

# Normal Science, Pathological Science and Psychometrics

---

**Joel Michell**

UNIVERSITY OF SYDNEY

**ABSTRACT.** A pathology of science is defined as a two-level breakdown in processes of critical inquiry: first, a hypothesis is accepted without serious attempts being made to test it; and, second, this first-level failure is ignored. Implications of this concept of pathology of science for the Kuhnian concept of normal science are explored. It is then shown that the hypothesis upon which psychometrics stands, the hypothesis that some psychological attributes are quantitative, has never been critically tested. Furthermore, it is shown that psychometrics has avoided investigating this hypothesis through endorsing an anomalous definition of measurement. In this way, the failure to test this key hypothesis is not only ignored but disguised. It is concluded that psychometrics is a pathology of science, and an explanation of this fact is found in the influence of Pythagoreanism upon the development of quantitative psychology.

**KEY WORDS:** measurement, normal science, pathology of science, psychometrics, quantification

*There is no safety in numbers, or in anything else.* (James Thurber)

I argued (Michell, 1997a, 1997b) that quantitative psychology manifests methodological thought disorder, eliciting from Lovie (1997) criticisms quite unlike those offered by others invited to comment on my argument (Kline, 1997; Laming, 1997; Luce, 1997; Morgan, 1997). Lovie follows the post-positivist tradition stemming from Kuhn (1970a) and, from that perspective, saw my approach as a 'hard-nosed (and very outdated) positivist and empiricist/realist line' (Lovie, 1997, p. 393). The view that positivism is a form of empirical realism remains widespread, despite Passmore's (1943, 1944, 1948) early critique and recent analyses (e.g. Friedman, 1991). Hence, there may be value in clarifying my argument regarding pathological forms of science and highlighting my reasons for so categorizing psychometrics.

My thesis is that psychometrists are not only uncritical of an issue basic to their discipline but that, in addition, they have constructed a conception of quantification that disguises this. If science is a cognitive enterprise, then I

maintain this way of doing it is not normal. It is pathological. I begin by describing the concept of a pathology of science. Then I consider concepts of normal science. Finally, the case of psychometrics is analysed in enough detail to identify some of the dynamics involved.

## Pathology of Science

The concept of pathology of science may be unfolded by analogy with that of a pathology of individual cognition. If cognition (in its most general sense) is taking something to be the case when it obtains, then error is taking something to be the case when it does not. Mere error is a failure of cognition, but this, by itself, is not pathological. Circumstances allowing, mere error will be corrected, but error involved in pathology of cognition may not. This distinction is not a dichotomy. There may be intermediate grades: cases of mere error slow to correction and cases of pathology eventually cured. Nonetheless, the extremes are clearly distinguishable. For example, the normal person who, momentarily distracted, mistakenly thinks that they had their usual breakfast this morning instead of remembering that, because of a medical appointment, they fasted will typically correct their error when pressed. However, the person suffering Korsakoff's syndrome typically will not. A pathology of cognition is error caused by a special factor: a relatively permanent condition (e.g. neural damage caused by thiamine deficiency) that not only interferes with the cognition of facts of a certain class, but also hinders correction of these errors. This special factor causes a breakdown not just in cognition itself, but also in processes of error-correction. This causal factor may be the result of direct damage to the cognitive apparatus or it may be the result of conflict between the motivational systems whose interests the cognitive apparatus serves (Freud, 1915/1957).

Science's processes of error-correction reside in its characteristic *modus operandi*. Scientific practices are oriented towards inquiry. Given that 'Nature loves to hide' (as Heraclitus aphorized [Burnet, 1957, p. 133]), that inquiry is burdened by 'the dullness, incompetency and deceptions of the senses' (Bacon, 1620/1960, p. 50), and that even the best of our scientific methods may only afford the 'twilight of probability' (Locke, 1690/1959, p. 360), the optimal form of inquiry in science is critical inquiry. It is rarely the case that scientists' first guesses are their best. Conjectures are sifted. What are now taken to be right answers were attained, in part, through eliminating apparently wrong ones. Scientific inference has a simple analogue in attempts to solve crossword puzzles: 'The clues are the analogues of the subject's experiential evidence; already filled-in entries, the analogues of his reasons' (Haack, 1993, pp. 81–82). A new entry needs to fit both. Perhaps

arrived at by trial and error, it may subsequently prove incorrect. Similarly, research data, always fallible, often indeed very dirty, are taken in conjunction with what the scientist thinks is already known, as providing clues to underlying structures. A picture is constructed (perhaps aspiring to the sorts of ‘virtues’ described by McMullin [1992]) of a system’s character or ways of working. Because these processes of observation and inference are fallible, any claims they have to superiority over other forms of inquiry (e.g. appeals to authority, conformity to established ideas) depend upon processes of error-correction. The method of critical inquiry deals with the possibility of error by attempting to put hypotheses to the test; in the first instance, to the test of logical coherence, and, in the second, the test of empirical adequacy.

Critical inquiry involves two forms of test because in general there are two kinds of error that can be made in conjecturing: logical and empirical. Hypotheses put forward to explain how some system works may be logically defective. For example, because of some incoherence, a hypothesis might not propose something that could ever be part of a workable mechanism. Or hypotheses might be empirically mistaken: a hypothesis might propose a workable mechanism but not the actual process involved in the natural system under investigation. Critical inquiry serves to identify such errors. We can grant that there are no foolproof ways of doing this, while recognizing that inferring total scepticism from mere fallibility uncritically draws too long a bow.

If science’s normal way of working is critical inquiry, then a *pathology* of science will involve some breakdown in that process. A breakdown in critical inquiry occurs when some hypothesis is accepted as true without a serious attempt being made to test it. Processes of critical inquiry break down frequently in science and not all such breakdowns are pathological. Breakdowns are inevitable given the difficulty of doing science, the cognitive limitations of scientists and their methods, and the multiplicity of interests and cross-purposes scientists always bring to any inquiry. Breakdowns are usually an affliction of individual scientists (e.g. scientific fraud) or particular research groups (e.g. exaggerated evaluation of research results) and they rarely become discipline-wide. Even when discipline-wide, they need not be pathological: for example, some hypothesis might be accepted as true, at least provisionally, without a serious attempt being made to test it because it is mistakenly thought not to be an empirical issue or, perhaps, it is not yet known how to test it. A breakdown in critical inquiry only becomes pathological when it includes a higher-order attitude, namely that of ignoring the first-order breakdown. That is, in a pathology of science not only is some hypothesis accepted within the mainstream of a discipline without a serious attempt to test it, but that fact is not acknowledged or, in extreme cases, is disguised.

## Normal Science

If pathology of science involves this kind of two-level breakdown in the processes of critical inquiry within the mainstream of a discipline, then (in relation to this feature) science works *normally* when there is no such breakdown. That is, in normal science a critical attitude is taken towards all matters and, so, if any hypothesis remains untested, then this fact is acknowledged. This concept of normal science is not consonant with that popular since Kuhn (1970a). Kuhn's views have become so much part of the fabric of discussion about science that they do not need detailed exposition here. Suffice it to note that according to Kuhn (1970a, p. 182), normal science is research conducted within the framework of a paradigm, a 'disciplinary matrix', that is, a structured set of guiding beliefs taken for granted by all scientists working within some discipline. It may include theoretical claims and empirical generalizations basic to the discipline and, as well, claims about appropriate methods of observation and data interpretation based upon exemplars of seminal research. According to Kuhn, during a phase of normal science, scientists do not criticize their paradigm. Taking the paradigm as a foundation, they build upon it, doing the sort of research that Kuhn calls 'puzzle-solving' or 'mopping-up operations'. While the paradigm provides criteria for determining what an acceptable research problem is, criteria for acceptable solutions and criteria for standards of evidence, Kuhn insists that the paradigm itself is not an object of criticism within normal science. The scientist acquires the paradigm, says Kuhn (1970a), 'less from the incomplete though sometimes helpful definitions in his text than by observing and participating in the application of these concepts to problem-solving' (p. 47). As a result, normal scientists 'are little better than laymen at characterizing the established bases of their field, its legitimate problems and methods. If they have learned such abstractions at all, they show it mainly through their ability to do successful research' (p. 47).

This rankled Karl Popper, famous for the view that 'science is essentially critical' (1970, p. 53). Kuhn (1970b) replied that critical inquiry is not essential for normal science: 'I hold that in the developed sciences occasions for criticism need not, and by most practitioners ought not, deliberately be sought' (p. 247). I infer from this that Kuhn does not require, as a necessary feature of normal science, that the paradigm be critically scrutinized. All the indications are that Kuhn only thinks that such criticism occurs when science is not normal. Suppose, then, that for some science, elements of the paradigm have never been seriously investigated (having been adopted, say, for ideological reasons), then this failure may go unnoticed by the scientists themselves. I am not aware of anything in Kuhn's writings that suggests he would not still see this as normal science. In my view, however, it would be

a pathology of science. It is instructive to locate the basis of this difference.

My view of normal science is linked to scientists' self-understanding. Scientists see themselves as engaged in finding out how natural systems work: physicists see themselves as discovering the characteristic ways of working of physical systems; biologists, the ways of working of biological systems; psychologists, psychological systems; and sociologists, social. Given the difficulties of this enterprise, science requires mechanisms for the detection of errors, and critical inquiry is the only effective way of doing this (which is not to say that it is always effective). When it works, critical inquiry is effective because nothing is taken to be true without its first having been satisfactorily tested. This applies as much to those propositions that Kuhn has identified as elements of a scientific paradigm as to any others.

Because the motives of individual scientists are always mixed, and because science as a social activity always interacts with diverse social movements and institutions (some hostile to critical inquiry), science's normal way of working may be compromised and critical inquiry may not be the mode within particular disciplines during phases of their history. Thus, while characteristic of science, and in this sense normal, critical inquiry may sometimes be suppressed in the work of individual scientists and may even, more rarely, be muted within entire disciplines. A good example is genetics in the Soviet Union between the 1930s and 1960s (Soyfer, 1994).

This sense of normalcy is not normative. It is entirely descriptive: critical inquiry is science's characteristic way of working, a way that unfolds from its being a cognitive enterprise of fallible knowers. When scientists are uncritical, errors are not sought and the effectiveness of the enterprise is diminished. The more critical scientists are, other things being equal, the more effective is the enterprise.

By a cognitive enterprise, I mean that science is an activity undertaken believing that certain ways of doing things will (at least sometimes) result in knowledge of how natural systems work and that does (at least sometimes) result in knowledge of this sort being attained. Furthermore, by knowledge I mean true belief, and I hold the realist view that a belief is true when and only when things are as believed (Mackie, 1973). Because Kuhn declines to see science as a cognitive enterprise in this straightforward, realist sense, his theory involves a different concept of normal science.

Kuhn's concept unfolds from his view that scientific research is possible only within paradigms. This is a philosophical view and controversial. According to Kuhn (1970a), 'Paradigms are not corrigible by normal science at all' (p. 122). Scientific inquiry into the merits of two 'competing' paradigms is impossible, according to Kuhn. A researcher working outside of paradigms would find that the world is just, 'in William James' phrase, "a bloomin' buzzin' confusion" ' (p. 113). That is, paradigms are our only

means of interpreting the world. According to Kuhn, there is no paradigm-neutral language and so the issue of the truth of any claims made in science only makes sense within a paradigm. Consequently, the concept of truth only has 'intra-theoretic applications' (Kuhn, 1970b, p. 266). This is a position that Kuhn never resiled from (see Kuhn, 1993). In adopting it he echoed (apparently unwittingly) the conventionalism of the positivists, that is, the view that the concept of truth applies only within linguistic frameworks (Carnap, 1950). If truth is relative to paradigms, then so is knowledge, and it follows that science is not a cognitive enterprise in the realist sense. According to such a view, we never know how natural systems work, where natural systems are understood as structures existing independently of us and our paradigms. If Kuhn's picture is correct, then we can never know what is really there in the world, existing independently of us and our paradigms. Kuhn thinks that there is 'no theory-independent way to reconstruct phrases like "really there"' (1970a, p. 206), and reality, independent of our paradigms, is 'ineffable, undescribable, undiscussible' (1991, p. 12).

Friedman (1998) has traced the origins of this sort of view to Kant's (1781/1978) idea that 'We have no insight whatsoever into the intrinsic nature of things' (A277). According to Kant, experience is a construction based partly upon the schemata of the mind and categories of cognition, and so there is no way that we can ever know things as they really are. As Kant put it, 'the order and regularity in the appearances, which we entitle *nature*, we ourselves introduce. We could never find them in appearances, had not we ourselves, or the nature of our mind, originally set them there' (A125). What Kant took to be fixed in 'the nature of our mind', the positivists took to be fixed by convention (Hibberd, 1999). They thought that scientific knowledge is always expressed relative to a linguistic framework and is always, in part, at least, constituted by the conventions defining that framework, in particular, the 'conventions' of logic and mathematics. Likewise Kuhn (1991), who thought of a paradigm as like a lexicon constructing the taxonomic structure of experience, commented that 'like the Kantian categories, the lexicon supplies preconditions of possible experience' (p. 12; see also Sankey, 1997). For them all, Kant, the positivists and Kuhn alike, 'things-in-themselves' are unknowable. They present variations on the same theme because they start from the same question: what sense can be made of science if 'things-in-themselves' are unknowable?

This question cannot be answered by studying science as an empirical phenomenon. Historical and sociological research into science cannot alone bear the weight of philosophical conclusions. Kuhn's conclusion that truth and knowledge are relative to paradigms is philosophical, and philosophical conclusions require philosophical premises. Hence, his argument must contain such premises, if only implicitly, and it is these that lead to his conclusions about the relativity of knowledge to paradigms.

If Kuhn's conclusion is correct and the concepts of truth and knowledge have no application outside of paradigms, then the realist understanding is mistaken. Science is not what it appears and scientists misunderstand themselves. Are we forced to this sceptical conclusion? In its support, the best that anyone can produce is a philosophical argument leading to the conclusion that science is not what it seems and, because of the perpetual inconclusiveness of all philosophical arguments, this leaves us, not with that conclusion established, but with a disjunction: either it is not the case that science is what it seems to be or it is not the case that this philosophical argument is correct. A disjunction, itself, proves neither disjunct and, at best, presents only a choice. Kuhn beckons along the Kantian path, but has not closed the realist gate. While Kuhn takes as his premise the proposition that 'things-in-themselves' are unknowable, scientists hold that they are knowable. From where the scientist stands, one would have to be either exceedingly sceptical or deeply ignorant to seriously doubt that science is now able to tell us the ways in which many natural systems really work. From this standpoint, the Kantian premise is adjudged false and Kuhn's philosophical conclusion likewise.

Anyone accepting the realist view is committed, of course, to a philosophical programme. The following question must be faced: if it is the case that science sometimes tells us the way things really are, then what must the world and our cognitive relation to it be like in their most general features? One feature must be the lack of an ontological divide between the meaning of true propositions and the existence of real situations. Real things are only able to be considered 'in terms of what can be said about them, i.e. in propositions' (Anderson, 1962, p. 4). Hence, 'for a statement to be true is for things to be as they are stated to be' (Mackie, 1973, p. 22). Since scientific discourse is always about either things having properties or things standing in relation to other things, what exists must be situations of these forms. It is required that the world consist of situations, situations possessing the logical structure exhibited in propositions (e.g. Armstrong, 1997). This is the minimal structure for reality if things in themselves are not to be unspeakable. Furthermore, for science to work, it is required that the cognitive relation exists: that is, it is required that humans are sensitive, at least sometimes, to the propositional structure and content of situations. These philosophical claims, when satisfactorily unfolded, constitute the minimal forms of ontological and epistemological realism necessary for scientific knowledge in the realist sense.

If the philosophical position sustaining Kuhn's concept of normal science excludes scientific knowledge in the realist sense, why was his view so popular? The answer is that much of what Kuhn says about paradigms is true. This is illustrated via Haack's (1993) crossword-puzzle analogy. This perspicuous analogy

... immediately suggests a way to come to terms with some Kuhnian themes about the process of scientific inquiry. Normal science can be thought of on the model of working on smaller, non-central entries while taking the correctness of intersecting, already-completed, long, central entries for granted. (Haack, 1997, p. 498)

The 'intersecting, already-completed, long, central entries' are like the paradigm in providing a foundation for subsequent research (i.e. work on 'smaller, non-central entries'), and it is a fact that much scientific research is based upon results and theories taken more or less for granted, at least provisionally. This is an empirical generalization, one supported by the sort of historical evidence Kuhn adduces. The fact that there is good reason to accept this empirical part of Kuhn's theory of paradigms misled some into thinking that the same reasons support his philosophical thesis. However, the fact that paradigms guide research entails no non-realist theories of truth. To sift what is true from Kuhn's theory, it is necessary to distinguish empirical from philosophical components.

Obviously, nothing in Haack's analogy implies that the intersecting, already-completed, long, central entries were initially inserted uncritically, nor that they are treated uncritically when working on smaller entries. At any time, a given scientist generally works on only a small number of problems. This may mean provisionally accepting answers to other questions. Cautious scientists, however, know to what extent their conclusions are so qualified. There is no necessary connection between the existence of paradigms and the holding of uncritical attitudes. Indeed, Haack's analogy helps expose just how pathological it is to ignore criticism of a paradigm. Those who proceed to work on smaller, non-central entries as if already-completed, longer entries are beyond question adopt a pathological crossword strategy.

Haack's analogy shows why some might have thought that the concept of truth applies only intra-theoretically:

Kuhn's thesis of the paradigm-dependence of observation can be reconstrued on the analogy of the way each entry depends, not only on a clue, but also on intersecting entries, i.e., not only on experiential evidence, but also on background beliefs. The thing observed doesn't change, . . . but the judgment he makes of what he sees changes, because of his changed background beliefs. (Haack, 1997, p. 499)

Judging may be paradigm-dependent: the same proposition might be thought to be true given one set of background beliefs and false given a different set. Thus, it might appear as if the concept of truth has only intra-theoretic application. This is an illusion. If background belief,  $p$ , and perception,  $q$ , jointly entail  $r$ , then a person believing  $p$  and seeing  $q$  may adjudge  $r$  true, but this does not entail that  $r$  is only true in a conditional sense. The fact that  $r$  may be true when either  $p$  or  $q$  is false (which it may be providing it does not entail either) shows this. That is, in terms of Haack's analogy, given the insertion of central entry  $p'$  and the presence of clue  $q'$  in the puzzle, smaller

entry  $r'$  might be entailed, but  $r'$  could still be correct even if  $p'$  were not or if  $q'$  were not given. The point is that while  $r'$ 's being judged correct might not be independent of  $p'$  and  $q'$ , what it means for it to be correct is. Applying this to science: while judging may be paradigm-dependent, what is judged (i.e. the independently existing real situation) is not, and so the paradigm-dependence of judging does not entail the paradigm-dependence of truth.

Haack's analogy contains a model of the realist concept of truth. An entry is objectively correct if it is the one stipulated by the puzzle-maker. These stipulations are the analogue of the independently existing facts investigated in science. We can see how different people attempting the puzzle can be led to see different possibilities as correct given differences in already completed entries (i.e. perceptions are guided by the paradigm). Likewise, scientists who accept different presuppositions may appraise the facts of an experiment contrarily, each in a coherent way. But we are not thereby forced to revise our concept of truth any more than with the crossword puzzle we are forced to revise our concept of correctness. At most, only one of the scientists will be seeing things veridically and the issue might be resolved by critical appraisal of their respective paradigms. Then again, the issue might not be so resolved because of difficulties testing the propositions constituting the paradigms, in which case we do not know who is seeing things as they are (if, indeed, either is).

If my analysis is correct and the empirical content of Kuhn's concept of a scientific paradigm is consistent with realism, then we can revise Kuhn's concept of normal science. There is nothing intrinsic to the concept of a paradigm that prevents critical inquiry into the truth of paradigms, no matter how difficult in practice. Likewise, in those chapters in the history of a science when a paradigm is replaced, there is no reason intrinsic to this process that rules out the possibility of critical inquiry. Thus, if normal science is defined as critical science, then Kuhn's distinction between normal and revolutionary science is really only a distinction between phases in the vicissitudes of normal science: research carried out under the provisional assumption of a paradigm and research into the relative merits of competing paradigms, resulting in the (perhaps tentative) replacement of one by its rival.

The normal way to work towards the solution of a crossword puzzle is not just to work away on smaller, non-central entries, taking for granted already-completed, long, central entries. It is also normal to revise longer entries. Viewed through the lens of this analogy, Kuhn's normal science and his revolutionary science are each normal ways of working. Likewise, within science, revolutionary thinking is as much normal science as is research under the auspices of a paradigm. In the crossword puzzle, what would be abnormal would be the behaviour of a person who insists that a particular word (say, the word 'quantitative') should appear in the solution to every

puzzle and inserts it into the first available 12-letter space, ignoring relevant clues and refusing to consider its replacement. Likewise, I claim, with science.

### **Psychometrics: The Symptom**

If a pathology of science involves a two-level breakdown in the processes of critical inquiry within the mainstream of a discipline (i.e. a hypothesis is accepted without a serious attempt being made to test it and this failure of critical inquiry is ignored), then the primary symptom of pathology must be the absence of any serious attempt within the mainstream of a science to test some hypothesis. Consider any attribute that psychometricians currently believe they are able to measure (such as any of the various intellectual abilities, personality traits or social attitudes that the textbooks mention), and ask the question, *Is that attribute quantitative?* The hypothesis that such an attribute is quantitative underwrites the claim to be able to measure it. However, there has never been any serious attempt within psychometrics to test such hypotheses.

The attributes that psychometricians aspire to measure are not directly observable (i.e. claims made about them can only [at present] be tested by first observing something else and making inferences). What psychometricians observe are the responses made to test items. Intellectual abilities, personality traits and social attitudes are theoretical attributes proposed to explain such responses, amongst other things. Typically, test scores are frequencies of some kind, and the hypothesized relations between these theoretical attributes and test scores are taken to be *quantitative* relationships (i.e. functional relationships between quantitative attributes). A good example is the class of factor analytic theories of mental ability, in which people's test scores are hypothesized to relate linearly to products of person and test factor scores, these being understood as measures on hypothetical attributes. It is a necessary part of such quantitative theories that the theoretical abilities hypothesized are understood as *quantitative attributes*. Thus, those accepting such theories as a basis for measurement are committed to the proposition that the psychological attributes theorized about are quantitative in structure.

Psychometricians admit as much in taking their theoretical attributes to be measurable on *interval* or *ratio* scales, as is typically done in modern psychometrics with respect to, say, factor analytic or item response theory approaches to measurement. The view expressed by Lord and Novick (1968) that 'The level of measurement most often specified in mental test theory is *interval measurement*, which yields an *interval scale*' (p. 21) remains the minimal position adopted within mainstream psychometrics. Since an interval scale presumes that at least differences between levels of the relevant

attribute possess quantitative structure, this position implies that the psychological attributes thought to be measured are quantitative.

Hölder (1901; see Michell & Ernst, 1996, 1997) made explicit how quantitative structure involves additivity and that an attribute's being additive is a specific empirical condition. Of any attribute hypothesized to be quantitative, it is always relevant to ask, *Is it really quantitative?* From the scientific point of view, the answer to this question can never be an automatic *yes* because it is an empirical issue. Once this is recognized, it is evident that scientific quantification is a two-stage process (Michell, 1997a). Stage one is the *scientific task of quantification*. This involves devising test situations that are differentially sensitive to the presence or absence of quantitative structure: if the attribute is quantitative, then the outcome of the test goes one way; if not, then it goes another. Stage two is what I call the *instrumental task of quantification*. This involves devising standardized procedures for estimating measures of the attribute involved. Without stage one, no critically minded scientist aware of the logic of quantification would claim to be able to measure the attribute involved. Numbers purporting to be measures of that attribute would only be so contingent upon a hypothesis, the truth of which, in this instance, is not yet tested.

For attributes that psychometricians claim to measure, stage one has never been seriously attempted. This fact is clear, not just because reports of research within journals such as *Psychometrika* or *Applied Psychological Measurement*, for example, contain little on this issue; it is also evident from the fact that syllabuses within psychometrics courses do not consider the issue of how to undertake stage one and, furthermore, from the fact that textbooks on psychometrics likewise neglect this question. This obtains despite the fact that at various points in the history of psychology, this issue has been raised (e.g. Adams, 1931; Boring, 1920; Ramsay, 1991; Reese, 1943; Suppes & Zinnes, 1963). *Prima facie*, this is a symptom of pathology of science.

### **Psychometrics: The Pathology**

From this symptom it does not follow that psychometrics is pathological. It may be that, before now, psychometricians had not penetrated far enough into this problem to recognize the need to complete the scientific task of quantification. Or the task might be recognized, but it may not be known how to complete it in contexts like those found in psychometrics. However, neither of these extenuating circumstances obtains now and has not for many decades.

It might be thought that if psychometricians considered what measurement means, they could reason as follows. Starting from some version of the traditional concept of measurement, such as, 'When we measure in any

department of natural science, we compare a given magnitude with some conventional unit of the same kind, and determine how many times the unit is contained in the magnitude' (Titchener, 1905, p. xix), it follows that any attribute measured must be additive in structure because the concept of how many rests upon that of a sum. Furthermore, because some natural attributes are additive (and, therefore, measurable) and others (e.g. kinship structures, grammatical structures) are not, it follows that there is an *empirical* question here, namely with what kind of attributes do we deal in psychometrics, quantitative or not?

Psychometricians are blocked from this line of reasoning by a false definition of measurement. When they define measurement there is strong uniformity, but the definitions given do not even slightly resemble the traditional concept (Michell, 1997a). They all derive from a definition first given by the psychologist S.S. Stevens in 1946, and repeated by him many times after (Stevens, 1946, 1951, 1958, 1959, 1967, 1968, 1975): *measurement is the assignment of numerals to objects or events according to rule*. This definition entails that all that is involved in measurement is the construction of a rule for assigning numerals to objects or events. Since, in relation to any attribute, be it quantitative or not, one can always locate frequencies of some sort to be counted or contrive other devices for making numerical assignments, 'the hazard of educational and psychological measurement is that almost anyone can devise his or her own set of rules to assign some numbers to some subjects' (Suen, 1990, p. 5). Those accepting Stevens' definition cannot comprehend the scientific task and go directly to the instrumental task.

Are not psychometricians free to define the concept of measurement any way they like? Psychometricians think they are measuring attributes like intellectual abilities, personality traits and social attitudes. These attributes are hypothesized to stand in quantitative relations with test scores of one sort or another and, as indicated, only quantitative attributes stand in quantitative relations. Hence, they are committed to the hypothesis that their attributes are quantitative. Because quantitative attributes are additive in structure, different levels of any such attribute stand in numerical relations to one another (Hölder, 1901). Measurement is just the discovery or estimation of these relations (always involving one particular level conventionally taken as the unit). In accepting Stevens' definition of measurement, psychometricians have adopted one inconsistent with the way they theorize about what it is that they think they are measuring (Michell, 1996).

Thus, in psychometrics we have a situation in which (a) a basic, empirical hypothesis (namely the hypothesis that psychological attributes are quantitative) is accepted as true without it ever having been seriously tested for its empirical adequacy, and (b) the fact that this hypothesis has never been satisfactorily tested is disguised. To return to Haack's analogy, it is as if psychometricians have not only inserted the long, central entry 'quantitative'

in the first available 12-letter slot, it is also as if they have pasted over the relevant clue another of their own invention, one implying that 'quantitative' is correct.

### Psychometrics: Case History of a Pathology of Science

All scientific work has social causes and conditions. However, to explain a pathology of science, special causes must be invoked. If the pathology involves some form of secondary gain for the discipline involved, linked to non-scientific interests, then the pathology is explained when this secondary gain is identified and these interests exposed. (My analysis is loosely based upon Michell, 1997a and 1999.)

When, in the 19th century, quantitative psychology finally emerged as an independent discipline, it was already grossly deformed by ideological pressures issuing from the Scientific Revolution of the 17th century. One important feature of this revolution was the emphasis upon quantification (Crombie, 1994). Aristotelian physics was qualitative and undervalued measurement; in the new science of Kepler, Galileo, Harvey and Newton, measurement ruled. This transformation of qualitative into quantitative theories was a triumph for Pythagoreanism.

The Pythagorean idea that nature is fundamentally quantitative in structure profoundly influenced Western thought. Surprisingly, the version that resurfaced during the Scientific Revolution excluded psychological phenomena. Why? And why did it take so long for psychologists to wish to be included? Work exemplifying a quantitative approach to physiology (e.g. by Descartes and Harvey) formed part of the Scientific Revolution itself. Also, Cohen (1994) has shown how social philosophers used quantitative theories of Galileo and Harvey as models for quantitative speculations about society as early as the 17th century. However, although philosophers like Locke and Hume found inspiration in Newton's physics (Hume [1739/1960, p. 12], for example, likening the association of ideas to gravitational attraction), their psychological speculations were not quantitative. On the surface, it seems that circumstances could hardly have been more favourable for the development of quantitative psychology. Some special factor must have delayed it. This impression is strengthened when the history of Pythagoreanism is considered.

Pythagoras, it is said (Hussey, 1997), founded a philosophical movement in Italy around the 6th century BC. He taught that mathematics reveals the underlying structure of reality. Aristotle, the earliest author to give an account of Pythagorean doctrines (Guthrie, 1962), claimed that it was in numbers that Pythagoreans found the principles explaining 'all things', 'such and such a modification of numbers being justice, another being soul and reason' (Aristotle, *Metaphysics* [1941a], 985b 28–30). Aristotle's reference

to soul and reason indicates how thoroughgoing the Pythagoreans were: numerical principles were thought to explain psychological as well as physical phenomena. Aristotle had little time for their doctrines, however, declaring of their thesis that the soul is a 'self-moving number' that it was, 'of all the opinions we have enumerated, by far the most unreasonable' (*De Anima* [1941b], 408b 33). This it may be, but the view that mathematics provides the principles by which all things can be understood is plausible. If mathematics is the science of structure (Parsons, 1990; Resnick, 1997), then it provides the resources for conceptualizing naturally occurring structures, even if which structures actually occur must be discovered observationally. The Pythagoreans were mistaken in their exclusive emphasis upon quantitative structures as opposed to other kinds of mathematical structures. Of course, throughout most of the history of science, because it was thought that 'mathematics is the science of quantity' (Kant, 1764/1992, p. 280), and because the investigation of non-quantitative structures did not really begin until the 19th century, this mistake cast a long shadow.

Pythagoreanism's influence was exerted through the philosophy of Plato, who, in his *Timaeus* (Plato, 1971), reduced the 'basic elements' of earth, fire, air and water to quantitative structures. Plato's Pythagoreanism included psychological phenomena: in his *Protagoras* (Plato, 1956) he considered measuring intensities of pleasure and pain. Pythagoreanism remained strong in late Antiquity (O'Meara, 1989) and was not extinguished during the period of Aristotelian dominance during the Middle Ages. While, within Aristotelian philosophy, the category of quantity was just one amongst many, and certainly not the most important, Pythagoreanism surfaced in controversies about the relation between quality and quantity (Grant, 1996; Sylla, 1972). Interestingly, these controversies arose when the 12th-century scholar Peter Lombard 'put the question whether the theological virtue of charity could increase and decrease in an individual' (Crombie, 1994, p. 410). One solution was to conceptualize such qualitative change as quantitative (i.e. to interpret the category of quality using that of intensive quantity). Nicole Oresme (see Clagget, 1968) conceptualized intensive quantities (including such psychological attributes as pleasure and pain) by analogy with length and claimed them measurable. Given the inclusion of the psychological within ancient and medieval Pythagoreanism, its exclusion from the much more triumphant Pythagoreanism of the Scientific Revolution looks even more puzzling.

The answer lies in the different character of the new science. Unlike their medieval counterparts, the natural philosophers of the Scientific Revolution actually measured. Crombie (1990) notes of the medieval Pythagoreans that 'a far greater need was felt that concepts and theoretical and mathematical procedures should be quantified than that actual measurements should be made' (p. 74). The practice of measurement had been more important in ancient science (Lloyd, 1987), but it was never then the engine for

Pythagorean speculations. However, the practice of measurement was a key ingredient driving the success of the Scientific Revolution, and this success shaped the new Pythagoreanism. Pythagoreanism was now not just a philosophical vision, it was a licence to construct a practical, quantitative, experimental science. The scientists of the 17th century measured what they could, attempted to make measurable what they could not, and what they could not make measurable, they doubted the reality of. Attributes found to be measurable they thought of as *primary qualities*. The remainder they called *secondary qualities*. It was the distinction between the measured and the unmeasured that formed the basis of Descartes' division between matter and mind (Buroker, 1991).

While Descartes endorsed Pythagoreanism, recognizing 'no matter in corporeal things apart from that which the geometers call quantity' (1644/1985, p. 247), his category of quantity extended no further. Descartes' view was that secondary qualities (or, as he termed them, *sensible qualities*), such as colours, do not really exist in the physical objects they appear to adhere to. Instead, he thought, they occur only in the mind when the brain is stimulated in certain ways. Furthermore, their features were held to be obscure and confused. It was thought that they could be ordered, but order alone is not sufficient for quantity and measure. Descartes recognized that such qualities are partially correlated with the corporeal attributes stimulating our brains, but concluded that this relationship could not be described mathematically because sensible qualities cannot be measured (Buroker, 1991). The operational distinction, based in measurement, between the so-called 'primary' and 'secondary' qualities was transformed by Descartes into a metaphysical distinction between separate realms of being, those of body and mind. Mental phenomena were excluded from science because they were excluded from quantity.

The influence of Cartesianism upon the education of successive generations of scientists should not be underestimated (Gascoigne, 1990). It had a debilitating effect upon psychology generally, but its influence upon the prospects of quantitative psychology were especially so. Overcoming the Cartesian exclusion of psychology from the realm of quantitative science required overcoming Cartesian dualism as well as constructing methods that could plausibly be thought of as mental measurement. Fechner overcame the obstacle of Cartesian-style Pythagoreanism because he presented a credible philosophical alternative. The distinction between mind and body, which Descartes had interpreted as ontological, Fechner saw as simply cognitive. According to Fechner, there is just one order of being, which we cognize in two ways: by introspection and by sensory observation. When by the former, we call the objects mental; when by the latter, physical. Thus, he wrote,

We count as mental, psychological, or belonging to the soul, all that can be grasped by introspective observation or that can be abstracted from it; as

bodily, corporeal, physical, or material, all that can be grasped by observation from the outside or abstracted from it. (Fechner, 1860/1966, p. 7)

Similar views were endorsed by Wundt (1896/1907) and James (1890). Adopting such a view, it can be argued (Heidelberger, 1994) that the attributes of mental phenomena must also be quantitative because they are identical to some of the features of the brain, the latter features being quantitative because the brain is a physical structure. In Fechner's opinion, the fact that they are perceived from the 'inside' meant that mental phenomena could not be measured directly, and so Fechner had to employ his various indirect psychophysical methods (such as the method of just-noticeable-differences) to measure them. However, by the 19th century, the use of indirect methods of measurement was familiar enough and measurement was no longer restricted to primary qualities. Also, the fact that Fechner's famous law linked the intensities of sensations to measurable, physical attributes, in much the same way as, say, thermometric attributes had come to be linked to other previously established physical quantities, gave psychophysics the flavour of being yet another extension of quantitative science to previously unquantified attributes. Fechner's achievement promised to restore the original vision of Pythagoreanism and Fechner thought he had shown that the mental, like the physical, was subordinate to the mathematical.

Given the influence of Cartesian philosophy, Fechner's achievements encountered opposition (e.g. Kries [1882] and Bergson [1887/1913]). However, having constructed methods of measurement and overcome Cartesian philosophy, at least to his own satisfaction, Fechner (1887/1987) treated objections as 'mere writing in the sand' (p. 215). And Fechner's methods were accepted by most of the founders of modern psychology as methods of psychological measurement, even if they disagreed about exactly what was measured (Fullerton & Cattell, 1892; Titchener, 1905). Against the backdrop of the Cartesian exclusion, it was Fechner's methods, as methods for measuring something psychological, that were seen as important. Exactly what was measured was seen as a matter for negotiation.

In one respect, Fechner's approach was sound. If psychology is to be treated scientifically, then psychological phenomena must be thought of as subject to the same categories as physical phenomena. The flaw in Fechner's reasoning was this: there is no necessity that all natural attributes must be quantitative. This flaw, however, was not easily discerned at the time. The progress of quantitative physics over the preceding two centuries had made it seem as if 'the extension of science from time to time is correspondent to the discovery of fresh measurable elements in nature' (Venn, 1889, p. 433), and the writings of the founders of modern psychology resonate with this theme. There was an advantage to the new science in not looking too closely at the logic of Pythagoreanism. It was born into a cultural milieu in which

science was strongly identified with measurement and the values of quantitative precision (Porter, 1995; Wise, 1995). As a new discipline, psychology was more easily absorbed by the scientific and wider academic community because it appeared to confirm the prevailing quantitative paradigm. These two factors, the widespread acceptance of Pythagoreanism and the advantage of being able to claim to be already quantitative, might explain why the psychologists of the late 19th and early 20th centuries (a) did not seriously investigate empirically the issue of whether psychological attributes are quantitative, and (b) ignored this failure.

These failures affected psychology's development when interest expanded to psychometrics. Many of those pioneering this interest (e.g. Spearman, James McKeen Cattell) were trained in psychophysics and accepted the Pythagorean vision. For psychometricians, that vision was encapsulated in Thorndike's (1918) credo, 'Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quantity' (p. 16). Psychological tests, as methods delivering numerical data, could, when taken in conjunction with Pythagoreanism, be thought of as solving the instrumental problem of quantification. Thorndike's one-time student, Kelley (1929), thought that

Our mental tests measure something, we may or may not care what, but it is something which it is to our advantage to measure, for it augments our knowledge of what people can be counted on to do in the future. The measuring device as a measure of something that it is desirable to measure comes first, and what it is a measure of comes second. (p. 86)

As Pythagoreans, psychometricians accepted it as an article of faith that psychological tests must measure something, even if it was not known what. Apparently absolved from obligations towards investigating the scientific task, this conviction left psychometricians free to develop their discipline in two other directions: (a) the articulation of statistical theories about the properties and components of test scores; and (b) the stipulation of conventions thought to identify the attributes measured by tests.

The second matter of interest in Kelley's comments was his emphasis upon using tests for prediction. The application of psychological tests in education, industry and the military was the main avenue through which psychometrics became a profession. Furthermore, the usefulness of psychological tests provided an excellent opportunity for secondary gains to accrue via the rhetoric of measurement. The usefulness of tests in prediction is not dependent on them being measures of anything. As Comrey (1951) and Cronbach and Gleser (1957) later recognized, predictive usefulness simply depends upon actuarial relationships between test scores and criteria. However, in a context in which tests are already believed to be instruments of measurement, applications will be discussed in measurement terms. Such rhetoric, employed within social contexts already valuing measurement,

could only improve the prospects of tests being accepted socially (Brown, 1991, 1992). In this way, the standing of psychology as a science and as a profession was enhanced. These gains would have been threatened had the scientific task of quantification been publicly recognized, so it is not surprising that such criticisms as were made of the rhetoric of measurement in psychometrics (e.g. Johnson, 1936) were generally ignored.

While the scientific task of quantification is an integral component of the logic of quantification, this fact was not often recognized prior to Helmholtz (1887) (a paper unpublished in English translation before 1971). However, the issue was more often discussed, at least amongst philosophers of science, following Campbell (1920). This increased the likelihood of psychologists being forced to confront this task.

When in 1932, the British Association for the Advancement of Science appointed Campbell to a committee to report upon the possibility of measuring intensities of sensations, it was inevitable that the failure of psychologists to investigate the scientific task of quantification would be exposed. The reports of this committee (Ferguson et al., 1938, 1940) also revealed the reluctance of psychologists to acknowledge this failure. By 1940, however, it was difficult to side-step these criticisms. Progress in the mathematics of non-quantitative structures meant that Pythagoreanism, as an ideological defence against them, had passed its use-by-date. If the failure of psychologists to investigate the scientific task of quantification was to remain hidden now, a disguise was needed.

Stevens' (1946) paper went half-way to constructing one. Stevens packaged his new definition of measurement, not as new, not even as his, but as Campbell's definition. Then he argued that the theory of scales of measurement (his, now familiar, theory of nominal, ordinal, interval and ratio scales) followed from this definition. The implication was that Campbell misunderstood his own definition, that it included psychophysical measurement, and that, therefore, Campbell had wasted the Ferguson Committee's time attempting to show that psychophysics is not scientific measurement. Campbell's own definition, Stevens argued, entailed the contrary.

By the mid-1950s, this new definition was widely accepted within psychometrics (e.g. Green, 1954; Guilford, 1954; Lorge, 1951). During the previous decade, Stevens had presented himself as an expert on measurement theory to emerging opinion leaders within the coming generation of psychologists (Benjamin, 1977), exposing them to his new ideas. This smoothed the way for its rapid acceptance. However, Stevens' definition alone was not a sufficient disguise. While it is true that non-quantitative structures can be coded numerically (giving rise to Stevens' nominal and ordinal scales), measurement in the sense claimed within psychophysics and psychometrics was measurement in the traditional sense (i.e. interval and ratio scales). The scientific task of quantification was, apparently, not

avoided by accepting this definition. Stevens' solution was to combine his definition with operationism.

According to operationism, the meaning of a concept is the set of operations used to specify it (Bridgman, 1927). Stevens (1935a, 1935b, 1936) was an early advocate for operationism in psychology. Applying operationism to his definition means that what is measured is defined by the rules for making the relevant numerical assignments. Where these rules involved ratios (as with Stevens' [1956] psychophysical method of magnitude estimation), ratio scales were consequently assumed. Psychometricians could not use this manoeuvre, however, because they counted responses and, apparently, had no ratios. However, test scores, being frequencies, are quantitative (which, of course, is not to say that they are measures of anything). If some theoretical, psychological attribute (say, general intellectual ability or extraversion) is said to be operationally defined by scores on some test (or, more complexly, by some mathematical function of test scores, as might be obtained, say, by factor analysis), then that concept is 'operationalized' quantitatively. By operationist logic, the numbers so assigned can be regarded as measures of that attribute. This kind of ploy was, as Kerlinger (1979) later noted, 'a radically different way of thinking and operating, a way that has revolutionized behavioral research' (p. 41). It does this by stipulating that the theoretical attribute is quantitative and that this attribute is quantitatively related to the relevant test scores. Since these issues are empirical, stipulation here substitutes for scientific investigation, a substitution disguised by the doctrine of operationism combined with Stevens' definition of measurement.

During the period in which Stevens' definition and operationism came to be accepted in psychology, no one had a clear idea of how to address the scientific task of quantification in psychometric contexts. In these new contexts, this task presented new challenges to quantitative science, challenges that no one yet knew how to attack. Because of this, it could be argued that the scientific task is not ignored, only deferred, and psychometrics is therefore not a pathology of science. However, if only deferred, then discovering how to investigate the matter would overcome this obstacle; but, if ignored, then such discovery would likewise be ignored.

Hitherto, the conventional wisdom deriving from Helmholtz (1887/1971) and Campbell (1920) was that the quantitative structure of an attribute could only be tested for in one of two ways: (a) fundamental measurement via fairly direct reflections of additivity in the concatenation of objects or events possessing levels of the attribute involved, as with extensive attributes like length or time; or (b) derived measurement through the discovery of system-dependent parameters, as with quantities like density. A leap beyond this understanding was made when Luce and Tukey (1964) proved that additive structure was sometimes testable via the detection of ordinal or equivalence relations.

The details of the theory of conjoint measurement are beyond the scope of this paper (see Michell, 1990, or Narens & Luce, 1986, for accessible expositions). Conjoint measurement theory applies in circumstances where a dependent attribute is a function of two independent attributes (e.g. performance on a test might be a function of both ability and motivation), but where nothing more than ordinal structure in any of these attributes can be detected directly. A key condition, diagnostic of hidden quantitative structure, is called *double cancellation*. A special case of double cancellation is the *Thomsen condition*, and I will use it to give an indication of how conjoint measurement works. The Thomsen condition is a generalization of the Euclidean axiom that in quantitative structures, equals plus equals gives equals.

Equal differences *within* the independent factors can be assessed via trade-offs between their effects upon the dependent factor. For example, suppose that performance on intellectual tasks of a specific kind improves as each of ability and motivation increase and that these two attributes combine in a non-interactive (e.g. additive) way to determine performance. Consider the performance of three people, *A*, *B* and *C*, who differ in ability (with, say,  $A < B < C$ ), upon sets of equivalent tasks under three different levels of motivation (say, level 1 < level 2 < level 3), and suppose that conditions are controlled such that no other factor contributes to differences in performance. If we are able to classify different performances as equally good (same) or not (different), then the claim that ability, motivation and performance are quantitative can be tested as follows. If *A*'s performance at motivation level 2 equals that of *B* at motivation level 1, then the positive difference between *B* and *A* in ability compensates for the negative difference between them in motivation. That is, in terms of effects upon performance,

$$\text{Motivation}_2 - \text{Motivation}_1 = \text{Ability}_B - \text{Ability}_A.$$

Similarly, if *C* at motivation level 2 performs the same as *B* at motivation level 3, then

$$\text{Motivation}_3 - \text{Motivation}_2 = \text{Ability}_C - \text{Ability}_B.$$

If motivation and ability are quantitative, then because equals plus equals give equals,

$$(\text{Motivation}_3 - \text{Motivation}_2) + (\text{Motivation}_2 - \text{Motivation}_1) = (\text{Ability}_C - \text{Ability}_B) + (\text{Ability}_B - \text{Ability}_A),$$

which, after simplifying, may be expressed as

$$\text{Motivation}_3 - \text{Motivation}_1 = \text{Ability}_C - \text{Ability}_A.$$

This is a new relationship, one following because quantity is hypothesized. If this hypothesis is true, then it is predicted that *A* at motivation level 3 will perform as *C* at motivation level 1. If this prediction is confirmed, then this

supports the hypothesis that ability and motivation are quantitative; if infirmed, then not.

Within psychometrics, little notice has been taken of this work since its publication, a fact noted by Cliff (1992). For example, Carroll (1982), in a review coordinating attempts to measure intelligence with developments in measurement theory, neglects to mention it. It has never been incorporated into the typical syllabus of courses on psychometrics, nor into relevant textbooks (e.g. Suen, 1990). Recent proposals for 'revitalizing the measurement curriculum' in psychology (Meier, 1993) also ignore it. Significantly, the appearance of this theory produced no recognition within mainstream psychometrics that the scientific task of quantification exists. As expounded in textbooks (e.g. Suen, 1990), the logic of quantification remains disguised.

This neglect is not merely a technical omission, one which, if attended to, would inevitably confirm existing quantitative theories. The hypothesis that exclusively quantitative mechanisms sustain differences in performance on psychological tests is, *a priori*, no more plausible than non-quantitative alternatives. Psychometric theories apply to test scores, but test scores supervene upon response patterns. A person's response pattern is the pattern of correct and incorrect responses obtained on a test. Without counting test scores, relations already exist between response patterns. For example, person *i* does better on the test than does person *j* if *i* correctly answers all questions answered correctly by *j* and more. Such a relation is transitive and asymmetric, but not generally connected, and so it does not yield a simple ordering of people, let alone a quantitative arrangement. Response patterns are more fundamental than test scores, for two reasons. First, test scores are always able to be deduced from response patterns, but not vice versa. Second, the underlying psychological processes cause the response pattern, the total score simply being one property amongst many that such patterns have. Hence, the fundamental structure with which psychometricians deal is the structure manifest in response patterns (a non-quantitative structure, specifically, a partial order) and not a quantitative structure (test scores). While theories postulating the existence of quantities underlying such partial orders might seem plausible to the Pythagorean mind, they are really no more plausible than theories postulating underlying non-quantitative structures (e.g. partially ordered levels of ability) would be.

Add to that the fact that intellectual performance always involves cognition, an apparently non-quantitative relation, and it cannot yet be ruled out that differences in intellectual performance come about because of a confluence of quantitative and non-quantitative causal processes. Moreover, while physics benefited from a revolution in which quantitative theories replaced qualitative ones, it cannot yet be ruled out that psychology will not benefit from a revolution in which qualitative theories succeed quantitative ones.

## Discussion

I have argued that when the early quantitative psychologists claimed to measure, they used the term 'measurement' for political reasons, not for scientific reasons. They presented psychology as a quantitative science. The secondary gain was admission into the scientific community. When psychometricians claimed to be able to measure, they used the term 'measurement' not just for political reasons but also for commercial ones. Again, there was no valid scientific reason. Psychometrics was presented as an applied, quantitative science. The secondary gain was the package of economic and social rewards reserved by society for applied scientists. Later, in order to preserve these gains, both the science and the profession adopted the operationist conception of quantification because it disguised this lack of scientific reasons. In the socio-historical context in which psychometrics developed, the economic, social and political costs of abandoning the rhetoric of measurement outweighed the scientific costs of abandoning the method of critical inquiry.

This is a fact worth identifying. If science is a cognitive enterprise, then pathologies of science work to subvert that enterprise. Those who support scientific research economically, socially and politically have a manifest interest in knowing that the scientists they support work to advance science, not subvert it. And those whose lives are affected by the application of what are claimed to be 'scientific findings' also have an interest in knowing that these 'findings' have been seriously investigated and are supported by evidence. Yet, in recent decades, many scholars who have turned their attention to the history, sociology and philosophy of science have done so utilizing concepts of science that, like Kuhn's, decline to understand science as a cognitive enterprise in the straightforward, realist sense. As Lovie (1997) has noted, those who start from such premises produce histories different to the kind reported here. The category of pathology of science is not one that they will identify because they decline to criticize science by identifying error. It may be the case that 'Most critics of science happen to be scientists, and I think they are far better placed to do that critical job than historians, sociologists, or philosophers' (Shapin, 1996, p. 165), but unless historians, sociologists and philosophers are also prepared to share the critical job and help identify scientific errors, they will not understand science for the cognitive enterprise that it is.

Attempting to understand cognitive enterprises while denying the concept of error is a heroic undertaking. One can admire those assuming this challenge, as one admires all who courageously face impossible odds, like Scott's expedition hauling their sleds to the South Pole. However, Scott's misjudgement was not an error of logic. Those who would understand science bereft of the concept of error err logically. They are like those who would understand perception without the concept of illusion. We know only

too well, from lived-experience, the fallibility of perception. However, the point is a logical one: if perception can be veridical, then it can be non-veridical also, and so an adequate account of perceptual systems must address the limits of their veridicality. Likewise, I claim, investigating how science works as a cognitive enterprise, ignoring the fact that error, as a possibility, always attends the occurrence of cognition, embraces failure at the start.

## References

Adams, H.F. (1931). Measurement in psychology. *Journal of Applied Psychology*, 15, 545–554.

Anderson, J. (1962). *Studies in empirical philosophy*. Sydney: Angus & Robertson.

Aristotle. (1941a). Metaphysics (W.D. Ross, Trans.). In R. McKeon (Ed.), *The basic works of Aristotle* (pp. 682–926). New York: Random House.

Aristotle. (1941b). De Anima (J.A. Smith, Trans.). In R. McKeon (Ed.), *The basic works of Aristotle* (pp. 534–603). New York: Random House.

Armstrong, D.M. (1997). *A world of states of affairs*. Cambridge: Cambridge University Press.

Bacon, F. (1960). *New organon, Vol. 1* (F. Anderson, Ed.; J. Spedding, Trans.). New York: Bobbs-Merrill. (Original work published 1620.)

Benjamin, L.T. (1977). The Psychology Round Table: Revolution of 1936. *American Psychologist*, 32, 542–549.

Bergson, H. (1913). *Time and free will* (F.L. Pogson, Trans.). London: George Allen & Co. (Original work published 1887.)

Boring, E.G. (1920). The logic of the normal law of error in mental measurement. *American Journal of Psychology*, 31, 1–33.

Bridgman, P.W. (1927). *The logic of modern physics*. New York: Macmillan.

Brown, J. (1991). Mental measurements and the rhetorical force of numbers. In J. Brown & D.K. van Keuren (Eds.), *The estate of social knowledge* (pp. 134–152). Baltimore, MD: Johns Hopkins University Press.

Brown, J. (1992). *The definition of a profession: The authority of metaphor in the history of intelligence testing, 1890–1930*. Princeton, NJ: Princeton University Press.

Burnet, J. (1957). *Early Greek philosophy*. London: Macmillan.

Buroker, J.V. (1991). Descartes on sensible qualities. *Journal of the History of Philosophy*, 29, 585–611.

Campbell, N.R. (1920). *Physics, the elements*. Cambridge: Cambridge University Press.

Carnap, R. (1950). Empiricism, semantics and ontology. *Revue Internationale de Philosophie*, 4, 20–40.

Carroll, J.B. (1982). The measurement of intelligence. In R.J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 29–120). Cambridge: Cambridge University Press.

Clagget, M. (1968). *Nicole Oresme and the medieval geometry of qualities and motions*. Madison: University of Wisconsin Press.

Cliff, N. (1992). Abstract measurement theory and the revolution that never happened. *Psychological Science*, 3, 186–190.

Cohen, I.B. (1994). The scientific revolution and the social sciences. In I.B. Cohen (Ed.), *The natural sciences and the social sciences* (pp. 153–203). Dordrecht: Kluwer.

Comrey, A.L. (1951). Mental testing and the logic of measurement. *Educational and Psychological Measurement*, 11, 323–334.

Crombie, A.C. (1990). *Science, optics and music in medieval and early modern thought*. London: Hambledon Press.

Crombie, A.C. (1994). *Styles of scientific thinking in the European tradition, Vol. 1*. London: Duckworth.

Cronbach, L.J., & Gleser, G.C. (1957). *Psychological tests and personnel decisions*. Urbana: University of Illinois Press.

Descartes, R. (1985). *Principles of philosophy*. J. Cottingham, R. Stoothoff, & D. Murdoch (Eds.), *The philosophical works of Descartes, Vol. 1*. Cambridge University Press. (Original work published 1644.)

Fechner, G.T. (1966). *Elements of psychophysics, Vol. 1* (D.H. Howes & E.G. Boring, Eds.; H.E. Adler, Trans.). New York: Rinehart & Winston. (Original work published 1860.)

Fechner, G.T. (1987). My own viewpoint on mental measurement. *Psychological Research*, 49, 213–219. (Original work published 1887.)

Ferguson, A., Myers, C.S., Bartlett, R.J., Banister, H., Bartlett, F.C., Brown, W., Campbell, N.R., Craik, K.J.W., Drever, J., Guild, J., Houstoun, R.A., Irwin, J.O., Kaye, G.W.C., Philpott, S.J.F., Richardson, L.F., Shaxby, J.H., Smith, T., Thouless, R.H., & Tucker, W.S. (1940). Quantitative estimates of sensory events: Final report of the committee appointed to consider and report upon the possibility of quantitative estimates of sensory events. *Advancement of Science*, 1, 331–349.

Ferguson, A., Myers, C.S., Bartlett, R.J., Banister, H., Bartlett, F.C., Brown, W., Campbell, N.R., Drever, J., Guild, J., Houstoun, R.A., Irwin, J.C., Kaye, G.W.C., Philpott, S.J.F., Richardson, L.F., Shaxby, J.H., Smith, T., Thouless, R.H., & Tucker, W.S. (1938). Quantitative estimates of sensory events: Interim report of the committee appointed to consider and report upon the possibility of quantitative estimates of sensory events. *British Association for the Advancement of Science*, 108, 277–334.

Freud, S. (1957). Instincts and their vicissitudes. In J. Strachey (Ed. and Trans.), *The standard edition of the complete psychological works of Sigmund Freud, Vol. 14* (pp. 109–140). London: Hogarth. (Original work published 1915.)

Friedman, M. (1991). The re-evaluation of logical positivism. *Journal of Philosophy*, 88, 505–519.

Friedman, M. (1998). On the sociology of scientific knowledge and its philosophical agenda. *Studies in History and Philosophy of Science*, 29, 239–271.

Fullerton, G.S., & Cattell, J.McK. (1892). *On the perception of small differences*. Philadelphia: University of Pennsylvania Press.

Gascoigne, J. (1990). A reappraisal of the role of the universities in the Scientific Revolution. In D.C. Lindberg & R.S. Westman (Eds.), *Reappraisals of the Scientific Revolution* (pp. 207–260). Cambridge: Cambridge University Press.

Grant, E. (1996). *The foundations of modern science in the Middle Ages*. Cambridge: Cambridge University Press.

Green, B.F. (1954). Attitude measurement. In G. Lindzey (Ed.), *Handbook of social psychology*, Vol. 1 (pp. 139–150). Reading, MA: McGraw-Hill.

Guilford, J.P. (1954). *Psychometric methods*. New York: McGraw-Hill.

Guthrie, W.K.C. (1962). *A history of Greek philosophy*, Vol. 1: *The earlier Presocratics and the Pythagoreans*. Cambridge: Cambridge University Press.

Haack, S. (1993). *Evidence and inquiry: Towards reconstruction in epistemology*. Oxford: Blackwell.

Haack, S. (1997). The puzzle of 'scientific method'. *Revue Internationale de Philosophie*, 202, 495–505.

Heidelberger, M. (1994). The unity of nature and mind: Gustav Theodor Fechner's non-reductive materialism. In S. Poggi & M. Bossi (Eds.), *Romanticism in science* (pp. 215–236). Dordrecht: Kluwer.

Helmholtz, H. von. (1971). An epistemological analysis of counting and measurement. In R. Kahl (Ed. and Trans.), *Selected writings of Hermann von Helmholtz* (pp. 437–465). Middletown, CT: Wesleyan University Press. (Original work published 1887.)

Hibberd, F. (1999). *Social constructionism, logical positivism and the continuity of error. Part 3: Phenomenalism and its analogue*. Manuscript in preparation. Department of Psychology, University of Sydney.

Hölder, O. (1901). Die Axiome der Quantität und die Lehre vom Mass. *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse*, 53, 1–46. (Translation in Michell & Ernst, 1996, 1997.)

Hume, D. (1960). *A treatise of human nature*. Oxford: Clarendon Press. (Original work published 1739.)

Hussey, E. (1997). Pythagoreans and Eleatics. In C.C.W. Taylor (Ed.), *From the beginning to Plato* (pp. 128–174). London: Routledge.

James, W. (1890). *Principles of psychology*. New York: Holt, Rinehart & Winston.

Johnson, H.M. (1936). Pseudo-mathematics in the social sciences. *American Journal of Psychology*, 48, 342–351.

Kant, I. (1978). *Critique of pure reason* (N.K. Smith, Trans.). London: Macmillan. (Original work published 1781.)

Kant, I. (1992). Inquiry concerning the distinctness of the principles of natural theology and morality. In D. Walford (Ed.), *The Cambridge edition of the works of Immanuel Kant: Theoretical philosophy, 1755–1770* (D. Walford with R. Meerbote, Trans.; pp. 243–286). Cambridge: Cambridge University Press. (Original work published 1764.)

Kelley, T.L. (1929). *Scientific method*. Columbus: Ohio State University Press.

Kerlinger, F.N. (1979). *Behavioral research: A conceptual approach*. New York: Holt, Rinehart & Winston.

Kline, P. (1997). Commentary on Michell, Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88, 385–387.

Kries, J. von. (1882). Über die Messung intensiver Größen und über das sogenannte psychophysische Gesetz. *Vierteljahrsschrift für wissenschaftliche Philosophie*, 6, 257–294.

Kuhn, T. (1970a). *The structure of scientific revolutions* (2nd ed.). Chicago, IL: University of Chicago Press.

Kuhn, T. (1970b). Reflections on my critics. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 311–341). Cambridge: Cambridge University Press.

Kuhn, T. (1991). The road since Structure. In A. Fine, M. Forbes, & L. Wessels (Eds.), *PSA 1990, Vol. 2* (pp. 2–13). East Lansing, MI: Philosophy of Science Association.

Kuhn, T. (1993). Afterwords. In P. Horwich (Ed.), *World changes: Thomas Kuhn and the nature of science* (pp. 311–341). Cambridge, MA: MIT Press.

Laming, D. (1997). A critique of a measurement-theoretic critique: Commentary on Michell, Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88, 389–391.

Lloyd, G.E.R. (1987). *The revolutions of wisdom: Studies in the claims and practice of ancient Greek science*. Berkeley: University of California Press.

Locke, J. (1959). *An essay concerning human understanding, Vol. 2*. Oxford: Oxford University Press. (Original work published 1690)

Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Lorge, I. (1951). The fundamental nature of measurement. In F. Lindquist (Ed.), *Educational measurement* (pp. 533–559). Washington, DC: American Council of Education.

Lovie, A.D. (1997). Commentary on Michell, Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88, 393–394.

Luce, R.D. (1997). Quantification and symmetry: Commentary on Michell, Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88, 395–398.

Luce, R.D., & Tukey, J.W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1, 1–27.

Mackie, J.L. (1973). *Truth, probability and paradox: Studies in philosophical logic*. Oxford: Clarendon Press.

McMullin, E. (1992). *The inference that makes science*. Milwaukee, WI: Marquette University Press.

Meier, S.T. (1993). Revitalizing the measurement curriculum: Four approaches for emphasis in graduate education. *American Psychologist*, 48, 886–891.

Michell, J. (1990). *An introduction to the logic of psychological measurement*. Hillsdale, NJ: Erlbaum.

Michell, J. (1996). S.S. Stevens' definition of measurement: The illogicality of an intellectual virus. In C.R. Latimer & J. Michell (Eds.), *At once scientific and philosophic: A Festschrift for John Philip Sutcliffe* (pp. 81–96). Brisbane: Boom-bana.

Michell, J. (1997a). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88, 355–383.

Michell, J. (1997b). Reply to Kline, Laming, Lovie, Luce and Morgan. *British Journal of Psychology*, 88, 401–406.

Michell, J. (1999). *Measurement in psychology: Critical history of a methodological concept*. Cambridge: Cambridge University Press.

Michell, J., & Ernst, C. (1996). The axioms of quantity and the theory of

measurement, Part I. An English translation of Hölder (1901), Part I. *Journal of Mathematical Psychology*, 40, 235–252.

Michell, J., & Ernst, C. (1997). The axioms of quantity and the theory of measurement, Part II. An English translation of Hölder (1901), Part II. *Journal of Mathematical Psychology*, 41, 345–356.

Morgan, M. (1997). Measurement in psychology: Commentary on Michell's Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88, 399–400.

Narens, L., & Luce, R.D. (1986). Measurement: The theory of numerical assignments. *Psychological Bulletin*, 99, 166–180.

O'Meara, D.J. (1989). *Pythagoras revived: Mathematics and philosophy in late Antiquity*. Oxford: Clarendon Press.

Parsons, C. (1990). The structuralist view of mathematical objects. *Synthèse*, 84, 303–346.

Passmore, J. (1943). Logical positivism I. *Australasian Journal of Psychology and Philosophy*, 21, 65–92.

Passmore, J. (1944). Logical positivism II. *Australasian Journal of Psychology and Philosophy*, 22, 129–153.

Passmore, J. (1948). Logical positivism III. *Australasian Journal of Psychology and Philosophy*, 26, 1–19.

Plato (1956). *Protagoras* (G. Vlastos, Ed.; B. Jowett & M. Ostwald, Trans.). New York: Bobbs-Merrill.

Plato (1971). *Timaeus and Critias* (B. Radice, Ed.; D. Lee, Trans.). Harmondsworth: Penguin.

Popper, K. (1970). Normal science and its dangers. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 51–58). Cambridge: Cambridge University Press.

Porter, T.M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.

Ramsay, J.O. (1991). Review of *Foundations of measurement, volumes II & III*. *Psychometrika*, 56, 355–358.

Reese, T.W. (1943). The application of the theory of physical measurement to the measurement of psychological magnitudes, with three experimental examples. *Psychological Monographs*, 55, 1–89.

Resnick, M.D. (1981). Mathematics as a science of patterns: Ontology and reference. *Noûs*, 15, 529–550.

Resnick, M.D. (1997). *Mathematics as a science of patterns*. Oxford: Clarendon Press.

Sankey, H. (1997). Kuhn's ontological relativism. In D. Ginev & R.S. Cohen (Eds.), *Issues and images in philosophy of science* (pp. 305–320). Dordrecht: Kluwer.

Shapin, S. (1996). *The Scientific Revolution*. Chicago, IL: University of Chicago Press.

Soyfer, V.N. (1994). *Lysenko and the tragedy of Soviet science*. Newark, NJ: Rutgers University Press.

Stevens, S.S. (1935a). The operational definition of psychological terms. *Psychological Review*, 42, 517–527.

Stevens, S.S. (1935b). The operational basis of psychology. *American Journal of Psychology*, 47, 323–330.

Stevens, S.S. (1936). Psychology: The propaedeutic science. *Philosophy of Science*, 3, 90–103.

Stevens, S.S. (1946). On the theory of scales of measurement. *Science*, 103, 667–680.

Stevens, S.S. (1951). Mathematics, measurement and psychophysics. In S.S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1–49). New York: Wiley.

Stevens, S.S. (1956). The direct estimation of sensory magnitudes—loudness. *American Journal of Psychology*, 69, 1–25.

Stevens, S.S. (1958). Measurement and man. *Science*, 127, 383–389.

Stevens, S.S. (1959). Measurement, psychophysics and utility. In C.W. Churchman & P. Ratoosh (Eds.), *Measurement: Definitions and theories* (pp. 18–63). New York: Wiley.

Stevens, S.S. (1967). Measurement. In J.R. Newman (Ed.), *The Harper encyclopedia of science* (pp. 733–734). New York: Harper & Row.

Stevens, S.S. (1968). Measurement, statistics, and the schemapiric view. *Science*, 161, 849–856.

Stevens, S.S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: Wiley.

Suen, H.K. (1990). *Principles of test theories*. Hillsdale, NJ: Erlbaum.

Suppes, P., & Zinnes, J. (1963). Basic measurement theory. In R.D. Luce, R.R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 1 (pp. 1–76). New York: Wiley.

Sylla, E. (1972). Medieval quantifications of qualities: The 'Merton School'. *Archive for History of Exact Sciences*, 8, 9–39.

Thorndike, E.L. (1918). The nature, purposes, and general methods of measurements of educational products. In G.M. Whipple (ed.), *Seventeenth yearbook of the National Society for the Study of Education*, Vol. 2 (pp. 16–24). Bloomington, IL: Public School Publishing.

Titchener, E.B. (1905). *Experimental psychology: A manual of laboratory practice*. London: Macmillan.

Venn, J. (1889). *The principles of empirical or inductive logic*. London: Macmillan.

Wise, M.N. (1995). *The values of precision*. Princeton, NJ: Princeton University Press.

Wundt, W. (1907). *Outlines of psychology* (C.H. Judd, Trans.). Leipzig: Engelmann. (Original work published 1896.)

ACKNOWLEDGEMENTS. This is a revision of a paper first presented in October 1997 while a Fellow of the Research Institute for Humanities and Social Sciences of the University of Sydney. Versions were given at colloquia at Macquarie University, the University of Melbourne and the University of Western Sydney in 1998 and 1999, and at the International Society for Theoretical Psychology Conference at Manly in April 1999. I am grateful for comments made at these venues and for the comments by reviewers for *Theory & Psychology*.

JOEL MICHELL teaches psychometrics and the history and philosophy of psychology at the University of Sydney. He is the author of *An Introduct-*

*tion to the Logic of Psychological Measurement* (Erlbaum, 1990) and *Measurement in Psychology: A Critical History of a Methodological Concept* (Cambridge University Press, 1999); co-author (with P. Bell and P. Staines) of *Logical Psych: An Introduction to Reasoning and Explanation in Psychology* (University of New South Wales Press, 2000) and co-editor (with C.R. Latimer) of *At Once Scientific and Philosophic: A Festschrift for John Philip Sutcliffe* (Boombana, 1996); and has published papers in both psychological and philosophical journals. ADDRESS: Department of Psychology, University of Sydney, Sydney, 2006, NSW, Australia. [email: joelm@psych.usyd.edu.au]