would correspond to the experimental method, k to the individual differences or correlational method, and no physicist would ever doubt that both are essential for a united science (Eysenck, 1975b).

Let us use the same argument in relation to psychology. Figure 2b illustrates an analysis of human behavior (physiological, verbal, expressive) in an experimental situation productive of emotion. Again the stress (independent variable) is plotted on the ordinate, and the strain (dependent variable) on the abscissa; A and B represent an emotionally stable and an emotionally unstable individual or group of individuals respectively. Identical stress θ_1 gives rise to quite different strains α and β. It would require stress θ_2 to make the strain in A individuals equal to that produced by θ_1 in B individuals. Differences between θ_1 and θ_2 are the kinds of differences traditionally studied by experimental psychologists; differences between A and B are the kinds of differences traditionally studied by personality psychologists, believers in the importance of constitutional factors, and clinical psychologists. In order to understand what is going on, in order to formulate theories, and to make predictions, we must incorporate both types of factors in our model.

This model may be called compensatory or substitutional in the sense that one type of variable can compensate for or be substituted for the other (Savage & Eysenck, 1964). To produce a certain elongation greater than x we can either increase the weight suspended from the wire or choose a wire the k of which indicates greater pliability. On the human side, consider a series of experiments carried out by Rosenbaum (1953, 1956). He found that threat of a strong shock led to greater generalization of a voluntary response than did threat of a weak shock; this would be a typical experimental approach. He also discovered that anxious subjects showed greater generalization to *identical* stimuli than did nonanxious subjects; this would be the typical correlational or individual differences approach.[5] To obtain a given degree of generalization, we can thus either change the strength of the threatened shock or choose more or less anxious subjects; the two can be traded against each other. Savage and Eysenck (1964) have carried out a series of experiments with rats illustrating the same point. Their chapter is headed "Definition and Measurement of Emotionality," indicating another important function

[5]It is interesting to note that Rosenbaum published one paper in the *Journal of Experimental Psychology*, the other in the *Journal of Abnormal and Social Psychology*, thus ensuring that very few psychologists would read *both* papers!

of this whole theoretical concept of experimental and individual variables as being to some degree interchangeable.

We may ask ourselves whether a given test or measure which is used to indicate differences in emotionality or anxiety, whether in humans or in animals, is in fact a good measure of that variable, and this whole question of validity of tests has of course always been a very difficult one. It can be solved, however, along the lines of the above argument by looking at experiments which are generally accepted as increasing an organism's degree of emotionality. We can then argue that if experimental modification x could produce an increment in emotionality, which can be indexed by performance y, then if our personality test a is a good measure of emotion or anxiety, people having high scores on this test should behave like persons subjected to a strong dose of x, that is, show strong y, whereas people having low scores on a should have low scores on y. Thus this conception of the relationship between experimental and individual differences variables gives us a powerful tool for demonstrating the validity of personality measures and adding substantially to the range of application of the experimental design.

Another important consequence of these theoretical considerations is that if the suggestions made above are true, then it follows that we cannot make verifiable predictions from general laws without incorporating specifically a variable k which refers to the constitution of the individual for whom the prediction is being made. Consider an experiment reported by Jensen (1962). He studied the number of errors in serial learning as a function of the rate of stimulus presentation; there were two rates, one of 2 seconds and one of 4 seconds. The traditional experimental psychologist would regard this problem as meaningful and soluble; either differences in rate of stimulus presentation cause a difference in the number of errors to criterion, or they do not. Jensen argued that this test imposes a stress on the subject and that the resulting strain would be indexed in terms of an increased number of errors when the shorter rate of stimulus presentation was employed, as compared with the longer rate. (There is an obvious pressure with the quick rate of presentation which might well increase the general stress of the test situation.)

This stress would impose differential strain on the individuals respectively high and low on trait anxiety, and Jensen measured this trait anxiety with the neuroticism scale of the Maudsley Personality Inventory. Contrasting subjects scoring high and low respectively on this scale, he found that for low scorers (i.e., nonemotional individuals) the added stress of shortening the rate of stimulus presentation produced no effect at all; they made 63 errors on the average for the long rate and 64 errors for the short rate. But for the high scorers there was a tremendous difference; in the long rate of presentation condition they only made 46 errors, far fewer than the unemotional subjects, whereas in the short rate of presentation conditions they made 90 errors! In other words, the results of the experiment can only be understood in terms of the person $\times$ condition interaction; leaving out the differential personality effect makes complete nonsense out of any simple averaging of the results. Such averaging would tell us that there was a mild and nonsignificant effect of shortening the rate of presentation, when in reality there was no effect for nonemotional subjects and a very strong effect for emotional ones.

I feel very strongly that just as it would be meaningless in physics to leave out the constant *k* in dealing with predictions of the kind considered, so it is meaningless in psychology to try and frame general laws without taking into account individual differences. In fact, this is even more so in psychology than in physics because of certain very relevant differences between these two sciences. In physics we can usually proceed along very analytic lines by experimentally excluding certain variables and by completely dissecting the objects of our interest until we are dealing with simple elements or alloys of known composition. In psychology, however, we cannot do this. By definition, we are dealing with organisms and their behavior; as a consequence, we are not allowed to cut up the organism in such a way as to isolate certain aspects. The integrity of the oranism must be maintained, and that means inevitably that the personality, the intelligence, and other important functions of the organism will play an important part in whatever measurement we may be concerned with. To relegate these individual differences to the error term, as experimentalists are wont to do, simply means that the error term will become enormously exaggerated in size and the main effects (unless in trivial and obvious experiments) will be much smaller than is acceptable in a scientific discipline. We can rescue these variables from the error term by looking at interactions, and many experiments have shown that these interaction terms can frequently be much more important than the so-called main effects (Eysenck, 1967, 1981). This general rule has certain important consequences for the design of experiments meant to test perfectly general hypotheses in experimental psychology which have been spelled out and illustrated elsewhere (Eysenck, 1976b). The many different experiments there quoted to illustrate this point suggest that the use of personality parameters in experimental psychology is not only permissive but *mandatory* (Eysenck, 1981). An experiment in which personality variables are theoretically likely to play a part should never be planned without either including these personality variables in the analysis of variance design or at least measuring them in order to use them as moderator variables in the final analysis.

Anything less is an abuse of the scientific method and is likely to lead to the absurd state in which much of experimental psychology finds itself, namely a failure to replicate, an inability to formulate wide-ranging theoretical conceptions, and an error term which far outweighs the main effects in the experiment. A combination of the experimental and the individual differences approaches is essential if psychology is to become a science at last.

It might be thought that the studies mentioned do not deserve the title of *experimental* if this term is interpreted in the traditional way of changing the independent variable and observing the resulting changes in the dependent variable in a laboratory setting. I believe that this interpretation is too narrow. Physicists and astronomers usually consider Eddington's observation of the gravitational effect of the sun on light coming from a star situated apparently close to the eclipsed orb as being an experiment, although the eclipse was not of course physically produced by Sir Arthur. In a similar way MZ and DZ twins are not actually produced by the experimenter; like Sir Arthur, he is making use of experiments conducted by nature, going to considerable pains to ensure that conditions of observation are such as to make the experiment as well controlled as if it had been conducted under laboratory conditions. Such an empirical trial of a given hypothesis is far removed from the usual correlational or factor analytic study of the psychometrist and deserves to be considered closer to the rubric of experiment considered in its narrowest aspect.

4. The Contribution of Experimental Psychology to Research in Individual Differences

In the last section we have discussed the need for experimental psychology (and also social, educational, industrial, clinical, and other areas of applied psychology) to take into account and incorporate in experimental designs the major relevant dimensions of personality. It should not be imagined, however, that trading between experimental and individual differences in psychology is entirely one way; nothing could be further from the truth. In this section we will consider the importance of introducing experimental methods and theories into the field of individual differences, using as our example the study of intelligence. Readers with a philosophical background will note that I have concentrated here as elsewhere on giving only a brief discussion of the general theoretical viewpoint and much more space to a discussion of specific application of that theoretical point of view. In addition, it will be noted that I have concentrated on a historical treatment. The reason for adopting these methods of presentation is twofold. In the first place, I am not a professional philosopher, although I have received some training in that discipline; neither are most of my readers likely to be philosophers. Consequently an *ad oculos* demonstration of the points made, in their historical development, will be more easily intelligible and more compelling than would a purely theoretical discussion. In addition, I have always believed that philosophy of science is crucially dependent on an accurate history of science; the history of science supplies the empirical foundations for any pronouncements that the philosopher of science may make in the hope that it might be helpful or enlightening for the working scientist. Hence my stress on historical examples and developments.

The developments here discussed have been documented extensively in *A Model for Intelligence* (Eysenck, 1983); hence no extensive references will be made to the literature. The topic to be discussed is the development of the notion of intelligence and the theories concerning it which have been developed by psychologists and others during the past hundred years or so. In looking at this development, we note from the beginning an antithesis which became quite clear in the views and methods of measurement advocated by Sir Francis Galton, on the one hand, and Alfred Binet, on the other. For Galton, intelligence was a single determinant of cognitive behavior, entering into all types of activity which could be called cognitive rather than emotional or conative. This activity he conceived to be determined largely by genetic causes, and as a consequence he recommended that it be measured by as direct an index of neural functioning as could be devised. In particular, he recommended the study of reaction time; the development of electrophysiological measurement, such as the EEG, would undoubtedly have attracted his interest and would have led him to advocate the use of these more recondite procedures.

Binet, on the other hand, believed that there were a number of *different* abilities determining cognitive functioning, such as verbal ability, numerical ability, suggestibility, memory, and so forth. His view was that these should all be measured and that intelligence was a rather artificial concept based on the sum of all these different abilities. Furthermore, he seemed to believe that an individual's abilities were determined to a large extent, if not entirely, by environment, socioeconomic factors, cultural influences, and education. Thus we see right from the beginning a difference in conception which has persisted ever since.

The Galtonian tradition was continued by Charles Spearman, who introduced factor analysis into the debate and believed that he had demonstrated the existence of a general factor of intelligence by means of his method of tetrad differences. Thurstone and later Guilford also used factor analysis and believed that they succeeded in demonstrating the existence of a number of primary abilities, to some extent resembling those suggested by Binet. Guilford at the moment claims in his model of intellect that already he has evidence for something like 80 or more of the 120 different abilities he postulates. It will be seen that factor analysis has been unable to settle the question, and this illustrates very clearly my own belief that problems of this kind cannot be settled by the individual differences approach only, using correlational or factor analytic methods, but that experimental tests of specific hypotheses are needed.

In my very first contribution to this field (Eysenck, 1939) I suggested that a reanalysis of Thurstone's original data indicated a compromise between the two extreme positions, namely, the existence of a strong general factor accompanied by a number of rather weaker and independent primary abilities. This view was later accepted by Thurstone and Spearman and is probably more widely accepted now than either extreme. However, it cannot be claimed that on the basis of correlational studies such a view can find any direct and incontrovertible *proof*; there are many psychometric and empirical reasons for preferring it to the extreme views originally adopted by Spearman and now by Guilford (in opposite directions!), but reasonable expectations and compromise solutions are a long way removed from the convincing proof that scientists rightly demand for the adoption of a particular hypothesis.

When we come to the question of tests used and the determination of genetic influences, we find that until quite recently the balance was very much in favor of Binet. He introduced tests of an educational character, depending to some or even to a large degree on acquired knowledge, skills, and habits, and consequently inevitably subject to a strong determination of scores indicative of individual differences by environmental factors. Genetic studies summarized elsewhere (Eysenck, 1979) indicate that even so the proportion of the total variance on such tests accounted for by genetic factors is 80%, that accounted for by environmental factors only 20%, but in practical terms this 20% is still a quite large proportion of the total variance and makes it impossible to draw any conclusions regarding individual persons. Heritability estimates are always population estimates and cannot be applied to individuals; hence tests of this type are inherently incapable of telling us very much about the specific genetic endowment of a given person.

It is interesting to consider historically why Binet won out and Galton lost in the type of measurement of intelligence mostly used by psychologists. An early study of the ability of reaction time experiments to discriminate between more and less intelligent individuals was carried out early this century by Clark Wissler at Columbia University with negative results, and this study so much influenced psychologists (and found its way into most textbooks) that until quite recently very little work had been done along these lines. Acceptance of Wissler's results can only be explained in terms of the zeitgeist, because his experiment must be considered to be one of the worst ever done in experimental psychology. It is well known that to measure reaction time properly a large number of determinations have to be made, as there is considerable variance around the mean; yet Wissler only tested each individual a very small number of times. It is well known that to obtain a reasonable correlation of any given test with intelligence the range of ability must be large and should be roughly equivalent to that obtaining in the whole population; yet Wissler chose advanced university students at Columbia University, showing a very small range of ability. And finally it is well known that as a measure of intelligence we must use a proper IQ test; Wissler chose achievement tests, which are known to correlate very little if at all with IQ tests in the kind of population he used. In these circumstances a low correlation was inevitable and proved nothing. Yet his research has been cited again and again to indicate the lack of value of Galton's suggestion!

Recent work has demonstrated conclusively that reaction time in fact correlates very highly with Binet-type measures of IQ such as the Wechsler test. Admittedly tests of simple reaction time only correlate less than .5 with IQ in children or adults showing the normal spread of IQs, but other measures of reaction time have been found much more useful. Hicks's law has been used to look at complex reaction times; this law simply states that for any individual reaction times will increase in a linear fashion as the logarithm to the base 2 of the number of stimuli presented. The slope thus defined differs from person to person, with people having a high IQ showing a smaller increase in reaction time with increase in the number of stimuli than do individuals having a lower IQ. Using this slope or simply correlating IQ with reaction times to some eight differential stimuli gives us correlations in the neighborhood of .50, uncorrected for attenuation, which means that the true correlation must be somewhere in the neighborhod of .70 or thereabouts.

Another variable which has been found to be highly correlated with IQ (even more so than reaction time) has been the *variability* in reaction times shown by a given individual. People with a high IQ tend to show low variability, people with a low IQ high variability. This feature thus can be added to our determination of IQ by reaction time, making the correlation even higher than before.

Other similar developments have been related to the measurement of what has come to be called *inspection time.* Here the subject is shown on the tachistoscope two lines, one very obviously much longer than the other; he has to say which is longer. (Backward masking is used to avoid the use of after-images, etc.). The lines are presented for very short periods of time a number of times, with the position of the long and short lines varying on a random basis, of course, and the experimenter determines the shortest duration of exposure at which the subject gives an accurate estimate 97.5% of the time. It has been found that high-IQ individuals have shorter inspection time than low-IQ individuals, and a similar finding has been made when the test was translated into the auditory field. Inspection time techniques work better with individuals of below- and above-average IQ, and the correlations with below-average IQ subjects again extend into the 70s. We thus have here another correlate of Binet-type IQ which follows the theories of Galton rather than those of Binet.

More important than these developments, however, have been recent applications of theories and methods of electrophysiological measurement of so-called evoked potentials. When the EEG of a given individual is taken and a sudden visual or auditory stimulus is introduced, a series of rapid waves can be seen which constitute the so-called evoked potential. These waves extend over a period of 500 milliseconds at most, but for measurement purposes only the first 250 milliseconds are considered. It has been known for some dozen years that the latencies and amplitudes of these waves show some slight correlation with IQ, not usually above .3; but as these correlations were difficult to replicate, and as they were not high enough to be of any great importance, little has been made of these findings. In addition, there seemed to be little theoretical basis for these findings, and purely pragmatic correlations of this kind are of little interest in science.

In recent years there has been a considerable change in this position, due largely to the theoretical work of Alan Hendrickson and the measurement of IQ by means of evoked potential, following this theoretical work, of Elaine Hendrickson (Eysenck, 1983). Essentially what these two have done is this: Alan Hendrickson developed a theory of information processing through the cortex, which goes into considerable physiological and biochemical detail. It would be going too far to reproduce this theory here, but two aspects of it are essential for an understanding of the new type of measurement undertaken by Elaine Hendrickson. The first point of this theoretical development is that the evoked potential, in all its details, is *homologous* with the detailed information concerning the stimulus which is being propagated through the nervous system.

The second aspect to be considered is the fact that in the propagation of the message errors will occur. What the theory now says is simply that people in whom many errors occur will have a low IQ, whereas people in whom few errors occur will have a high IQ. We thus have here a very simple causal hypothesis about the reasons why certain individual differences occur, and this theory is eminently testable.

How does this theory influence our measurement? Clearly, latencies and amplitudes are largely irrelevant to the measurement of IQ following this particular hypothesis. What is important is the following: As is well known, evoked potentials have a low signal-to-noise ratio and hence have to be averaged over a number of time-locked presentations; in our own work we have usually used 90 such presentations. Now if the resulting averaged curve is homologous with the message in all its details, under ideal conditions of no error, then the presence of error should reduce the *complexity* of the curve, until with individuals in whom a great deal of error is present only the broadest outline should be shown, that is, simple waves without any of the numerous small ups and downs so characteristic of evoked potentials. In other words, what we should be measuring is the *complexity* of the waveform, and this is what in effect has been done by the Hendricksons.

The theory predicts that this complexity measure should correlate highly with IQ, and in one of their studies Elaine Hendrickson found a correlation of .83 with the WAIS over a population of some 200 school children, with a mean IQ and a standard deviation closely resembling the normal population. (In other studies she found similar correlations for adult populations, also using other tests of IQ such as the Matrices.) These results thus show a very high correlation between evoked potential and IQ, a correlation which would be higher still if corrected for attenuation.

A factor analysis of the correlations between the 11 subtests of the Wechsler and the evoked potential complexity measure showed a very prominent general factor on which the evoked potential had a loading of .77; the highest loading observed for any of the Wechsler Tests was .82 (for the Similarities test). We must now turn to a theoretical consideration of the consequences which may be derived for a better understanding of intelligence from all these data (Eysenck, 1985).

Following Binet, most American psychologists have tended to look for an environmental interpretation of the individual differences found on ordinary IQ tests. Most would probably admit that heredity plays a very important part, but interest has largely centered on the possibility of changing a person's IQ by educational, cultural, and other modification of the environment. Head Start is but one example of the interest shown in this possibility. Such an interest is understandable in view of the mixed nature of the typical IQ test, giving as it does scores which are determined partly by genetic, partly by environmental factors. It is obvious that results from such tests will be difficult to interpret and will give rise to endless controversies, such as those attending Jensen's suggestion of genetic factors' being responsible for racial differences in IQ. In the nature of things it is almost impossible to answer such questions in a direct manner, and we are reduced to complex statistical argumentation which is unlikely to prove totally convincing to the adherents of either side. Clearly it would be very advantageous if, by going back to Galton's approach, we could derive a measurement technique which would as far as possible approach the genotype without the inclusion of cultural, socioeconomic, and educational factors. The study summarized above suggests that this possibility is a very real one; measures of reaction time, inspection time, and evoked potential all correlate between .70 and .80 with traditional IQ tests, and a look at the detailed statistics suggests that such a test as evoked potential must be very near to being an almost pure measure of innate ability, without the addition of culturally determined factors (Eysenck & Barrett, 1985).

Such a hypothesis can be tested and has been tested in the following manner. Twenty-five children from low socioeconomic status homes and 25 children from high socioeconomic status homes were given the Wechsler, with a difference of 23 points between them, which amounts to 1.64 standard deviations using the standard deviation found in the total group from which these were extracted. If we accept the estimate of 80% of the total variance on the Wechsler being contributed by heredity, 20% by environment as correct, and if the evoked potential measure is a true measure of genetic differences only, then we would predict that on this evoked potential measure the difference between the children will be reduced by 20% (i.e., the 20% contributed to the Wechsler by environmental factors), and this is actually what we do find (Eysenck, 1983). It is not suggested that this single small-scale experiment *proves* the accuracy of the evoked potential as a measure of the genotype; it is merely quoted as an example of the type of deduction which can be made from hypotheses of this kind. Obviously many more such experiments are needed and much larger numbers of subjects will be required before we can rest content with demonstration as actively proving the point.

The implications of these findings for a theory of intelligence are considerable. On Galton's hypothesis, the high correlations between reaction time and evoked potential on the one hand and the different tests constituting the Wechsler on the other are to be expected; indeed, one would have expected that Wechsler subtests having high loadings on the general factor of intelligence would have high correlations with these external criteria and tests having low loadings would have low correlations. This application of *criterion analysis* (Eysenck, 1950) has indeed been shown to work along these lines, and it is difficult to see how such proportionality could arise if we disallow the existence of a strong and general factor of intelligence running through all the subtests of the Wechsler. (The very high loading of the EEG measure on a factor analysis of the Wechsler subtests is another way of stating the same set of interrelations.) Thus introduction of an external, experimentally defined criterion makes it possible for us to go beyond the vicious circle of purely relational and factor analytic investigations and to make a more formal test of the Galtonian hypothesis (Eysenck & Barrett, 1985). Clearly, the finding does not deny the existence of additional "primary abilities" such as are proposed in the compromise solution, but the data completely rule out solutions such as those suggested by Guilford, that is, the absence of a general factor of intelligence and the distribution of the total variance over a large number of special abilities.

Criterion analysis is a quite general and very powerful test of certain types of psychological hypotheses, linking psychometric research on individual differences with direct experimental analysis. Two illustrations may be given, both deriving from work reported by Jensen (1981a). The first relates to inbreeding depression. This is a genetic phenomenon manifested in the offspring of parents who are genetically related (father-daughter, cousins, etc.). Such inbred offspring show a depression or diminution in those characteristics which are in some degree genetically influenced by directional dominance, such as intelligence. This is due to the fact that when recessive alleles detract from the positive expression of a trait, inbreeding increases the chances that recessive alleles from each parent will be paired at the same loci on the chromosomes, thereby diminishing the phenotypic expression of the trait. The degree of inbreeding depression for any given trait is normally assessed by comparing measurements of the trait in the inbred offspring of genetically related parents with measurements in the offspring of unrelated parents.

Now it is known that intelligence is inherited in a mode which includes dominant and recessive genes (Eysenck, 1979), and consequently one would expect that tests which correlate highly with intelligence (have high *g* loadings) would show a *greater* degree of inbreeding depression than tests having a low correlation with intelligence (having low *g* loadings). Jensen plotted the *g* factor loadings of 11 WISC subtests as a function of the percent of inbreeding depression on subtest scores. (The tests are *I*nformation, *C*omprehension, *A*rithmetic, *S*imilarities, *V*ocabulary, *P*icture Completion, *P*icture *A*rrangement, *B*lock *D*esign, *O*bject *A*ssembly, *D*igit *S*ymbol, and *M*azes.) It will be seen (Figure 3) that there is a very high correlation of 0.76 between the *g* loading and the percentage of inbreeding depression, verifying this particular deduction from the theory. (The correlation is unexpectedly high because the range of *g* loadings is of course very restricted, going from .5 to .8. Had the range been greater, clearly the correlation would have been very much higher.)

The second example, also taken from Jensen (1981a), presents a test of what he calls the Spearman hypothesis of white–black differences. Spearman (1927, p. 379) hypothesized that the varying magnitudes of the mean differences between whites and blacks in standardized scores on a variety of mental tests were directly related to the size of the tests' loading on *g*. Jensen re-analyzed nine independent studies giving data making such an analysis possible and found strong confirmation for Spearman's hypothesis. The largest set of relevant data was based on the GATB, a widely used and well-standardized battery of tests of cognitive abilities; the comparison used populations of white and black subjects of well over 1,000. Figure 4 shows the *g* loadings, the white–black mean difference (in standard score units), and the average correlation of each aptitude scale with 12 well-known standard tests of IQ or general intelligence, for each of the nine aptitudes measured by the GATB. There is clearly a close similarity between the profiles for these three variables, and correlations calculated by Jensen among the three

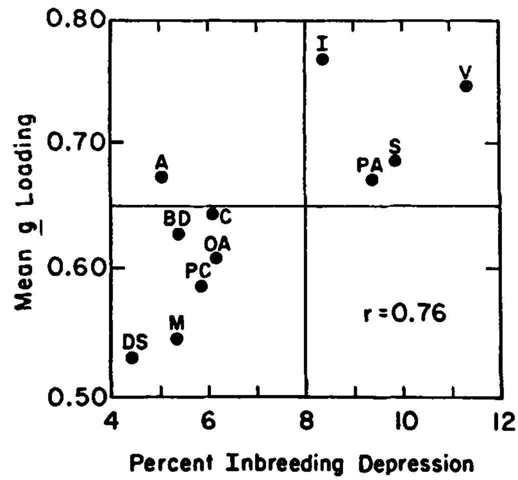

Figure 3. Mean *g* loadings and percentage inbreeding depression, as shown in 11 Wechsler subscales. From Jensen, 1981a, with permission.

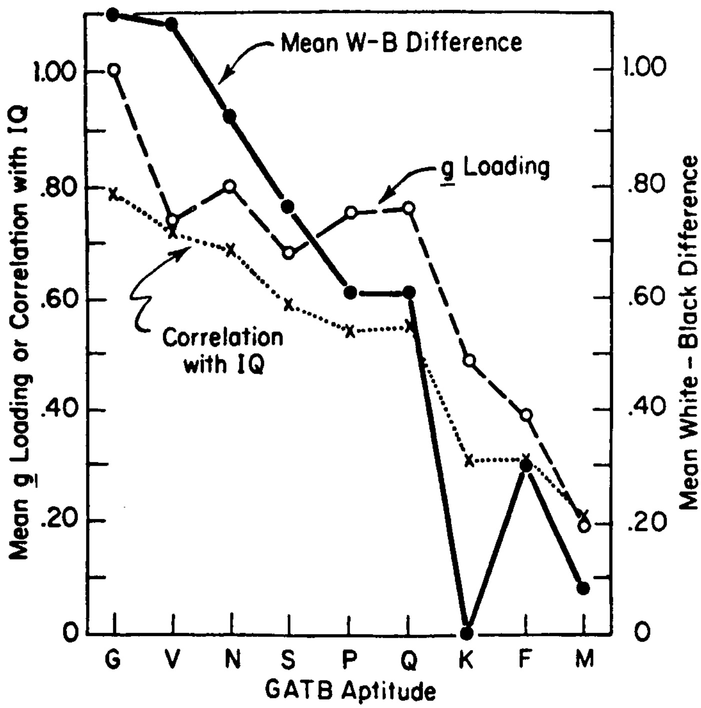

Figure 4. Mean g loadings or correlations with IQ and mean white–black differences on subtests of the GATB aptitude test. From Jensen, 1981a, with permission.

are astonishingly high, ranging from .9 to 1.0. Thus complex psychometric theories are capable of resolution by combining the psychometric (factor analytic) approach with direct experimental testing along the lines of criterion analysis.

The implications of the findings on the electrophysiological measurement of IQ go much further than already noted. The contamination of orthodox IQ tests with educational and cultural material has always made difficult the study of questions which are of considerable interest to psychologists, such as the development in young children of intellectual ability, the decline of this ability in old age, and differences in ability between males and females or between different national or ethnic groups. Differences are usually less than the 20% of the variance allocated to environmental factors in the genetic analysis of IQ data, and consequently it can be argued endlessly that they are due entirely to the environmental factors. Having available tests which are independent of environmental factors of an educational or cultural kind would make it possible to resolve these problems in a direct, scientific manner, without having to have recourse to complex statistical arguments which make assumptions which can always be challenged. The only such problem to have been attacked hitherto is that of sex differences (Eysenck, 1983). It was found that among the boys and girls tested there were no mean differences in evoked potential scores, but boys had a larger *variance* than girls, which is in good accord with the literature on orthodox IQ tests. Thus on this important question the evoked potential gives results similar to those previously obtained with IQ tests, but as expected the difference in variance is greater for the EEG measure than for the IQ. Differences in variance on the ordinary IQ would be likely to be lessened by the introduction of cultural and educational factors which would tend to equalize variances; here the argument is exactly the opposite to that presented previously for differences in SES.

Even more important from the theoretical point of view are the repercussions these findings have on our conception of the *nature* of intelligence. It is well known that there is little agreement among psychologists on the definition of intelligence or on its nature; what most definitions appear to agree on, however, is that intelligence is related to problem solving, the induction of relations and correlates, cognitive-type work of one kind or another, the learning of complex and difficult tasks, whether verbal or nonverbal, and so on. All these definitions emphasize culturally determined types of activities which are based in large or small part on education and previous experience of one kind or another. The new theory suggests that the only important variable in intelligence is the individual's innate structure of the CNS enabling this system to propagate messages correctly across axons and synapses. This is an elegant and intuitively appealing notion, but it opens up certain important problems. Thus we would imagine that in solving a problem more would be involved than simply passing along information correctly through the central nervous system; some kind of use would have to be made of this information in order to obtain a correct solution. Yet the observed correlation coefficients leave very little leeway for any additional factors over and above the simple propagation of correct information through the cortex; it would appear that bringing together correct items of information relevant to the problem automatically produces the solution! This is certainly counterintuitive and may deserve some discussion.

Let us note first of all that the information in question is not necessarily only information coming in through the senses and coming from the outside; we would include equally information obtained through memory retrieval, that is, items of information stored in the cells of the CNS and addressed during the information-processing stage. It is equally difficult to think of a solution's being arrived at directly from the free interplay of the bits of information involved as it is to think of a separate problem-solving stage—such a stage easily conjures up the view of a homunculus sitting in the nervous system, obtaining the necessary information and then producing the solution! It has to be said that at the moment the theory has no real contribution to make to this fundamental problem, other than to state that in mathematical terms very little of the variance, if any, is left over for a separate problem-solving stage.

One further theoretical problem requires discussion, namely, that of the relation between the different types of noncognitive tests we have enumerated as being so closely related to cognitive ability. If evoked potentials, reaction time, and inspection time all correlate around .80 with IQ, then clearly these different measures must correlate quite highly together themselves, and the question arises as to which is fundamental. An additional question is whether we can account for the observed findings on reaction and inspection time in terms of what we have found or hypothesized with respect to evoked potentials. The following is thus a brief sketch of a possible solution: If we agree that IQ (in its genetic aspects) is determined completely by the probability in a given CNS that information will be processed more or less accurately, than it would appear to follow that high-IQ individuals should have short reaction and inspection times, because erroneous processing of information would require a large number of redundant messages to be passed before reaction is possible. Similarly, we can account for the fact that variance in reaction times is negatively correlated with IQ in the same terms; when the signal is passed correctly through the cortex, then reaction times will be short; when it is in error, reaction times will be long because further messages will have to be passed before reaction is possible. Thus the greater the error propensity of a given CNS, the more variable will the results be. This explanation also accounts for the fact that even for retardates, the fastest reaction times are just about as fast as the fastest reaction times of highly intelligent subjects. This can be explained in terms of the fact that even in a CNS given to making many errors occasionally these errors will be absent. We can thus see that there is considerable agreement between these different noncognitive tests of cognitive ability, pointing to a common causal theory, namely, that of error-proneness. Obviously all these hypotheses are in an early stage of development and much further work will be required to provide backing and support. However, judging simply on the basis of existing knowledge we now seem to have a unified theory of intelligence based on firm physiological foundations, which enables us to measure intelligence without the usual adulteration provided by socioeconomic, cultural, and educational factors irrelevant to "pure" intelligence (Eysenck, 1984).

This whole chain of reasoning, against its historical background, has been introduced in order to demonstrate most forcibly how the introduction of experimental methods into the correlational field can deliver us from the inherent impossibility of allocating the total variance in a meaningful and convincing manner to different hypothetical causal elements. In order to do this we must formulate hypotheses that can be tested in a more direct manner than is possible through factor analysis itself, that is, along experimental lines. This does not mean, of course, that the elements involved in such hypotheses cannot also be introduced into a factorial design; it should be possible to make precise predictions (e.g., along the lines of criterion analysis) as to the outcome of such studies, and as we have seen the outcome hitherto has been very positive. But the essential point is that such theories and experimental studies have an explanatory value which does not appear in purely statistical "factors" of the kind usually studied by psychometrists. Here, then, we have a demonstration that just as the study of individual differences is indispensable for the experimentalist, so experimental studies are indispensable to the psychometrist. The two disciplines of scientific psychology stand and fall together.

5. Theory or Weltanschauung

Theories in science generally, and in psychology in particular, can of course be of varying types, and it is important to bear in mind the precise degree of generality of a theory. Clearly such theories as those regarding the nature of intelligence or personality, as discussed in previous sections, are very general; they enable a large number of deductions to be made and tested. In experimental psychology theories are usually on a much smaller scale; an example might be the theory of reminiscence (Eysenck & Frith, 1977). But we also have even more wide-ranging and far-reaching types of theory which begin to meld into something more closely akin to a weltanschauung. As an example we might take the doctrine of behaviorism.

Behaviorism, particularly the naive construction of Watson but also the neobehaviorism of later practitioners, clearly had its roots in French empiricism (Condillac, Bonnet, and Descartes properly understood) and French materialism (La Mettrie, Helvétius, and Cabanis). The objections of behaviorists to introspection sound in many ways similar to the French schools' objections to the concept of a soul, that is, a revolt against idealism in any of its forms. This philosophical stance has been the mainspring of behaviorism and is much more important than some of the accidental features which were the product of Watson's and Skinner's particular outlook, such as the denigration of physiology (shared with Condillac) and of genetic factors. Insofar as behaviorism is more than a methodology, it must be regarded as a philosophical doctrine and hence more in the nature of a weltanschauung rather than a scientific theory.

Another type of weltanschauung (or zeitgeist, as Boring prefers to call it) has already been mentioned in connection with work on intelligence; this is the belief in equality, also originating in its modern form in France (Rousseau). This belief has powerfully influenced American psychology in particular and has led to the neglect of genetic studies and the disregard of many empirical findings such as those mentioned in the last section. Egalitarianism of this kind is clearly a philosophical and political belief, not a scientific theory, although it may masquerade as such. The test of a scientific theory, of course, is the possibility of making deductions from it which are capable of being tested. Insofar as egalitarianism denies the importance of genetic factors it might be said to give rise to testable deductions, but clearly the refusal of egalitarians to carry out such tests, or to consider in their writings the results of tests carried out by geneticists, removes their beliefs from disproof. Theirs is not what Lakatos (1968) would call a program of research; their concern is entirely with what Lakatos calls a "protective belt of auxilliary hypotheses" which can be modified when empirical difficulties arise. But for any kind of research program that Lakatos envisages, such a protective belt surrounds a hard core of theoretical postulates and intepretations; it is this hard core that is completely missing in the weltanschauung of the egalitarian. His method is simply to deny the observed facts and to denigrate those whose experiments are responsible for the emergence of these facts.

Consider what Lakatos has to say about his general view:

> One of the crucial features of sophisticated falsificationism is that it replaces the concept of *theory* as the basic concept of the logic of discovery as a concept of *series of theories*. It is the succession of theories and not one given theory which is appraised as scientific or pseudo-scientific. Thus members of such series of theories are usually connected by a remarkable *continuity* which welds them into *research programmes*. This *continuity*— reminiscent of Kuhnian "normal science"—plays a vital role in the history of science; the main problems of the logic of discovery cannot be satisfactorily discussed except in the framework of a methodology of research programmes. (p. 84)

But the weltanschauung of the egalitarian does not give rise to such a series of theories, or such a methodology of research programs. It therefore departs decisively from science altogether, and egalitarianism should therefore not be regarded as a theory in the scientific sense; it is an expression of the zeitgeist, inspired by ideological and political beliefs and hostile to empirical demonstration in scientific discovery.

The degree to which egalitarianism and environmentalism are weltanschauungen is clearly shown in the vagaries of Marxist thought. During Stalin's reign, environmentalism and egalitarianism were fashionable among communists on both sides of the Iron Curtain, and Lysenkov's heresy was widely accepted. Since then there has been a return to a more correct interpretation of the thinking of Marx and Lenin, as shown for instance in the books on intelligence published by Guthke (1978) and Mehlhorn and Mehlhorn (1981) in East Germany, the work of Lipovechaya, Kantonistova, and Chamaganova (1978) on the heritability of intelligence among Russian school children, and many other sources. Mehlhorn and Mehlhorn, to take but one example, warn against the danger of calling scientists reactionary who point to the biological limitations in the development of intellectual functions and call such an attitude "*unmarxistisch*" (p. 7); they quote Marx and Engels' *Die Deutsche Ideologie* to illustrate the importance of genetic factors in Marxist thinking and go on to a quotation from Lenin (1965):

> When one says that experience and reason demonstrate that men are not equal, one understands under equality the equality of intelligence, or the similarity of physical strength and cognitive ability of men. It is of course clear that in this sense men are not equal. No single reasonable man and no single socialist can forget this. *This* equality has *nothing* to do with socialism. (p. 137)

Lenin goes on to say that to extend the notion of equality into this field is an "absurdity," and further that "when socialists speak of equality they always understand *social* equality, i.e., equality of social position but not equality of physical and mental abilities of individual persons" (p. 140). And the Russian psychologist Krutezki (1974, p. 140) concludes after a review of classical Marxist writings: "When it is said, 'from each according to his abilities', then it is clearly indicated that human beings in this respect are not equal." The conclusion Mehlhorn and Mehlhorn draw from their discussion is as follows:

> From the point of view of Marxism it is equally important to uncover the biological foundations and conditions of mental development of human beings as to study the specific influence of environment to the extent that these influence the development of mental capacities. (p. 8)

This view is largely identical with that always presented by western psychologists whose tendency to include genetic as well as environmental influences in their explanation of individual differences in intelligence has often led to their being castigated as anti-Marxian and even fascist. *Sic transit gloria mundi!*

An important consequence of these studies is a revision of the concept of discrimination, as widely used in sociology and politics. It is often suggested that the simple fact that a given group (women, blacks, children of low SES parents) is more or less frequently represented in certain social groups (university students, educationally subnormal classes, etc.) is sufficient evidence for the existence of discrimination of a racial or sexual kind in that field. Boudon (1973) defines inequality of educational opportunity and mechanisms of social immobility in these terms. By simply demonstrating that children of low SES parents are less frequently represented in student populations, he believes that he has demonstrated the existence of discrimination. But of course this would only be true if the children of high and low SES groups had similar abilities which enabled them to profit equally from university education; this is clearly not so. Yet throughout his book Boudon does not even mention the demonstrated existence of such differences, and his argument throughout is based on the unstated hypothesis of complete human equality with respect to intelligence. In a similar way political arguments are often based on the observed differences in proportions of blacks or other racial minorities in various socially defined or educationally relevant groups, but without precise knowledge of the actual IQs of the groups in question no deductions can be made about discrimination.

A recent government-inspired study of school performance of white, black, and Indian children in England demonstrated quite clearly the inferior performance of black West Indian children; the conclusion drawn was that racial discrimination was at the base of this. Yet the Indian children, against whom equal discrimination is shown in England, had school achievement scores actually superior to the white children, if anything! This agrees with the known IQ performance of these three groups of children, but the report made no mention of these IQ differences, nor did it draw attention to the fact that one group of socially disadvantaged and discriminated against children did better than the whites whereas the other did much worse. Thus the definition of discrimination has to be very carefully phrased in terms of known biological differences between groups.

We might say that generalized world concepts like egalitarianism are not even a "degenerating problem shift," to use Lakatos's term, but goes right outside the field of science altogether. What is interesting in much modern debate is this argument between rationality and irrationality, between science and zeitgeist, with the advocates of the latter trying to use the language of science but failing to come up to its spirit.

Such battles are of course not unheard of even in the field of the "hard sciences." Galileo's fight against the inquisition and Darwin's battle against the fundamentalists are still well remembered. A somewhat less well known case is that of atomism, the theory that nature is made up of atoms and their interaction. This theory, originating with Democritus and finally entering modern science through Dalton (Greenaway, 1966), was, as Bernal (1969) points out, not only a scientific but also a philosophical theory introducing materialism and radicalism into science. As he goes on to point out, in the nineteenth century the knowledge of thermodynamics began to reach into chemistry and even into biology, thanks to the work of Le Chatelier and Gibbs, and for a while it seemed that the whole of natural phenomena could be explained in terms of simple observables of mechanical energy and heat. In the hands of philosopher scientists like Mach (Blackmore, 1972), and chemists like Ostwald this produced a new positivism which stated that matter and physical hypotheses such as atoms were no longer necessary and that the whole of science could be deduced directly from elementary observations. Dumas in France and many others followed this lead and denied the existence of atoms until the work of Einstein and Perrin made such recalcitrance untenable. The kinetic theory of heat, evolved by Maxwell in 1866, had indeed implied the existence of atoms,but these were entirely hypothetical, and new evidence was required before these atoms could be accepted as measurable and countable material objects. The whole history of the atomic theory, and its belated acceptance by physicists, is a lesson in the determination of scientific theories by the zeitgeist and the prevailing weltanschauung.

The link between weltanschauung and such an all-embracing theory as behaviorism (if indeed this can be called a theory in the scientific sense) is of course complex and difficult to trace. Behaviorism as a methodology does not carry the implications of behaviorism as doctrine, but it is the latter which has had much the greater impact. After all, behaviorism as methodology simply preaches the virtue of the scientific approach; it adds little which is specific to psychology. Behaviorism as doctrine, however, does make a number of statements which, although not logically implicit in a behaviorist-methodological approach, are in good agreement with the American weltanschauung of the early 1920s. The ability to mold and manipulate people's behavior; the control over responses by controlling stimuli; the specificity of responses, and the absence of organized personality and cognitive structures; the determination of behavior by environment rather than by heredity; finally, the influence of social rather than physiological factors—all these mirror the hopeful and optimistic American outlook characteristic of that period, in which the application of science to psychological and social problems seemed to hold the answer to all the ills of mankind. A similar kind of zeitgeist prevailed in the Soviet Union, where the optimistic hope of being able to create "Soviet man" in the new image favored the conditioning and behavioristic proclivities of Pavlov, who became an honored hero.

Fundamental in this general approach was the notion of *specificity*. This thesis proclaimed essentially that all human behaviors were specific products of learned S–R connections and hence capable of environmental influence and change; this view, of course, goes back to Locke's *tabula rasa* image of the human mind.

Thorndike (1903) already preached this philosophy when he held that:

> There are no broad, general traits of personality, no general and consistent form of conduct which, if they existed, would make for consistency of behaviour and stability of personality, but only independent and specific stimulus–response bonds or habits. (p. 29)

This doctrine of "sarbondism" as McDougall used to call it, with its attending notions of the equipotentiality of the conditioned stimulus, has been decisively disproved in the fields of personality and intelligence (Eysenck, 1970, 1979), but it still persists in the public domain and is supported by ideologically motivated psychologists either unaware of or disregarding the established facts. Court judgments, based on evidence given by professional psychologists in contravention of the established facts, may illustrate the prevalence of this attitude.

First, consider the *Hobson* v. *Hansen* case, which abolished ability grouping in the schools of Washington, D.C., in 1967. Judge J. Skelly Wright stated:

> The skills measured by scholastic aptitudes tests are verbal. More precisely, an aptitude test is essentially a test of the student's command of standard English and grammar. The emphasis or these skills is due to the nature of the academic curriculum, which is highly verbal; without such skills a student cannot be successful. Therefore, by measuring the student's present verbal ability the test makes it possible to estimate the student's likelihood of success in the future. Whether a test is verbal or non-verbal, the skills being measured are not innate or inherited traits. They are learned, acquired through experience. It used to be the prevailing theory that aptitude tests—or "intelligence" tests as they are often called, although the term is obviously misleading—do measure some stable, predetermined intellectual process that can be isolated and called intelligence. To date, modern experts in educational testing in psychology have rejected this concept as "false". Indeed, the best that can be said about intelligence insofar as testing is concerned is that 'it is whatever the test measures.' . . . The IQ tests must be recognized as artificial tools to rank individuals according to certain skills.

In more recent years, in the well-known *Larry P.* v. *Wilson Ryles* case (1979), which resulted in prohibiting the use of standardized intelligence tests for the placement of black and Hispanic pupils in classes for the mentally retarded in California schools, Judge Robert Packham, in his final decision, expressed the specificity doctrine very succinctly: "IQ tests, like other ability tests, essentially measure achievement in the skills covered by the examinations . . . the tests measure the skills tested, and each of the tests subject to this litigation assesses very similar skills." He went on to quote the psychologist Leon Kamin as a witness for the plaintiffs as follows: "IQ tests measure the degree to which a particular individual who takes the test has experience with the particular piece of information, particular bits of knowledge, the particular habits and approaches that are tested in these tests."

The law, of course, as Dickens's Mr. Bumbles observed, is an ass; we do not expect judges to give expert rulings on scientific topics. That psychologists and educationalists should still be found to give evidence along these lines is proof of the overriding strength of the zeitgeist when confronted with indisputable facts; the facts are simply disregarded, as are the theories tested experimentally and giving rise to these facts. If IQ tests measure nothing but specific items of information, how is it that this information correlates so highly with nonverbal and culture-fair tests of the type exemplified by Raven's Matrices? How can items of information, each specifically acquired in isolation, give such high correlations with reaction time, inspection time, or evoked potentials? The specificity theory is not a theory in the scientific sense; it is an expression of a weltanschauung, a child of the zeitgeist, an ideological commitment. As such it is clearly proof against any factual disconfirmation, just as religious beliefs counter Galileo's and Darwin's factual evidence with irrelevant theological arguments. It is important for psychologists to recognize the difference between ideology and theory, the evocation of the zeitgeist and a scientific hypothesis. It is the absence of such recognition which makes the reading of psychological journals such a sad experience.

One of the worst consequences of the invasion of science by weltanschauung has been the interpretation of scientific data along preconceived lines, without even a consideration of alternative hypotheses. The developmental, educational, and clinical literature is full of examples of the *post hoc ergo procter hoc* fallacy, in which theoretically neutral findings are interpreted along the lines of one of several competing theories. Explanations are practically always framed in terms of the specificity-environmentalist hypothesis. Consider a study in which a child shows behavior *Y*, which is correlated with parental behavior *A*. This simple correlation is practically always interpreted causally in terms of the influence of *A* on *Y*; *A* is considered the independent variable, *Y* the dependent variable. However, clearly many other causal chains are conceivable and indeed have strong empirical support.

Suppose that the child's behavior is one of aggression, cruelty, and sadism, and the alleged environmental influence is that the parent cruelly beats the child at various times. Now it is, of course, theoretically conceivable that beatings of this kind may influence the child to behave in an aggressive and cruel manner later; if such a correlation is found, the data certainly do not contradict the hypothesis. But an alternative hypothesis would be that the genes which are partly responsible in causing the parents to beat the child savagely are inherited by the child and produce similarly cruel and sadistic behavior in him. There is much evidence for the genetic determination of aggressive and similar impulses, and the hypothesis certainly has stronger *a priori* support than does the much more widely accepted environmental one.

A third hypothesis is also tenable, of course, namely, that children who behave in an abnormally aggressive and sadistic manner are so difficult to manage that some parents are reduced to beating them savagely in order to establish their authority and preserve some degree of discipline. Other theories are of course also possible, and these three theories in turn are not mutually exclusive—all three causal chains may apply in individual cases, or even in all cases. What is suggested is simply that the existence of a correlation between A and Y does not *establish* any of these theories as correct; it simply fails to disconfirm any of them but gives no clue as to which (if any) is the correct causal theory—assuming (*pace* Hume) that causal theories are meaningful! In these circumstances, therefore, no causal interpretation can be given, and the fact that such causal interpretations along environmentalist lines are practically always given by the author (and accepted by the referees and the editor, who ought to be acting as guardians of the scientific conscience!) indicates the prevalence of a weltanschauung which renders both authors and judges blind to the realities of the situation. Clearly the theory itself is a scientific one in the sense that it can be disproved; but certainly the influence of the zeitgeist is felt in the *interpretation* of the facts along predetermined lines. Yet such studies are presented as being scientific in the strict sense, that is, an hypothesis is presented, an experiment is conducted to test the hypothesis, and the results are declared to bear out the hypothesis. What is neglected, of course, is an essential element of all scientific theory testing, namely, the consideration of alternative theories and the design of "crucial" experiments to decide between these theories. Scientists nowadays are aware of the difficulty of constructing a truly crucial experiment, but the complete neglect even to consider alternative theories is certainly not in line with the modern philosophy of science.

I have devoted a whole section to this discussion of differences between theories and weltanschauung and the disastrous effect which the latter can have on the testing of the former. Such a consideration is particularly relevant in psychology whereas in the "hard" sciences we no longer suffer from this type of indoctrination, at least to any noticeable degree. In biology the battle between fundamentalism and evolution has again emerged, at least in the United States, but biologists overwhelmingly reject the interjection of theological arguments based on a non-scientific weltanschauung. It is in psychology (and even more so in sociology) that the battle is joined most crucially. It is here that weltanschauung dictates to a large extent the interpretation of results, and even the very kinds of results which are acceptable. Jensen (1981b) gives some horrifying examples of this tendency. This is what he has to say on one case in a very large research project, one of potentially great public importance regarding the educational effects of school busing. He reports that the research was suddenly halted before it was half-completed, and the explanation given to him by a school official was that "The school system is a political unit, not a research institute, and cannot ignore political pressures in the community." Just two weeks before being treated to this shocking announcement, Jensen reports having been in Washington and having been told by a high government official in the White House that he was being overly naive to think that, at that time, he would be allowed to carry out bona fide research on the effects of school busing.

Other points made by Jensen (1981b) concern the funding of research. Referring to some instances, he says:

> They raise the question of the ethics of accepting research funds when there are strings attached as to the possible outcomes of the study, or restrictions on the reporting of results. (p. 15)

In one case, for example, the funding agency said they would consider supporting the research only if they could know beforehand the conclusions that the investigators intended to reach. In another instance, the funding agency would make the grant only if different racial groups than those originally proposed by the researchers were used in the study (whites and Asians had to be used instead of whites and blacks). In yet another case, a granting agency stipulated:

> Although data could be obtained on different racial groups, the researchers could not report group means or standard deviations, or any other statistics that might reveal the direction or magnitude of the group differences in scholastic abilities, but could report only correlations and factor analyses among different test scores. (p. 18)

These and many other examples quoted by Jensen show the grip which the zeitgeist has on research, whether concerning the funding of research, the organization of research, or the interpretation of results. It appears clear that explicit instruction in the differentiation between weltanschauung on the one hand and scientific theory on the other should form a very important part of the training of psychologists and sociologists. Only in this way will the vicious consequences of the disregard of this distinction be avoided.

6. The Biosocial Nature of Man: Theory or Metatheory?

If the very notion of *theory* is somewhat uncertain in the minds of psychologists, that of *metatheory* is probably even less clear. The term has been used by social psychologists who follow Marxian dialectics to denote the hypothesis that scientists are influenced in their work by social and economic factors and by the structure of the society in which they live, but the more critical followers of this view (e.g., Mackenzie, 1981) have acknowledged that while their views can be supported to some extent by arguments drawn fron the consideration of historical developments, theirs is not a scientific theory in the ordinary sense and cannot be falsified in ways that would satisfy Popper. Philosophical notions such as idealism or materialism might be thought of as metatheories, but they are too far removed from the nitty gritty of psychological research to be easily accommodated under that heading; it would be better to retain the title of *philosophical* for views of this kind. The same might perhaps be said about questions such as the body–mind relation; here the connection is closer, but testability is equally remote.

As an example of metatheory in psychology one might perhaps choose the conception of man which informs a psychologist's choice of theories, topic of research, and method of approach. We have already had occasion to refer to various different approaches to the nature of man, such as the respective weight given to genetic or environmental factors, but in an even more general way one might say that psychologists tend to be biophilic or sociophilic—in other words, they tend to prefer concepts of man emphasizing his *biological* nature or his *social* nature. It would here be contended that man is a biosocial animal and that there is no way in which we can understand man or interpret psychological data correctly unless and until we take into account this fundamental conception. This is not properly speaking a theory in the ordinary sense because it would be difficult to prove in a quantitative sense and specific deductions from it are difficult to make because the conception is such a very wide one, embracing the whole of psychology. However, the concept finds uses in directing research along the right channels and may be judged in relation to the extent of its success in doing so. The concept of man as a biosocial animal would certainly put an end to the absurd squabbles between those who embrace a 100% environmentalism and those who favor an equally exclusive biological reductionism.

It might be objected that such extremes are not really to be found in modern psychology, and that to oppose them would be to erect a man of straw. Up to a point, of course, such an objection would be justified; few psychologists are as incautious as Kamin (1974, 1981) who carries environmentalism to extremes by denying that there is any evidence for genetic factors in the causation of individual differences in intelligence. However, as pointed out in the last section, in practice many if not most psychologists certainly act as if they did hold such a view and interpret empirical data which are quite ambiguous in this respect entirely in terms of environmentalist concepts.

Few psychologists would be found at the opposite end of this continuum, but the recent doctrines of sociobiology as advocated by Wilson (1975, 1978), some of the authors in Caplan's book (1978) and many others certainly tend in that direction, although they are certainly less dogmatic and doctrinaire than extreme environmentalists like Kamin (1974). Wilson and his followers have taken up, at a higher level, the arguments of William McDougall originally advanced in his series of debates with Watson. McDougall's theory of instincts, of course, would not now be seriously supported by any biologist in its very primitive form, but in essence he was right and Watson was wrong in the importance attributed to the biological factors in human conduct.

What ended the supremacy of the behaviorists, of course, was the rise of the ethological school in Europe; Tinbergen, Lorenz, and many others demonstrated beyond any doubt the existence, importance, and specificity of mammalian instincts. McDougall had been right, in principle if not in detail, and Watson had been wrong; Watson's success in the argument had been a disaster for psychology, and we shall have to make good the years the locusts ate. With a realization of this sad calamity has come a realization of the importance of individual differences and of biological and genetic factors in psychology. In the treatment of mental disorder, to take but one example, behavior therapy, based on principles of conditioning pioneered by Pavlov, is taking the place of psychoanalysis, demonstrating greatly superior powers of alleviating distress (Eysenck, 1977). Personality theory, relating individual differences to biological factors (limbic system; reticular formation) is again getting into its stride (Eysenck, 1976a). Above all, genetic research into individual differences among human beings is again taking its rightful place, using new and much improved methods that were unheard of even a few years ago (Mather & Jinks, 1971).

These new developments are of crucial importance to any appraisal of sociobiology, although curiously enough this relevance has not hitherto been brought out clearly by Wilson or any of his followers. We may see how this comes about by looking at the three alternative states for our species which Wilson discusses in his foreword to Caplan's (1978) book. Either, he says, natural selection has exhausted the genetic variability underlying social behavior, or else the social genotype is uniform but prescribes a substantial amount of instinct-like behavior; or finally, some variability in human social behavior has a genetic basis and as a consequence at least some behavior is genetically constrained. He concludes:

> The evidence immediately available seems to leave room only for the last conclusion, that human social behaviour is to some extent genetically constrained over the entire species and furthermore subject to genetic variation within the species. (p. 3)

With this conclusion it would be difficult to quarrel (although as Caplan's book of readings shows, many people have managed to do just that!) It rests securely on two legs, one the phylogenetic type of evidence surveyed in Wilson's (1975) book, using evolutionary theory to account for human social behavior, the other the ontogenetic evidence of modern behavioral genetics, using the methods of biometrical genetical analysis to sort out the contributions to phenotypic variance of genetic and environmental factors.

Curiously enough, Wilson relies almost exclusively on the weaker of these two sources and seems to shun the stronger. In his first book he hardly ever mentions biometrical genetics; in his second book hardly more than two pages out of 260 are devoted to a desultory discussion of this evidence, and even this discussion is unsystematic, inaccurate, and not integrated with the remainder of the book. If there is to be a criticism of sociobiology, then I think it must be this failure to see that it stands securely on both feet, rather than totter insecurely around on one foot, with very little help from the other. If Wilson's argument had to rest on one line of evidence alone, then surely he has made the wrong choice; the ontogenetic argument is inherently the stronger because it rests on direct, experimental evidence rather than on brilliant argument from possibly shaky foundations, impossible in the nature of things to prove directly.

The major difference between Wilson's standpoint and mine is brought out very clearly in a sentence in his 1978 book:

> Human social behaviour can be evaluated . . . first by comparison with the behaviour of other species and then, with far greater difficulty and ambiguity, by studies of variation among and within human populations. The picture of genetic determinism emerges most sharply when we compare selected major categories of animals with the human species. (p. 84)

I would suggest that the argument from comparison with other species is beset by far greater difficulty and ambiguity than that from studies of variation among and within human populations; Wilson's own admission that "sociobiological theory can be obeyed by purely cultural behaviour" is ample evidence for this view.

Critics have sometimes suggested, as does Kamin (1974), that biologically oriented researchers favor this view because it supports the status quo, whereas socially oriented researchers favor environmentalism because it permits more freedom for social change. This belief that a person's scientific stance is determined by his political view is not borne out by historical fact. Watson, the archenvironmentalist, was also an archconservative; J. B. S. Haldane, one of the leaders of the genetic-biological camp and a precursor of sociobiology, was one of the leaders of the Communist Party in Great Britain! Noam Chomsky, too, is of the left wing politically, but favors genetic theories. *Argumenta ad hominem* arising from this ancient and often disproved notion should be laid to rest now; even if the correlation were perfect between social views and political affiliation, nevertheless the arguments in favor of either side would still have to be answered—throwing doubts on the scientist's motivation does not disprove his argument.

It is perhaps an ironic comment on the ideological onslaught which the presentation of genetic hypotheses in biology (Wilson, 1975), psychology (Eysenck, 1975a), history (Darlington, 1969), the study of race (Baker, 1974), and other social fields has provoked that ideology itself has been found to have strong genetic roots and to be intimately linked with personality factors genetically determined (Eaves & Eysenck, 1975, 1977; Eysenck & Wilson, 1978). In a large-scale twin study, Eaves and Eysenck (1974) found that radicalism–conservatism had a heritability of 65%; toughmindedness, a factor identifiable with ideological commitment, had a heritability of 54%. The tendency to voice extreme views, irrespective of right- or left-wing bias, had a heritability of 37%. This tendency and toughmindedness were found to be genetically connected with appropriate personality variables. It would thus appear that not only are left-wing ideologues wrong in assuming that scientists hold genetic views because they have been environmentally conditioned to defend the status quo; their own antigenetic views would appear to have a genetic basis! *Difficile est non satiram scribere.*

What is the upshot, substantively, of these considerations as far as the nature of human nature is concerned? In one sense, empirical studies simply support what common sense would unhesitatingly proclaim: man is a biosocial animal, whose aims and motives are shaped in part by his ancestral inheritance, in part by the pressure of the society in which he grows up and has his being. Curiously enough such a generalization would probably be approved by almost all geneticists, psychologists, biologists, sociologists, psychoanalysts, historians, and anthropologists who have given serious consideration to the problem; unfortunately such approval would be little but lip service in the majority. Even so, such lip service is the homage that vice pays to virtue; fundamentally we all know that nature and nurture are but the opposite sides of one and the same coin and that neither could exist without the other. The only real problem is a quantitative one; for particular groups and situations, what is the relative contribution of either? Such quantitative considerations demand a quantitative reply, and at present only the methods of biometrical genetical analysis can give us such an answer—qualified by the smallness of samples, their unrepresentative nature, and the unreliability of our measuring instruments, but nonetheless a first step in the unending quest for more precise information.

The general notion that man is a biosocial animal is of course too general to be of direct use in framing hypotheses. However, it can be rendered more precise by seeking the help of sociobiology and modern physiological research. As an example of how this can be done, consider the work of MacLean (1973) on "a triune concept of the brain and behaviour." As MacLean points out:

> Perhaps the most revealing thing about the study of man's brain is that he has inherited the structure and organisation of three basic types which, for simplification, I refer to as reptilian, old mammalian, and new mammalian. . . . It cannot be over-emphasized that these three basic brains show great differences in structure and chemistry. Yet all three must intermesh and function together as a *triune* brain. The wonder is that nature was able to hook them up and establish any kind of communication among them. (p. 7)

The hierarchy of these three brains is shown in Figure 5. Man's oldest brain is basically reptilian, forming the greater part of the upper brainstem and comprising much of the reticular system, midbrain, and basal ganglia. It is characterized by greatly enlarged basal ganglia which resemble the striatopallidal complex of mammals. The old mammalian brain (paleocortex) is distinctive because of the marked development of a primitive cortex which corresponds to the limbic system. Finally, there appears late in evolution a more complicated type of cortex, neocortex,

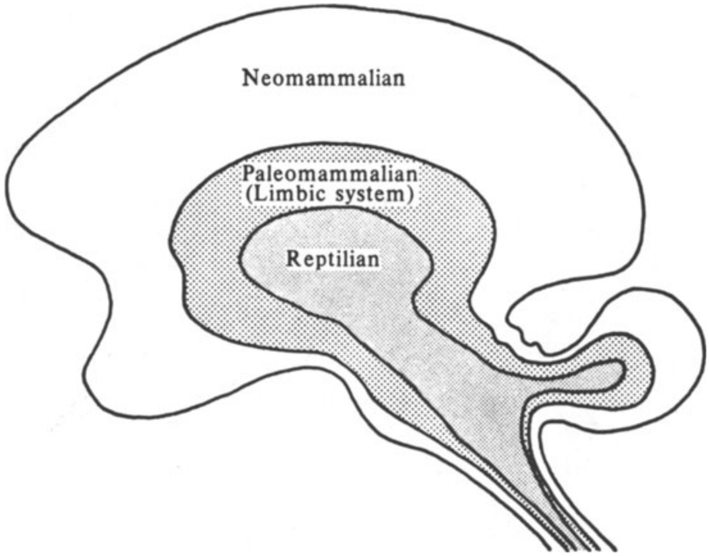

Figure 5. The triune theory of the brain.

which is the cerebral development characteristic of higher mammals and which culminates in man to become the organ mediating reading, writing, and arithmetic.

According to MacLean, the reptilian brain programs stereotype behaviors according to instructions based on ancestral learning and ancestral memories. He suggests that the mammalian counterpart of the reptilian brain is fundamental to such genetically constituted forms of behavior as selecting home sites, establishing territory, engaging in various types of display, hunting, homing, mating, breathing, imprinting, forming social hierarchies, and selecting leaders.

MacLean goes on to point out that the evolutionary development in the lower mammals of a rudimentary cortex appears to represent nature's attempt to provide the reptilian with a better means for adapting to its internal and external environment. In all mammals, most of this primitive cortex is found in a large convolution which Broca called "the great limbic lobe" because it surrounds the brainstem. (*Limbic* means "forming a border around.") This brain, as Papez was the first to show in 1937, plays an important role in elaborating emotional feelings that guide behavior in relation to two basic life principles of self-preservation and the preservation of the species. This limbic system has strong connections with the hypothalamus which plays a basic role in integrating emotional expression. MacLean goes into great detail regarding the structure and function of this system, but it cannot be part of this paper to follow him there. Suffice it to say that his discussion gives ample justification for regarding the limbic system as being concerned with emotion and with conditioning.

The neocortex, of course, is the characteristic mark of man and is concerned with reasoning, language, and number. Because of its great development in man, it has become customary to speak of man as *Homo sapiens,* but of course this is an absurd oversimplification. Man has a triune brain; the three parts are morphologically and functionally differentiated, and communication between them is relatively restricted. The neocortex may tell us how to achieve our wishes and desires, but the source of these is the reptilian brain, and the limbic system. Culture and other social concepts relate almost exclusively to the neocortex, but more primitive forms of behavior, immensely more powerful, relate to the reptilian brain and to the paleocortex. It would be equally wrong to deny the importance of the neocortex as it would be to deny the importance of reptilian brain and the paleocortex; recognizing the importance of all three gives us a picture of man as a biosocial animal, driven by primitive impulses and instincts but also possessing the means of rational adjustment to circumstances, evaluation of rewards and punishments, and powers of rational decision making.

The best example of the relative independence of these different parts of the triune brain is the existence of neurotic disorders. The behavior of neurotic patients is clearly determined by the paleocortex, probably through conditioned emotional reactions of one kind or another, and there is little or no communication with the neocortex. The patient *knows* perfectly well that his phobias, anxieties, and fears are irrational; this knowledge does not help him in the slightest to overcome his difficulties. The whole development of behavior therapy (Kazdin, 1978; Kazdin & Wilson, 1978; Rachman & Wilson, 1981), which has proved immensely more powerful in dealing with neurotic disorders than previous theories and methods, was premised on some such conception of man; insight therapies of the Freudian kind, which have been found seriously wanting, were based essentially on neocortical views of behavior emphasizing the importance of cognitive factors such as insight. Behavior therapy bases itself on laws of conditioning, and it has been shown that concepts such as those of extinction are basic to all forms of therapy insofar as these are successful (Eysenck, 1976a, 1979, 1980). It cannot, of course, be said that these developments prove the correctness of the view of man as a biosocial animal; all that can be claimed is that such a view leads to research which has been found to be eminently useful in this and other fields.

One obvious consequence of regarding man as a biosocial animal, and not looking upon his behavior exclusively from one point of view or the other, is clearly the abandonment of categorical viewpoints and the adoption of a quantitative way of looking at rival claims. We cease to regard behavior as being determined by environment or heredity and adopt rather the viewpoint that clearly all behavior is determined to some extent by both, as are individual differences; the task of science is to assess on a quantitative basis the contribution made by each of these factors. Indeed, analysis must go further and look at different ways of interaction and their contribution to the total variance. Also important is a breakdown of the total genetic variance into additive and nonadditive factors such as dominance, epistasis, and assortative mating. Environmental variance must be broken down into within-family and between-family, and so on. It is interesting to note that such a program of research (which is clearly a "progressive problem shift" in Lakatos's sense, is advocated and implemented by geneticists (Mather & Jinks, 1971). Nothing corresponding to it has been suggested by the environmentalists, who thus demonstrate their one-sidedness, their refusal to look at the total picture, in the most clear-cut fashion of all, that is, by using a methodology clearly inappropriate for the problems it is designed to solve. The conception of man as a biosocial animal, adopted as a guiding principle for research and not as a shibboleth deserving lip-service only, may thus be seen as a vitally important kind of metatheory for psychology (and sociology).

From the practical point of view, too, as has already been mentioned in relation to behavior therapy, such a conception is of extreme importance. Education, industry, government, and all areas of life involving intercourse with people could clearly improve their effectiveness by basing themselves on this principle, conducting research in conformity with it, and acting on the results of such research. To take but one example, consider the work of Hunter and Schmidt (1980), who have looked at methods of estimating the effectiveness of different selection strategies for various jobs. At the moment selection is largely disregarded and judicially frowned upon as leading to discrimination; furthermore, it is clearly a breach of the egalitarian shibboleth. Their analysis is essentially concerned with the probable cost of ignoring the kinds of information that can be provided by psychological tests of intelligence, personality, and so forth for job selection in all spheres of our national economy. They find that employee differences in job proficiency correspond to considerable difference in the actual dollar value of their performance.

Thus in a study of budget analysts, Schmidt and Hunter (1980) estimated that the dollar value productivity of superior performers (top 15%) was 23,000 dollars per year greater than that of low proficiency (bottom 15%). Computer programmers showed a comparable difference. Hunter and Schmidt point out that when these dollar losses are multiplied by the number of employees in an organization and by the number of years they are employed the losses quickly mount into millions of dollars.

In a study of the Philadelphia Police Department with 5,000 employees Hunter (1979) estimated that the abandonment of a general ability test for the selection of police officers would cost a total of $180 million over a 10-year period. The estimated gains in productivity resulting from one year's use of a more valid selection procedure for computer programmers in the federal government range from $5.6 million to $92.2 million for different sets of estimation parameters (Hunter & Schmidt, 1980). For the whole federal government with four million employees Hunter and Schmidt conservatively estimated that optimal selection procedures would save $16 billion per year!

The use of such tests is of course based on the concept of man as a biosocial animal, with genetically determined differences in abilities, personality, and so forth; these differences can be measured with considerable success and predictions made about future behavior on the basis of such tests. Failure to use these tests is based essentially on a picture of man vitally different from that given by modern science, and the resulting loss to society is of a size that should not be tolerated any longer. These figures are a clear indication of the deprivation which society imposes on itself by adopting a radically false picture of man. The concept of man as a biosocial animal may hence be regarded as a psychological metatheory which generates extremely important and testable consequences; it leads us to more specific theories, many of which have been tested and found to be borne out by the facts. Hence the importance of the metatheory and the need for defending it against ideological onslaughts of an unscientific kind.

7. Epilogue

This paper does not fit neatly into one of several categories which might be considered relevant, such as the philosophy of science, the history of science, the sociology of science, or the psychology of science. Having read fairly widely on all these topics, I believe that none of them can properly be studied without reference to the others; the dividing lines are arbitrary, and a proper understanding of science requires all these aspects to be taken into account. As a practicing scientist I have of course been most interested in seeing to what extent the debates, teachings, or pronouncements of these various groups are relevant to my own work and throw some light on the various areas of research in which I have been engaged. This paper is the outcome of such an active interchange between experiences gained in taking part in several scientific "revolutions" (Kuhn) or (hopefully) progressive problem shifts (Lakatos), on the one hand and consideration of the writings of philosophers, sociologists, and historians interested in these various aspects of science on the other.

Inevitably the outcome is somewhat untidy when looked at from the point of view of the professional philosopher, historian, or sociologist. It is my hope, nevertheless,that the paper will add to a better understanding of the place of theory in psychology. Most published works deal with theory in the more advanced sciences, and what is said there is not always relevant to a science which clearly is in a much less advanced stage. It is for this reason that I consider the developmental scheme outlined in Figure 1 to be particularly important and regard the distinction betweek weak and strong theories as potentially useful to research workers in psychology. Many a good theory is cut off in its prime because it does not come up to the standards of strong theories; had it been allowed to develop without having undue demands made upon it, it might have benefitted psychology considerably. Altogether psychology appears to oscillate between two extremes, both of which are equally unacceptable from the scientific point of view. One is the excessive leniency towards theories such as the Freudian; pronouncements along dynamic lines are uncritically accepted, without the usual requirement of any form of acceptable and reproducible proof. This whole area of psychology, which includes projective tests like the Rorschach, interpretations (as of dreams and stories), and therapies (psychotherapy and psychoanalysis), is clearly in a prescientific stage and should not be admitted to the group of scientific disciplines that make up psychology. There are certain minimal requirements which are not fulfilled by these dynamic theories, and until they are it would be wise to exclude theories of this kind from serious consideration.

The other extreme is the narrow exclusiveness often shown by so-called experimental psychologists who believe that strict adherence to relatively arbitrary criteria of methodology is the be-all and end-all of science and suffices to decide the acceptability or otherwise of a given study. Along these lines we have a long series of quite unimportant papers reporting experiments which are technically competent but do not make any valuable addition to psychological theory. Typically what happens is that someone invents a slightly different mode of investigating memory, or motivation, or whatever; this original paper is then followed by large numbers of replications, slight changes, and complex discussions and arguments, without any consideration of whether or not the original paradigm makes any real contribution to psychological theory. If psychology can learn to avoid the excesses of rigorous dogmatism, leading to *rigor mortis*, as well as the delinquencies of overspeculation of a Freudian kind, we may finally succeed in making psychology into a science, rather than, as William James expressed it, the hope of a science.

8. References

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

Baker, J. R. (1974). *Race.* London: Oxford University Press.

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

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

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

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

Bernal, J. P. (1969). *Science in history* (4 vols.). London: C. A. Watts & Co.

Blackmore, J. T. (1972). *Ernst Mach.* Berkeley: University of California Press.

Boudon, R. (1973). *Education, opportunity, and social equality.* London: Wiley.

Caplan, A. C. (Ed.). (1978). *The sociobiology debate.* London: Harper & Row.

Carnap, R. (1962). *Logical foundations of probability.* New York: Basic Books.

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.

Darlington, C. D. (1969). *The evolution of man and society.* London: Allen & Unwin.

Dashiell, J. F. (1938). Some rapprochements in contemporary psychology. *Psychological Bulletin, 36,* 1–24.

Eaves, L. J., & Eysenck, H. J. (1974). Genetics and the development of social attitudes. *Nature, 249,* 288–289.

Eaves, L. J., & Eysenck, H. J. (1975). The nature of extraversion: A genetical analysis. *Journal of Personality and Social Psychology, 32,* 102–112.

Eaves, L. J., & Eysenck, H. J. (1977). Genotype-environmental model for psychoticism. *Advances in Behaviour Research & Therapy, 1,* 5–26.

Eddington, A. (1920). *Space, time and gravitation.* Cambridge: Cambridge University Press.

Eysenck, H. J. (1939). Primary mental abilities. *British Journal of Educational Psychology, 9,* 270–275.

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

Eysenck, H. J. (1957). *Dynamics of anxiety and hysteria.* London: Routledge & Kegan Paul.

Eysenck, H. J. (1959). Serial position effects in nonsense syllable learning as a function of interlist rest pauses. *British Journal of Psychology, 50,* 360–362.

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

Eysenck, H. J. (1970). *The structure of personality.* London: Methuen.

Eysenck, H. J. (1975a). *The inequality of man.* London: Temple Smith.

Eysenck, H. J. (1975b). The measurement of emotion: Psychological parameters and methods. In L. Levi (Ed.), *Emotions—Their parameters and measurement.* New York: Raven Press.

Eysenck, H. J. (1976a). The learning theory model of neurosis—A new approach. *Behaviour Research and Therapy, 14,* 251–267.

Eysenck, H. J. (1976b). *The measurement of personality.* Lancaster: Medical and Technical Publishers.

Eysenck, H. J. (1977). *You and neurosis.* London: Temple Smith.

Eysenck, H. J. (1979). The conditioning model of neurosis. *Behavioral and Brain Sciences, 2,* 155–199.

Eysenck, H. J. (1980). A unified theory of psychotherapy, behaviour therapy and spontaneous remission. *Zeitschrift für Psychologie, 188,* 43–56.

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

Eysenck, H. J. (Ed.). (1983). *A model for intelligence.* New York: Springer.

Eysenck, H. J. (1985). The theory of intelligence and the psychophysiology of cognition. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence.* New York: Academic Press.

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

Eysenck, H. J., & Frith, C. D. (1977). *Reminiscence, motivation and personality.* New York: Plenum Press.

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

Eysenck, H. J., & Wilson, G. D. (1978). *The psychological basis of ideology.* Lancaster: Medical and Technical Publishers. Baltimore: University Park Press.

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

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

Grünbaum, A. (1976). Is falsifiability the touchstone of scientific rationality? Karl Popper versus inductivism. In R. S. Cohen, P. K. Feyerabend, and M. W. Wartofsky (Eds.), *Essays in memory of Imre Lakatos.* Dordrecht: R. Reidel.

Grünbaum, A. (1977). How scientific is psychoanalysis? In R. Stern, L. Horowitz, and J. Lynes (Eds.), *Science and psychotherapy.* New York: Haven Press.

Grünbaum, A. (1979). Is Freudian-psychoanalytic theory pseudo-scientific by Karl Popper's criterion of demarcation? *American Philosophical Quarterly, 16,* 131–141.

Grünbaum, A. (1981). Can psychoanalytic theory be cogently tested "on the couch"? In A. Grünbaum & L. Landon (Eds.), *Pittsburgh series in philosophy and history of science.* San Francisco: University of California Press.

Guthke, J. (1978). *Ist Intelligenz Messbar?* Berlin: Deutscher Verlag der Wissenschaften.

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

Harré, R. (1975). *Problems of scientific revolutions.* Oxford: Clarendon 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. E. (1979). *An analysis of validity, differential validity, test fairness, and utility for the Philadelphia Police Officers Selection Examination prepared by the Educational Testing Service.* Report to the Philadelphia Federal District Court, *Alvaraz* v. *City of Philadelphia.*
Hunter, J. E., & Schmidt, F. L. (1980). Fitting people to jobs: The impact of personnel selection on national productivity. In E. A. Fleishman (Ed.), *Human performance and productivity.* New York: Academic Press.
Jensen, A. R. (1962). Extraversion, neuroticism and social learning. *Acta Psychologica, 20,* 69–77.
Jensen, A. R. (1981a). *Test validity,* g *versus the specificity doctrine.* Invited address, Division 14, presented at the annual convention of the A.P.A., Los Angeles, August 26.
Jensen, A. R. (1981b). Taboo, constraint and responsibility in educational research. *New Horizons,* No. 22, 11–20.
Kamin, L. J. (1974). *The science and politics of I.Q.* London: Wiley.
Kamin, L. J. (1981). In H. J. Eysenck & L. J. Kamin (Eds.), *The intelligence controversy.* New York: Wiley.
Kazdin, A. E. (1978). *History of behavior modification.* Baltimore: University Park Press.
Kazdin, A. E., & Wilson, G. T. (1978). *Evaluation of behavior therapy.* New York: Ballinger.
Krige, J. (1980). *Science, revolution and discontinuity.* Sussex: Harvester Press.
Krutezki, W. A. (1974). Die Entwicklung Leninscher Ideen in der sowjetischen Psychologie der Fähigkeiten. In W. A. Krutezki (Ed.), *Lenins philosophisches Erbe und Ergebnisse der sowjetischen Psychologie.* Berlin: Springer.
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.
Lenin, W. I. (1965). Ein liberaler Professor über die Gleichheit. In W. I. Lenin (Ed.), *Werke* (Bd. 20). Berlin: V.E.B.
Lipovechaya, N. G., Kantonistova, N. C., & Chamaganova, T. G. (1978). The role of heredity and environment in the determination of intellectual functions. *Medicinskie Problemy Formirovanija Livcnosti,* 48–59.
Mackenzie, D. A. (1981). *Statistics in Britain: 1865–1930.* Edinburgh: University of Edinburgh Press.
MacLean, P. D. (1973). A triune concept of the brain and behavior: Including psychology of memory and sleep and dreaming. In T. J. Bord & D. Campbell (Eds.), *Brain and behavior.* Toronto: University of Toronto Press.
Marrow, A. (1969). *The practical theorist.* New York: Basic Books.
Mather, K., & Jinks, J. L. (1971). *Biometrical genetics.* London: Chapman & Hall.
Mehlhorn, G., & Mehlhorn, H. (1981). *Intelligenz.* Berlin: Deutscher Verlag der Wissenschaften.
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.
Putnam, H. (1962). What theories are not. In E. Nagel, P. Suppes & 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., & Wilson, T. (1981). *The effects of psychological therapy.* New York & London: Pergamon Press.

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

Rosenbaum, G. (1953). Stimulus generalization as a function of level of experimentally induced anxiety. *Journal of Experimental Psychology, 45,* 35–43.

Rosenbaum, G. (1956). Stimulus generalization as a function of clinical anxiety. *Journal of Abnormal and Social Psychology, 53,* 281–285.

Savage, R. D., & Eysenck, H. J. (1964). The definition and measurement of emotionality. In H. J. Eysenck (Ed.), *Experiments in motivation.* London: Pergamon Press.

Schmidt, F. L., & Hunter, J. E. (1980). *New research findings in personnel selection: Myths meet realities in the 1980s.* Public Personnel Administration: Policies and Procedures for Personnel. New York: Prentice-Hall.

Spearman, C. (1927). *The abilities of man.* New York: Macmillan.

Spence, K. W. (1944). The nature of theory construction in contemporary psychology. *Psychological Review, 51,* 47–68.

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.

Thorndike, E. L. (1903). *Educational psychology.* New York: Teachers College.

Toulmin, S. (1953). *The philosophy of science.* London: Hutchinson.

Vaucouleurs, G. de (1957). *Discovery of the universe.* London: Faber.

Williams, L. P. (1965). *Michael Faraday.* London: Chapman & Hall.

Wilson, E. O. (1975). *Sociobiology: The new synthesis.* London: Harvard University Press.

Wilson, E. O. (1978). *On human nature.* London: Harvard University Press.