British Journal of Psychology (1997), **88**, 355–383 *Printed in Great Britain*

© 1997 The British Psychological Society

Quantitative science and the definition of *measurement* in psychology

Joel Michell*

Department of Psychology, University of Sydney, Sydney NSW 2006, Australia

It is argued that establishing quantitative science involves two research tasks: the *scientific* one of showing that the relevant attribute is quantitative; and the *instrumental* one of constructing procedures for numerically estimating magnitudes. In proposing quantitative theories and claiming to measure the attributes involved, psychologists are logically committed to both tasks. However, they have adopted their own, special, definition of measurement, one that deflects attention away from the scientific task. It is argued that this is not accidental. From Fechner onwards, the dominant tradition in quantitative psychology ignored this task. Stevens' definition rationalized this neglect. The widespread acceptance of this definition within psychology made this neglect systemic, with the consequence that the implications of contemporary research in measurement theory for undertaking the scientific task are not appreciated. It is argued further that when the ideological support structures of a science sustain serious blind spots like this, then that science is in the grip of some kind of thought disorder.

> ...unluckily our professors of psychology in general are not up to quantitative logic...
>
> E. L. Thorndike to J. McK. Cattell, 1904

Psychologists resist the intrusion of philosophical considerations into their science, as if such considerations could somehow threaten its genuine achievements. Resistance is especially stiff in the methodological area where the tone was set by the founder of quantitative methods in psychology, G. T. Fechner, who, to criticisms of his psychophysical methods, responded, 'all philosophical counter-demonstrations are, I think, mere writing in the sand' (1887, p. 215). The same resistance is noted by Meier (1994) from a reviewer of a critical paper: 'So much for what the paper says, I am even more concerned about what it implies; namely, that we should cease and desist because applied psychological measurement has flaws.... Applied measurement is the primary contribution that psychology has made to society' (p. xiii). The attitude that any science should insulate itself against criticism is anti-scientific. If principled criticism is not answered in a principled way then the doubts raised remain.

Philosophical criticism in the methodological area has a special function. If the

* Requests for reprints.

methods of science are not sanctioned philosophically then the claim that science is intellectually superior to opinion, superstition and mythology is not sustained. Psychologists have as much at stake in this as other scientists. The methods of science (e.g. observation, experiment, measurement, etc.) involve some of the deepest of philosophical problems, problems in which definitive solutions seem as elusive as ever. In seeking solutions to these problems, the scientist will prefer a *scientific* philosophy, i.e. one that is capable of justifying science as a rational cognitive enterprise. The natural scientific attitude and the one that promises the most coherent defence of science is that of *empirical realism* (i.e. that of an independently existing natural world which humans are able to successfully cognize via observational methods, at least sometimes). From this perspective, philosophical research now enables a clear view of what some methods, especially *measurement*, amount to.

Taking this view seriously, it will be argued that psychology is in danger of losing contact with the great intellectual tradition of quantitative, experimental science. Many psychological researchers are ignorant with respect to the methods they use. This ignorance is not so evident at the *instrumental* level, i.e. that of using techniques of data collection and analysis, although there is a surprising degree of ignorance even here. The ignorance I refer to is about the *logic* of methodological practices, i.e. about understanding the rationale behind the techniques. Knowing the logic of methodological practices is not a matter of icing on the cake, icing which, if neglected, leaves the substance unaffected. Ignorance of this logic may mean not knowing the right empirical questions to ask or, even, that there are any in this context. The history of science shows that the insights underlying quantification were hard won. The history considered in this paper, by comparison, shows that they are very easily lost and, once lost, not easily regained. When the attitude of turning a deaf ear to criticism becomes entrenched, it can alter patterns of thinking and the way words are used, with the result that criticism may be treated as irrelevant.

1. The concept of scientific measurement

1.1 *Centrality of the concept of quantity*

In quantitative science attributes (such as velocity, temperature, length, etc.) are taken to be measurable. That is, it is theorized that an attribute, such as length, has a distinctive kind of internal structure, viz., *quantitative structure*. Attributes having this kind of structure are called *quantities*. Following a well-established usage, specific instances of a quantity are called *magnitudes* of that quantity (e.g. the length of this page is a magnitude of the quantity, length). Magnitudes of a quantity are measurable because, in virtue of quantitative structure, they stand in relations (*ratios*) to one another that can be expressed as real numbers.

While quantitative science has existed since ancient times, quantitative structure, itself, was only explicitly characterized late in the nineteenth century and its best known formulation is given by Hölder (1901). Hölder's set of seven axioms define a continuous quantity and the following is a slightly more succinct definition of the same concept (Michell, 1994). A range of instances of an attribute, $\mathcal{Q}$, constitutes a continuous quantity if and only if the following five conditions obtain (in each case

an attempt has been made to state first a more accessible explanation of what the condition means, free of mathematical symbols and technical terms).

1. Any two magnitudes of the same quantity are either identical or different and, if the latter, there must exist a third magnitude, the difference between them, i.e. for any a and b in $\mathcal{Q}$, one and only one of the following is true

(i) $a = b$,
(ii) there exists c in $\mathcal{Q}$ such that $a = b + c$,
(iii) there exists c in $\mathcal{Q}$ such that $b = a + c$;

2. A magnitude entirely composed of two discrete parts is the same regardless of the order of composition, i.e. for any a and b in $\mathcal{Q}$, $a + b = b + a$;
3. A magnitude which is a part of a part of another magnitude is also a part of that same magnitude, the latter relation being unaffected in any way by the former, i.e. for any a, b and c in $\mathcal{Q}$, $a + (b + c) = (a + b) + c$;
4. For each pair of different magnitudes of the same quantity there exists another between them, i.e. for any a and b in $\mathcal{Q}$ such that $a > b$, there exists c in $\mathcal{Q}$, such that $a > c > b$; and
5. Given any two sets of magnitudes, an 'upper' set and a 'lower' set, such that each magnitude belongs to either set but none to both and each magnitude in the upper set is greater than any in the lower, there must exist a magnitude no greater than any in the upper set and no less than any in the lower, i.e. every non-empty subset of $\mathcal{Q}$ that has an upper bound has a least upper bound.

Note that one magnitude is greater than another if and only if the latter is a part of the former, i.e. for any a and b in $\mathcal{Q}$, $a > b$ if and only if (ii) above is true. Conditions 4 and 5 ensure the density and continuity, respectively, of the quantity, which intuitively may, thus, be thought of as containing no gaps in the sequence of its magnitudes.

Some words of caution should be added about the use of the mathematical symbol, '+', in the above conditions. Readers will be most familiar with the use of this symbol in arithmetic contexts, where the terms added are numbers. My first warning, then, is that in the above conditions the addition is not of numbers but of *magnitudes of a quantity* (e.g. specific lengths, say). My second warning is this: '+' is often understood as a mathematical operation and this interpretation, when applied to magnitudes, has sometimes (e.g. by Campbell, 1920, 1928) been understood as requiring an empirical operation of concatenation (i.e. an operation of putting magnitudes together in some way). Such an interpretation is not intended here and to forestall it I recommend the alternative of interpreting $a + b = c$ as a relation between the magnitudes a, b and c. The relation I have in mind is this: magnitude c is entirely composed of discrete parts, magnitudes a and b. This interpretation is suggested by Bostock (1979). The point of making this distinction is that just because magnitudes stand in this relation, it does not follow that suitable operations of either concatenation or division will obtain for objects possessing the magnitudes so related. This may be so, as with length and other convenient quantities, or it may not, as with density or temperature. That is, the additive relation between magnitudes is a theoretical one and how we gain access to it may often be indirect.

1.2 *The concept of quantity entails that of measurement*

If an attribute is quantitative then it is, in principle, measurable. This was the main theorem of Hölder's (1901) paper. He showed that given such structure, for any a and b in $\mathcal{Q}$, the magnitude of a relative to b may always be expressed by a positive, real number, r, where $a = r.b$. That is, the ratio of a to b (a positive, real number) is the measure of a in units of b. This fact, in turn, makes it meaningful to hypothesize the existence of quantitative relations between attributes (like that between density, mass and volume). The practice of measurement requires getting some grip, either directly or indirectly, upon the additive structure of the attribute in order that ratios between magnitudes of the attribute may be discovered or estimated. Hence, scientific measurement is properly defined as *the estimation or discovery of the ratio of some magnitude of a quantitative attribute to a unit of the same attribute.* It is invariably along such lines that measurement is, and always has been, defined in the physical sciences (see, for example, Beckwith & Buck, 1961; Clifford, 1882; Cook, 1994; Massey, 1986; Maxwell, 1891).

This definition of measurement is a logical consequence of the structure that quantitative attributes are taken to possess. Given that structure, the fact that magnitudes of a quantity stand in numerical relations to one another is a provable mathematical theorem. Measurement is nothing more or less than the attempt to discover or estimate such numerical relations. This is the logical basis of quantitative science with all its mathematical beauty, conceptual scope, empirical power and practical utility (see Appendix II).

1.3 *The empirical commitments of the concept of quantity*

In conceptualizing an attribute as quantitative a scientific hypothesis is proposed. There is no logical necessity that any attribute should have this kind of structure. Hence, accepting this hypothesis is speculative, unless there is evidence specifically supporting it. The issue of evidence for quantity is always complex and some of the conditions 1 to 5, mentioned in section 1.1, are never separately, directly testable (such as 5, the continuity condition). However, in the case of some quantities (e.g. length) conditions 2 and 3 are directly testable, at least for humanly manageable lengths. For example, if x, y and z are straight rigid rods and x exactly spans the rod entirely composed of discrete parts, y and z, linearly concatenated in a particular order, then if 2 is true, x must exactly span the length of the rod entirely composed of z and y, linearly concatenated in the opposite order.

It should be stressed that for many physical quantities, the existence of which is now taken for granted (e.g. temperature and density), the evidence that they are quantitative is entirely indirect. That is, the additive structure of the attribute is not directly reflected via a relation of physical concatenation, as with the case of length (at least for lengths of humanly manageable sizes). Given the fact that much of science involves theories that are likewise only indirectly testable, such a point would hardly be worth stressing had it not been an almost permanent source of confusion over the last century, both in psychology and measurement theory generally. It would seem that measurement has been mistakenly thought of by some philosophers

as being an atheoretical, purely observational base upon which science's more theoretical structures stand. It is not. Measurement always presupposes theory: the claim that an attribute is quantitative is, itself, always a theory and that claim is generally embedded within a much wider quantitative theory involving the hypothesis that specific quantitative relationships between attributes obtain. Because the hypothesis that any attribute (be it physical or psychological) is quantitative is a contingent, empirical hypothesis that may, in principle, be false, the scientist proposing such an hypothesis is always logically committed to the task of testing this claim whether this commitment is recognized or not.

1.4 *The two tasks of quantification: The scientific and the instrumental*

Establishing a quantitative science involves two tasks. First, there is the logically prior *scientific* one of experimentally investigating the hypothesis that the relevant attribute is quantitative. Second, there is the *instrumental* task of devising procedures to measure magnitudes of the attribute shown to be quantitative. Failure to investigate the scientific task prior to working upon the instrumental one and failure to confirm the hypothesis that the relevant attribute is quantitative means that treating the proposed measurement procedures as if they really are measurement procedures is at best speculation and, at worst, a pretence at science.

2. Measurement in psychology

2.1 *The measurement of psychological attributes and the commitment to quantitative structure*

Even a superficial perusal of relevant psychological publications reveals that psychologists believe that they are able to measure many distinctly psychological attributes, such as cognitive abilities, personality traits, social attitudes and sensory intensities. These attributes are distinctly psychological in the sense that they form part of psychology's subject matter and, also, in the sense that they do not belong to the network of quantitative attributes measurable using the methods of the physical sciences (see, for example, Jerrard & McNeill, 1992; Sena, 1972). While these psychological attributes do not form part of this network, it is clear that quantitative psychology was first modelled upon quantitative physics (Fechner, 1860). That is, in both disciplines alike, certain attributes are supposed to have quantitative structure.

2.2 *The consequent commitment to the scientific task of quantification*

Psychologists, in their attempt to construct a quantitative science by analogy with quantitative physics, hypothesize that some of their attributes are quantitative and, furthermore, that some of these attributes relate quantitatively, either amongst themselves or with physical quantities. For example, Fechner (1860), in proposing his psychophysical theory, conceived of the intensity of sensations as a quantity and hypothesized a particular functional relationship between it and the physical intensity of the stimulus. Or, Spearman (1904), in proposing that level of performance on an

intellectual task was due to a combination of the level of general ability and the level of the ability specific to that task, conceived of general ability and the various specific abilities as quantitative attributes and proposed a functional quantitative relation between these quantities and test scores. Fechner's and Spearman's quantitative speculations provided the model for many later developments, for example, those by Thurstone (1938), Hull (1943), Stevens (1956), Cattell (1943) and others. In every case, the above concept of continuous quantity was necessarily presumed, although it must be said that this was not often explicitly acknowledged. It was presumed, however, as a necessary concomitant of quantitative theorizing. Hence, presumed along with it, as part of the same conceptual package, was the traditional concept of scientific measurement.

3. The definition of measurement in psychology

3.1 *Stevens' definition of measurement*

Even though quantitative psychologists (by whom I mean those who either theorize about or attempt to measure psychological quantities) hypothesize that their attributes are quantitative and, so, commit themselves to the concept of scientific measurement, the definition of measurement actually endorsed by most of them is radically different. This definition is the one formulated by Stevens (1946): *measurement is the assignment of numerals to objects or events according to rule.* It is easily verified that this, or something very similar, is the definition they prefer. Many psychological texts, especially those on research methods or relating in some way to psychological measurement, offer their readers a definition of measurement. These definitions are surprisingly uniform and, while they do not all match Stevens' definition word for word (although many of them do), they wear the mark of Stevens.

3.2 *Its widespread acceptance within psychology*

I recently surveyed the psychology library of a major European university. I was easily able to locate 44 books (see Appendix I), each providing a definition of measurement, ranging in publication date from the early 1950s to the early 1990s. Of these, 39 offered a definition of this form: *measurement is the assignment of X to Y according to Z.* Within this schema, which derives directly from Stevens, X was typically either *numerals* or *numbers*, although other cognate terms sometimes appeared (e.g. numerical values, scores, abstract symbols). Stevens' phrase, *objects or events*, was mostly retained for Y, although a wide range of other terms appeared as well, including things, situations, individuals, behaviour, observations, attributes, properties and responses. Where authors departed from Stevens' preference for Z, it was generally (in about a dozen cases) to offer a specific rule (in most cases a representational rule). None of the 44 definitions even remotely resembled the traditional scientific concept. These observations confirm that psychology, as a discipline, has its own definition of measurement, a definition quite unlike the traditional concept used in the physical sciences.

3.3 *Its relation to the concept of quantity and to the consequent scientific task of quantification*

As will be shown in section 4, Stevens' definition of measurement entered quantitative psychology at a particularly crucial stage of its history. The definition was accepted by psychologists, not innocently, out of ignorance of the truth about quantity and measurement: it was accepted and became entrenched because it appeared to solve a conceptual problem which had existed since Fechner's time and which in the 1940s was particularly pressing. The acceptance of this definition involved a quite deliberate turning away from traditional concepts and it resulted in a systemically sustained blind spot, one which has persisted to the present. These are claims that I will support shortly by considering historical evidence.

However, first, it should be noted that (as I will show in more detail in section 5.2) the facts about quantity and measurement are readily evident to anyone motivated to find them. More than that, the idea that the practice of measurement is underwritten by the concept of a quantitative attribute and the knowledge of the structure of such an attribute is almost entirely absent from books on psychological measurement (I exclude from this generalization the publications of R. D. Luce, P. Suppes and their associates—e.g. Krantz, Luce, Suppes & Tversky, 1971; Luce, Krantz, Suppes & Tversky, 1990; Narens, 1985; Suppes, Krantz, Luce & Tversky, 1989—which are exceptional in this respect and, also, in having escaped the notice of the majority of quantitative psychologists (Cliff, 1992)). Such books (see, for example, Lord & Novick, 1968; Thorndike, 1982) typically begin with an account of the procedures used to measure psychological attributes and move on to a consideration of relevant quantitative theories, but nowhere do they explicitly discuss the empirical commitments implicit in such theories regarding the internal structure of the attributes involved. Nor do they discuss ways in which these commitments can be tested experimentally.

These two facts, the widespread acceptance of Stevens' definition of measurement amongst psychologists and the failure of books on psychological measurement to note the character of quantitative attributes, mean that the true nature of scientific measurement and the empirical content of the hypothesis that an attribute is quantitative are almost universally overlooked within psychology. If a quantitative scientist (1) believes that measurement consists entirely in making numerical assignments to things according to some rule and (2) ignores the fact that the measurability of an attribute presumes the contingent (and therefore, in principle, falsifiable) hypothesis that the relevant attribute possesses an additive structure, then that scientist would be predisposed to believe that the invention of appropriate numerical assignment procedures alone produces scientific measurement. This is exactly the situation that exists in quantitative psychology, a situation that Stevens' definition serves to justify.

4. The measurement tradition in psychology and the scientific task of quantification

4.1 *Fechner, pythagoreanism and the scientific task of quantification*

From its inception, modern quantitative psychology was more concerned with the implementation of a quantitative programme than with the pursuit of answers to fundamental scientific questions about its hypothesized quantities. While there had been attempts prior to Fechner's (1860) to establish a quantitative psychology, most notably Herbart's (1816), these attempts failed where Fechner's succeeded because (1) at the theoretical level, he linked his quantitative psychology to quantitative physics, via his psychophysical law; (2) at the practical level, he supplemented his law with a range of alleged measurement methods; and (3) at the level of rhetoric, he persuaded others that these methods were *measurement* in exactly the same sense as this term is used in physics. In doing this Fechner's motives were *pythagorean*, i.e., like Pythagoras (Burnet, 1955) and, following him, many of the greatest scientific and philosophic minds in history (Crombie, 1994), Fechner believed that reality is fundamentally quantitative. Both the physical and mental realms, in common, were subordinate, he believed, 'to the principle of mathematical determination' (Fechner, 1887, p. 213). As such, psychology and, in particular, psychophysics must be quantitative.

> As an exact science psychophysics, like physics, must rest on experience and the mathematical connection of those empirical facts that demand a measure of what is experienced or, when such a measure is not available, a search for it (Fechner, 1860, p. xxvii).

However, Fechner's commitment to 'experience' as the basis of science was not as firm as his commitment to pythagoreanism and when this deviation from empiricism was combined with his inadequate understanding of the nature of measurement, the result was a dogmatic *a priorism*.

The pythagoreanism that Fechner had inherited from his education in the physical sciences was only one side of the ideological legacy bequeathed to psychology by the scientific revolution of the 17th century. The other was the *quantity objection*, the thesis that psychological attributes are not quantitative. To some extent, this objection had its origins in the 17th-century relegation of the secondary qualities (like colours, flavours, odours, etc.) to the mind (while the allegedly real, physical (i.e. non-mental) qualities of things were held to be quantitative) and in a later, 18th-century, Kantian view (Kant, 1786) that, as a matter of fact, psychology can never be a quantitative science. Fechner's main critic from this perspective was von Kries (1882), who argued that sensations do not stand in additive relations to one another and, so, the claim that one sensation is, say, ten times another in intensity, is meaningless. James (1890) and other psychologists (see Titchener, 1905; and Boring, 1921 for reviews) made similar criticisms.

From the scientific point of view, the only way to counter such a criticism is to present evidence that intensities of sensations are additive. Fechner did not see the need to do this. He believed that because he could, using his psychophysical methods, determine a series of stimuli in which each is just noticeably different from its immediate predecessor, the elements of the corresponding series of sensations would

each differ from its predecessor by one unit of sensory intensity. Thus, thought Fechner, the psychophysicist is simply counting units, albeit indirectly via stimulus intensities, in a manner analogous to the physicist counting units of some physical quantity. In a reply to von Kries he put the matter this way.

> Given several values, in any field, which may be taken to be magnitudes inasmuch as they can be thought of as increasing or decreasing; given the possibility of judging the occurrence of equality and inequality in two or more of these values when they are observed simultaneously or successively; and given that n values have been *found* equal or, if they can be varied freely, have been *made* equal: then it is self-evident (because it is a matter of definition and therefore a tautology) that their total magnitude,which coincides with their sum, equals $n \times$ their individual magnitudes. It follows that each single value, or each definite fraction or each definite multiple of the magnitudes that have been found equal (no matter which), can be taken as the unit according to which the total magnitude, or every fraction of it, can be measured. The n equal parts that can be thought of as composing a total magnitude of course have the same magnitude as the n equal parts into which the total magnitude can be thought to be decomposable. All physical measurement is based on this principle. All mental measurement will also have to be based on it (Fechner, 1887, p. 213).

As revealed in this quotation, Fechner's understanding of the logic of measurement was seriously defective. Given an ordered series of elements, $a_1, a_2, \ldots, a_n$, claiming to show that $a_{i+1} - a_i = a_i - a_{i-1}$ (for all i) does not amount to showing that $a_{i+1} - a_{i-1} = 2(a_{i+1} - a_i)$, unless it is also shown that $a_{i+1} - a_{i-1} = (a_{i+1} - a_i) + (a_i - a_{i-1})$, i.e. unless it is shown that the series possesses an additive structure. That is, Fechner needed to show that differences between sensation intensities are additive in order to justify his claim that counting jnds is counting units of measurement. Fechner, believing all attributes to be quantitative, thought that the only task required of him as a scientist was the instrumental one of identifying units and counting them. That apparently done, it was 'self-evident', he thought, that a series of what he took to be n equal and contiguous intervals would equal n of each. The flaw in Fechner's thinking was his pythagoreanism. It caused him to presume incorrectly that psychological attributes must be quantitative. Thus set, his mind could not see the force of von Kries' quantity objection.

One way in which the additivity of differences can be tested experimentally was revealed by Hölder (1901) in his axioms for stretches of a straight line. Hölder was thinking of the geometric case, but it applies to Fechner's case by analogy. This test, in fact, is a special case of the Thomson condition in conjoint measurement (see Krantz *et al.*, 1971). If (for any $f > g > h$ and $i > j > k$ ranging over $l, \ldots, n$) $a_f - a_g = a_i - a_j$ and $a_g - a_h = a_j - a_k$, then the hypothesis that these differences are additive entails that $a_f - a_h = a_i - a_k$. If this test is represented geometrically, as in Fig. 1, it is obvious that the additivity of the component distances entails the conclusion stated. Sensory differences are not geometric distances, however, and what is obviously true of the latter is not necessarily true of the former. If individuals judge the sense differences $a_f - a_g$ and $a_i - a_j$ equal, and also judge sense differences $a_g - a_h$ and $a_j - a_k$ equal, it does not follow from that alone that they must judge $a_f - a_h$ equal to $a_i - a_k$. That only follows if sense differences are quantitative. This experiment would test this proposition.

This test requires participants making direct judgments about the equality of sensory differences, a task which Fechner allowed could be done (Fechner, 1887).

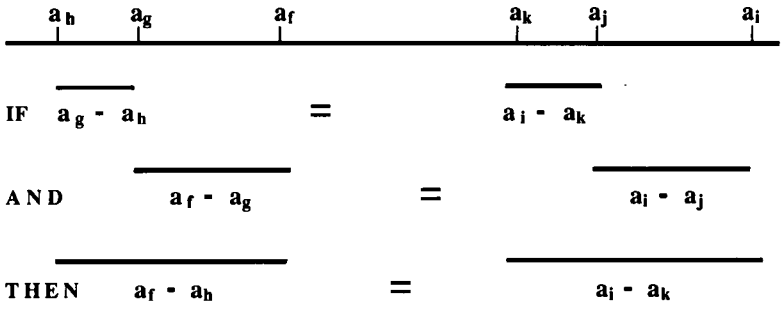

Figure 1. If the component distances are additive then, given the two antecedent conditions, the consequent condition follows.

The fact that it is in principle possible for a participant's judgments to violate this prediction and, hence, falsify a consequence of additivity, means that this is an empirical issue. It must be investigated experimentally before any claim to scientific measurement is justified. The logic of such tests was not known to Fechner and, indeed, was not generally accessible in the psychological literature until Suppes & Zinnes (1963). However, as already argued, Fechner's mind was effectively closed to the possibility of such empirical tests of additivity and the extent of this closure is revealed in his comment in relation to the Plateau–Delboeuf method that 'We simply call a total difference twice as large as each of two equal partial differences of which it is composed' (Fechner, 1887, p. 214).

4.2 *Fechner's psychophysics as the exemplar of psychological measurement*

Fechner's mind was closed because of his commitment to the doctrine of pythagoreanism and, being thus, he presented his psychophysical methods as methods of measurement within a milieu that shared similar views (Hacking, 1983; Michell, 1990). As the founding father of quantitative psychology, Fechner's work established the quantitative paradigm in psychology and became the definitive exemplar emulated by others. In this it established the trends of (*a*) dismissing the quantity objection as 'mere writing in the sand' (Fechner, 1887, p. 215) and (*b*) concentrating exclusively upon the instrumental task of quantification. Psychologists after Fechner also ignored the first task and concentrated upon the second, constructing number-generating procedures which, they thought, measured psychological attributes.

4.3 *Spearman, the quantitative imperative, and practicalism*

The truth of this last claim is shown clearly in Spearman's pioneering research on the measurement of intelligence. Spearman had been trained in psychophysics under Wundt and, so, naturally, had been influenced by Fechner's example. As part of that training it is almost certain that he would have been exposed to the controversy

surrounding the quantity objection. Spearman proposed a quantitative theory which purported to explain intellectual performance. He carried out tests of this theory (such as his predictions relating to tetrad differences), but these were to do with the number of abilities involved in solving tasks of specific kinds and were not sensitive to the issue of whether or not the postulated abilities were quantitative in structure. Like Fechner, he believed that psychological attributes had to be quantitative and the primary problem was to devise procedures for their measurement.

In doing this Spearman's motivation was not quite the same as Fechner's. He was not explicitly committed to pythagoreanism, but he did endorse the view that lacking measurement 'any study is thought by many authorities not to be scientific in the full sense of the word' (Spearman, 1937, p. 89). This view, that measurement is a necessary feature of all science, has been called the *quantitative imperative* (Michell, 1990). However, Spearman was also moved to promote ability tests as measurement instruments by another reason as well. In his epoch-making 1904 paper, he expressed his disappointment in psychology as a basis for applied science, especially in education and psychiatry: '... when we without bias consider the whole actual fruit so far gathered from this science—which at the outset seemed to promise an almost unlimited harvest—we can scarcely avoid a feeling of great disappointment' (1904, p. 203). Spearman believed that psychology should provide a quantitative basis for practical applications in these areas and he expressed the hope that his research would produce 'practical fruit of almost illimitable promise' (1904, p. 206).

Practicalism, the view that science should serve practical ends, when stronger than the spirit of disinterestedness, can corrupt the process of investigation. Science, as the attempt to understand nature's ways of working, knows nothing of practicalism, for scientific knowledge is neither useful nor useless, in itself. It only becomes so when taken in relation to interests other than the scientific, interests which may shift with the tide of social change. When scientific questions seem intractable, an impatient practicalism may presume upon nature. Unable to admit that such questions cannot yet be answered, the practicalist may blindly collude in the pretence that they do not exist by ignoring them, especially when the psychological expectations and social rewards are high. Spearman thought psychological measurement a necessity, practicalism an imperative, and he ignored the scientific issue of quantification. Hence, the hypothesis of a causal connection cannot be ruled out. In these attitudes he was not alone.

4.4 *Applied psychology, practicalism and the instrumental task*

Practicalism became a powerful motive driving the instrumental side of quantitative psychology. In America, the impetus for applied psychological measurement came from the work of J. McK. Cattell and E. L. Thorndike. Cattell believed, like Spearman, that 'Psychology cannot attain the certainty and exactness of the physical sciences, unless it rests on a foundation of experiment and measurement' (1890, p. 373) and wrote freely about *mental measurements* (1893) with respect to a range of psychological procedures, with no acknowledgement of the scientific task of quantification, even though he endorsed the traditional scientific concept of

measurement ('all measurement depending on ratios'; 1893, p. 321). In advocating mental measurements, he had applied psychology firmly in mind:

> Control of the physical world is secondary to the control of ourselves and our fellow man... If I did not believe that psychology affected conduct and could be applied in useful ways, I should regard my occupation as nearer to that of the professional sword-swallower than to that of the engineer or scientific physician (Cattell, 1904, as quoted in Brown, 1992, p. 3).

That engineering and medicine should have been consistently selected as the guiding metaphors for applied psychology (Brown, 1992) reveals a connection between practicalism and quantification. Engineering is applied physics and physics is the paradigm of quantitative science. If applied psychology was seen in this light, then the pressure exerted upon psychology by practicalism to claim to be able to measure would have been great indeed. This metaphor occurred repeatedly (e.g. Terman, 1916; see also Brown, 1992). Medicine had also recently made great strides through the introduction of quantitative methods derived from physiology and, likewise, provided an example that applied psychologists strove to emulate (Brown, 1992).

Thorndike, agreeing with Fechner's pythagoreanism, had said that 'Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quantity' (quoted in Clifford, 1968, p. 283). His view that 'Any mental trait in any individual is a variable quantity' (1904, p. 22) is an immediate consequence. Conjoined with the engineering metaphor it entails that 'Education is one form of human engineering and will profit by measurements of human nature and achievement as mechanical and electrical engineering have profited by using the foot-pound, calorie, volt and ampere' (quoted in Brown, 1992, p. 119). While Thorndike was aware that measurement in psychology ('by relative position') was different from that in physics ('by amount of some unit'), he believed that 'Measurement by relative position in a series gives as true, and may give as exact, a means of measurement as that by units of amount' (1904, p. 19). Of course, to assert that, in a quantitative order, a particular value, b, falls between two others, a and c, is never as exact a specification of its relation to other values (to a, for example) as is given by specifying its numerical relation to a unit (for the latter will always entail that $b = ra$, where r is a positive real number, while the former never does). However, in asserting this Thorndike was giving psychologists permission to use the term *measurement* for practices which were not supported by any scientific evidence of quantity. The advantages of this terminology to applied psychologists attempting to present themselves as applied scientists is obvious. It allowed them to present themselves as applied scientists in what was an easily identified manner. This terminology rapidly became standard and it was a commonplace observation that 'Tests are the devices by which mental abilities can be measured' (Viteles, 1921, p. 57).

This way of thinking, combined with the successful incorporation of tests in American society following World War I, meant that the scientific task of quantification was easily ignored. Some idea of how extensive the use of tests in applied psychology was may be gauged by a number of indices. By 1922, three million children a year were subjected to one form or another of mental measurement (Thorndike, 1923). By 1937, '5005 articles, most of them reports of new tests, which [had] appeared during the fifteen year period between 1921 and 1936' (South, 1937,

as quoted in Hornstein, 1988) were available for use by psychologists. Terman (1921) reported that 'more than half of the psychological research which is being carried on by members of the American Psychological Association (which includes practically all the psychologists of the United States) falls in one or another of the fields of applied psychology' (p. 3).

Some idea of the proportion of applied psychological research that was devoted to measurement can be gained from considering the number of publications reporting mental measurements of one kind or another published in the *Journal of Applied Psychology* from its inception in 1917 to 1946 (the year of the publication of Stevens' definition). From 1917 to 1926 the number was 150 (out of 338, or 44.4 per cent); from 1927 to 1936, 234 (out of 585, or 40 per cent); and from 1937 to 1946, 253 (out of 668, i.e., 38 per cent). In these studies, scores (generally obtained via counting responses of a certain kind) or transformations of scores were typically treated as measures of some psychological attribute or other. In the spirit of both Fechner and Spearman, the scientific question of the additivity of the attributes involved was hardly ever addressed. That is, decades prior to the publication of Stevens' definition, the practices of applied psychologists already conformed to it.

4.5 *Psychological measurement as a scientific anomaly*

The logic of applied science is such that if there is no science there can be no applied science, so whatever these psychologists were doing it was not applied *science*. Actually, they were doing what so-called 'applied psychologists' have, in general, always done. They were applying a *methodology* to what were thought of as practical problems (Freyd, 1926; Terman, 1924). This methodology was based upon the construction of procedures that yield numerical data. Such procedures, innocent enough in themselves, were then packaged and marketed as forms of 'scientific measurement' and, as such, constituted a pretence of applied science, rather than applications of empirically confirmed scientific theories. Constructing psychological tests for practical applications may be a useful thing to do. However, its relation to psychology as a science needs to be clarified. Even if it is found that performance upon a particular test (say, test A) is useful for predicting some criterion (say, success in a training course, X), this by itself is not *applied science* in any meaningful sense. The discovery that A predicts X *raises* a scientific issue, it does not solve one and neither does it amount to the application of a well-confirmed body of scientific theory to a practical problem as, for example, engineers apply physics to the building of a bridge. The scientific issue raised by such a discovery is this: why does performance on A predict performance on X? In attempting to answer such a question, a psychologist may theorize that test A *measures* intellectual ability I and I, in turn, is a cause of performance on X. This is a perfectly respectable way to theorize, but it is only one possible theory out of an indefinitely large array of possible theories and remains so until thoroughly investigated empirically. A part of this process of investigation must involve testing the hypothesis that ability I is a quantitative attribute. If this is not done then the claim that A measures I remains completely speculative. In this case, the use of test A to 'measure' I is not applied *science* and it

is misleading to think of it in this way. Until the scientific task of quantification is completed, claiming that a procedure measures anything is premature.

Hence, both quantitative psychology and 'applied psychological measurement' stood as anomalies within a discipline purporting to be a science because psychologists either declined to or did not know how to consider fundamental scientific issues and persistently presented their procedures as amounting to more than was justified scientifically. This anomaly was repeatedly noted by those who took the care to understand the character of scientific measurement. I have already mentioned the quantity objection to Fechner's work. Again, in 1913, at a joint symposium organized by the Mind Association, the British Psychological Society and the Aristotelian Society, the issue of whether or not sensory differences are quantitative was considered by Brown (1913), Dawes Hicks (1913), Myers (1913) and Watt (1913). Boring, who was to consider the quantity objection in some detail in his paper on the stimulus error (Boring, 1921), had written a year earlier that psychologists were 'not yet ready for much psychological measurement in the strict sense' (1920, p. 32), a theme he reportedly continued to maintain more than a decade later (Newman, 1974). Psychological tests were subjected to similar criticisms (e.g. McCormack, 1922). In the 1930s, other scholars (e.g. Adams, 1931; Brown, 1934; Johnson, 1936) made many of the same criticisms again. Thus, there remained within psychology a critical trajectory that surfaced from time to time, from von Kries until the 1930s. The mainstream of quantitative psychologists paid it no heed, however. As long as the criticism was internal to the discipline and consisted of only a few voices, it could be ignored with impunity. However, when it became external, with some kind of official backing, notice was taken of it.

4.6 *The Ferguson Committee*

When in 1940, a committee established by the British Association for the Advancement of Science to consider and report upon the possibility of quantitative estimates of sensory events published its final report (Ferguson *et al.*, 1940) in which its non-psychologist members agreed that psychophysical methods did not constitute scientific measurement, many quantitative psychologists realized that the problem could not be ignored any longer. Once again, the fundamental criticism was that the additivity of psychological attributes had not been displayed and, so, there was no evidence to support the hypothesis that psychophysical methods *measured* anything. While the argument sustaining this critique was largely framed within N. R. Campbell's (1920, 1928) theory of measurement, it stemmed from essentially the same source as the quantity objection.

At this juncture in the history of psychology two avenues were open. One was to admit the validity of these criticisms and so admit that the scientific issue of whether or not psychological attributes are quantitative had not been adequately addressed. In doing this, psychologists would have been admitting that their claim to be able to measure their attributes rested upon theory and speculation rather than upon direct scientific evidence. They could have put a brave face to the scientific world and claimed with Bartlett (1940, p. 441) that 'Scientific insight, as everyday perception,

has ever run ahead of measurement and mathematical proof'. The next step would have been to explore ways of testing the question begged throughout their history and a good starting point would have been Hölder (1901) (which Nagel, 1932, had recently brought to the attention of philosophers).

4.7 *Stevens' attempt at a rational reconstruction of psychological measurement*

However, this was not the path taken. The combined pressure of scientism (in the guise of pythagoreanism and the quantitative imperative) and practicalism was too strong. One finds in the psychological journals from 1940 to 1950 a rash of papers attempting to defend psychological practices and, sometimes, to redefine measurement in a way that legitimized psychology's claim to the scientific high ground it had never actually occupied (e.g. Bartlett, 1940; Bergmann & Spence, 1944; Brower, 1949; Comrey, 1950; Coombs, 1950; Cureton, 1946; Gulliksen, 1946; Nafe, 1942; Perloff, 1950; Reese, 1943; Thomas, 1942). This process culminated in Stevens' early papers on measurement theory (Stevens, 1946, 1951).

Like other quantitative psychologists, Stevens endorsed the quantitative imperative: 'It can be said that the history of science is the history of man's efforts to devise procedures for measuring and quantifying the world around him' (1967, p. 734). Thus, like Fechner, he presumed that psychophysical measurement was possible. It was Stevens' *sone scale* of loudness that the Ferguson Committee considered as a putative example of psychophysical measurement. He believed that his psychophysical methods produced scales of 'true numerical magnitude' (1936*b*, p. 406), so the contradiction of this claim by eminent members of that committee spurred him to defence. Short of undertaking the intellectual and scientific labour necessary to test that claim empirically, what Stevens required was an effective rationalization which would render such empirical tests apparently redundant. One can admire Stevens' resourcefulness in constructing this rationalization. As an exercise in rhetoric it displayed considerable creativity. However, in science all is not rhetoric. It is the role of the scientist, as far as possible, to let the facts speak for themselves. Instead, Stevens constructed a solution that obscured the facts from view.

His rationalization was a two-tiered ideological structure. The first layer was a reconstruction of the representational theory of measurement (the theory that measurement is the numerical representation of empirical relations). Russell (1897) had already attempted to undermine the traditional concept of measurement by trying to disengage the concept of number from that of quantity (Michell, in press). Following this he (1903) provided the first systematic presentation of the representational theory of measurement (Michell, 1993). However, it was not a completely thoroughgoing representationalism. Even less so was Campbell's (1920, 1928) later version, which had guided the thinking of many on the Ferguson Committee, for it really attempted little more than to translate the traditional concept of measurement into representational terms, requiring additivity as a necessary condition for all measurement. Stevens was, I think, the first to see clearly that basing measurement upon the concept of numerical representation freed it from exclusive dependence upon the concept of additivity or that of any specific relation beyond

equivalence. For Stevens, 'measurement is possible in the first place only because there is a kind of isomorphism between (1) empirical relations among objects and events and (2) the properties of...' numerical systems (Stevens, 1951, p. 1). From this starting point he developed his theory of the four possible types of measurement scales (nominal, ordinal, interval and ratio) (he later, 1959, added the log-interval scale) and the associated doctrine of permissible statistics (see Michell, 1986).

This layer of his reconstruction was a masterstroke because it at once disarmed Campbell and his associates on the Ferguson Committee of their most powerful weapon. Their criticism of psychophysical measurement, that it was not based upon the demonstration of any relevant additive relation between sensory intensities, was made to look as if it depended upon an unnecessarily restrictive version of the representational theory of measurement. However, this variety of liberalized representationalism also posed a threat to psychological measurement and, especially, to Stevens' psychophysical methods. If measurement involves the numerical representation of empirical relational structures and such structures are understood realistically (i.e. as structures existing independently of the scientific observer), then measurement still requires a logically prior scientific stage in which the hypothesis is tested that relations of the required kind hold within a particular empirical domain. Even the humble nominal scale would require the demonstration of a reflexive, transitive, and symmetric empirical relation of equivalence (or *sameness* with respect to some attribute) and for most putative instances of psychological measurement not even this much scientific work had been done. Hence, the second layer of Stevens' reconstruction required a repudiation of such a realist interpretation of the empirical structures numerically represented in measurement.

If Stevens' first layer was a masterstroke, his second was audaciously bold, for he replaced the natural scientific attitude of realism with a form of relativistic subjectivism. The 1930s had been a time of ferment within the philosophy of science, with the newer views of logical positivism and operationism challenging older positions (see Passmore, 1957). Stevens adopted both operationism (Hardcastle, 1995) and logical positivism (Stevens, 1939). His representational theory of measurement, with its emphasis upon the numerical representation of directly observable empirical relations and its formalist conception of numerical systems was essentially positivistic. His bolder vision, however, was to construct an *operational* interpretation of the representational theory. He took Bridgman's (1927) operationism, according to which the meaning of a concept is synonymous with the *operations* used to identify it and applied it to psychology (Stevens, 1935*a*, *b*, 1936*a*, *b*, 1939), agreeing with Bridgman that 'the meanings of our words can never transcend the operations which went into their determination' (1936*a*, p. 93). Operationism has more in common with Berkeley's idealism than with empirical realism (Michell, 1990) and this is not a judgment with which Bridgman would have disagreed (see Bridgman, 1950) and Stevens (1936*a*) also endorsed its implied subjectivism. What is of interest here, however, is the use to which Stevens put Bridgman's philosophy.

In the first place, operationism was woven by Stevens into an elaborate, self-serving, philosophy. Across a number of papers in the 1930s, he argued (1) that because operationism implies a relativity to the observer in all science, psychology is the 'propaedeutic science', the study of the observer (1936*a*); (2) all scientific

operations are reducible to the operation of sensory discrimination (1935*a*); (3) the operational methods of psychophysics are central to the study of sensory discrimination (1935*b*); and (4) the scaling methods advanced by Stevens allow the central question of psychophysics to be decisively answered (1936*b*). Operationism enabled Stevens to believe that his research was the rock upon which all science stood.

Secondly, operationism enabled him to believe that his psychophysical methods yielded ratio scale measurement of the intensity of sensations without research into the scientific task of quantification being necessary. If the meaning of a scientific concept is given by the operations (i.e. procedures) used to identify it, then it follows that the empirical relations numerically represented in measurement must likewise be defined by the operations used to identify them (Bergmann & Spence, 1944). Thus, for Stevens, it could just as truthfully be said that measurement is possible because of an isomorphism between empirical *relations* and numerical ones, as because of an 'isomorphism between the formal system and empirical *operations*' (Stevens, 1951, p. 23, my italics). The point here is that ordinarily a distinction is made between a relation into which things enter (e.g. the relation of object x being heavier than object y) and an operation or procedure used to identify such a relation (e.g. by placing x and y in the pans of a balance and observing that the arm of the balance supporting x tilts down). The fact that x is heavier than y would normally be thought of as a *necessary* condition for this outcome of the operation. However, because of his operationism, Stevens would confuse two such facts, asserting that the operation with its outcome is all that is meant by asserting that the relation holds. If ordinal numerical assignments were made to x and y they could, on Stevens' view, be taken to represent the above *operation* as easily as the above *relation*. When the operation itself involves making numerical assignments (as, say, with Stevens' psychophysical methods or with mental testing), these assignments may be taken, according to this operationist logic, to both define the relation represented and to represent it. It is only if this point is grasped that one can appreciate the conceptual unity between Stevens' definition of measurement as *the assignment of numerals to objects or events according to rule* and his representationalism.

Given operationism, any rule for assigning numerals to objects or events could be taken as providing a numerical representation of at least the equivalence relation operationally defined by the rule itself. Hence, any rule for making numerical assignments always defines at least a nominal scale, according to Stevens' view. This is why Stevens was able to rephrase his definition as 'the assignment of numerals to objects or events according to rule—any rule' (1959, p. 19) without feeling that he had shifted his ground one inch. Of course, to the realist, this way of thinking is viciously circular and there appears to be an hiatus between Stevens' definition of and theory of measurement (Michell, 1986). This is because the realist views the relations represented in measurement as having an existence independent of human observations or operations.

When this operational way of thinking was applied to his psychophysical methods it gave Stevens what he wanted. In the first place, it enabled him to reject the concept of 'private or inner experience for the simple reason that an operation for penetrating privacy is self-contradictory' (1936*a*, p. 95) and to conclude that what psycho-

physicists had hitherto thought of as 'a subjective scale is a scale of response' (1936*b*, p. 407). Then, further applying operationism, the relation of one tone's sounding half as loud, say, as another will be defined by the operation used to determine it, i.e. by a subject judging it to be so. Thus, Stevens was able to believe that '... the response of the observer who says "this is half as loud as that" is one which, for the purpose of erecting a subjective scale, can be accepted at its face value' (1936*b*, p. 407). Then it follows, he thought, that such a scale is additive because 'With such a scale the operation of addition consists of changing the stimulus until the observer gives a particular response which indicates that a given relation of magnitudes has been achieved' (1936*b*, p. 407). According to Stevens, an additive (or ratio) scale is obtained because the person is both instructed to judge and taken to be judging additive or numerical relations. In this case, the operation by which the numerical assignments are made was taken by Stevens to define the ratio scale that he believed was produced.

4.8 *Its implications for the scientific task of quantification*

Thus Stevens had both deflected the criticisms of the Ferguson Committee by adopting a thoroughgoing representationalism and protected his own psychophysical methods from scientific challenge by adopting an operationist interpretation of representationalism. Both Fechner and Stevens thought that their methods could be taken as methods of measurement without any further scientific justification. In so doing, Fechner established the psychological tradition of regarding number-generating procedures as measurement, a tradition strengthened by Spearman and set in concrete by applied psychometrics. Stevens then enshrined the practices of that tradition within an explicit definition of measurement. Prior to its formulation psychologists were already in the habit of regarding as measurement procedures for making numerical assignments to objects or events and were in the habit of ignoring the scientific issues relating to quantification. Stevens' definition perfectly matched this practice and appeared to legitimize it. For this reason it was rapidly absorbed into psychology's ideological support structures, soon after publication being cited in major texts as the only definition of measurement (e.g. Green, 1954; Guilford, 1954; Lorge, 1951), being referred to in leading journals as the 'classical' theory of measurement (Coombs, Raiffa & Thrall, 1954), a characterization that persisted (Fraser, 1980), and eventually projected back upon Fechner himself (Adler, 1980).

That it blinded the majority of psychologists to the scientific necessity of testing via experiment that psychological attributes are quantitative was dramatically revealed over the subsequent four decades. What may be described as a broadly realist interpretation of Stevens' thoroughgoing representational theory of measurement was worked out only just over a decade later. Suppes & Zinnes (1963) considered a wide range of numerically representable empirical relational structures. In many cases they stated necessary and sufficient empirical conditions for numerical representations of one kind or another. Their theory implied a much stricter definition of measurement than did Stevens'. Measurement was a homomorphism between independently existing empirical and numerical relational systems. Because of the required independence of the two systems involved, such a homomorphism is

never automatically obtained merely by having a rule for making numerical assignments. Furthermore, if the empirical relational system is quantitative (as required within quantitative theories), then the scientific task of quantification is made explicit. This made clear to the psychological scientific community the fact that measurement required attention to fundamental empirical issues and, by implication, condemned the non-empirical measurement tradition of Fechner, Spearman and Stevens. Then, a year later, perhaps the most important development in measurement theory since Hölder (1901) occurred when Luce & Tukey (1964) published the theory of conjoint measurement (which was anticipated to some extent by Adams & Fagot, 1959, and Debreu, 1960). This theory and its subsequent developments (see Krantz *et al.*, 1971) revealed a range of decisive, indirect tests for the hypothesis that attributes are quantitative. This research programme culminated in the publication of volume three of *Foundations of Measurement* (Luce *et al.*, 1990). It makes explicit the conditions under which apparently non-additive empirical structures are really additive at a deeper level (viz. when automorphisms of such structures constitute a simply ordered, Archimedean group; see also Luce, 1987). While this research was directly relevant to the scientific task of quantification, it inspired only a relatively small number of empirical studies towards that end, leaving that task still seriously incomplete. For the most part, mainstream quantitative psychology remained oblivious to these revolutionary developments in measurement theory despite the fact that they occurred within the discipline and were published in one of its leading journals (*Journal of Mathematical Psychology*). It is of interest to note, also, that Stevens' preferred methods of psychophysical measurement (magnitude estimation and cross-modality matching) were analysed theoretically by some of those associated with this programme (e.g. Krantz, 1972; Luce, 1990; Narens, 1996). These analyses state empirical conditions which must be true if these methods are actually measuring the psychological attribute of sensation intensity.

Almost at the inception of this research programme, Stevens, still oblivious to the scientific issues involved, belittled its achievements, claiming that 'measurement models drift off into the vacuum of abstraction and become decoupled from their concrete reference' (Stevens, 1968, p. 854) and, thereby, demonstrated his persisting blind spot. So complete is this failure to see the obvious that a recent commentary (Cliff, 1992) upon the failure of mainstream quantitative psychology to absorb the conceptual breakthroughs of this programme effectively laid blame upon that programme itself (a charge, incidentally, rejected by Narens & Luce, 1993). Cliff argued that its lack of influence resulted primarily from the difficulty level of the mathematics used in reporting its research results, the lack of demonstrated empirical power of the results obtained, and the difficulty of dealing with error. He admitted that 'factors that have to do with the zeitgeist and the habits of work and thought among psychologists' (p. 189) also played a role, but there was no recognition of the fact that those who understand measurement in the way defined by Stevens must of necessity regard ways of addressing the scientific task underpinning quantification as not only irrelevant to psychological measurement but as placing an unnecessary obstacle in the way of its development and application.

It should be noted that the minority tradition of those critical of psychological measurement, spawned a small but continuous series of experimental studies

engaging the scientific task, especially in psychophysics (e.g. Beck & Shaw, 1967; Gage, 1934*a*, *b*; Garner, 1954; Gigerenzer & Strube, 1983; Levelt, Riemersma & Bunt, 1972; Reese, 1943; Zwislocki, 1983). Here is not the place to review this work but, in general, its thrust does not warrant its consistent neglect by those endorsing Stevens' definition of measurement and advocating psychological methods as measurement procedures.

5. Psychological measurement and methodological thought disorder

5.1 *The concept of thought disorder*

By methodological thought disorder, I do not mean simply ignorance or error, for there is nothing intrinsically pathological about either of those states. Ignorance has many causes and not all indicate a cognitive fault. Likewise, when error occurs under certain conditions, it can be construed as part of the normal functioning of the cognitive system (e.g. the standard geometric, visual illusions). Ignorance and error are only pathological when some mechanism within that system sustains them under external conditions favourable to their correction. Hence, the thinking of one who falsely believes he is Napoleon is adjudged pathological because the delusion was formed and persists in the face of objectively overwhelming contrary evidence. I take thought disorder to be the sustained failure to see things as they are under conditions where the relevant facts are evident.

Hence, *methodological thought disorder* is the sustained failure to cognize relatively obvious methodological facts. It is well known that many psychologists are ignorant of important methodological facts and their methodological thinking is often erroneous (e.g. Rosnow & Rosenthal, 1989; Zuckerman, Hodgins, Zuckerman & Rosenthal, 1993). That, itself, is sufficient cause for deep concern and a subject worthy of serious scientific research. I am interested, however, not so much in methodological ignorance and error amongst psychologists *per se*, as in the fact of *systemic* support for these states in circumstances where the facts are easily accessible. Behind psychological research exists an *ideological support structure*. By this I mean a discipline-wide, shared system of beliefs which, while it may not be universal, maintains both the dominant methodological practices and the content of the dominant methodological educational programmes. This ideological support structure is manifest in three ways: in the contents of textbooks; in the contents of methodology courses; and in the research programmes of psychologists. In the case of measurement in psychology this ideological support structure works to prevent psychologists from recognizing otherwise accessible methodological facts relevant to their research. This is not then a psychopathology of any individual psychologist. The pathology is in the social movement itself, i.e. within modern psychology.

5.2 *The systemic character of methodological thought disorder in psychological measurement*

Stevens did not only give modern psychology his definition of measurement. He also gave it his theory of scales of measurement (e.g. Stevens, 1946, 1951, 1959). As with his definition of measurement, this theory has been largely absorbed into the

collective wisdom of psychology and most psychological researchers and students are familiar with its details. One thing that is clear from this theory is that postulating quantitative relations between attributes presumes, at least, what he called, 'interval scale' measurement. This is why most quantitative theories in psychology, if they pay heed to the issue at all, stipulate this level of measurement (in Stevens' sense). The following comment is typical:

> The level of measurement most often specified in mental test theory is *interval measurement*, which yields an *interval scale*. This scale presupposes the ordering property of the ordinal scale but, in addition, specifies a one-to-one correspondence between the elements of the behavioral domain and the real numbers, with only the zero point and the unit of measurement being arbitrary. Such a scale assigns meaning not only to scale values and their relative order but also to differences of scale values (Lord & Novick, 1968, p. 21).

Now, if an interval scale requires an isomorphism between elements of the behavioural domain (by which, I take it, Lord and Novick simply mean the psychological attribute) and the real numbers, then it should be obvious that an additive structure within the psychological attribute is being presumed. That is, Stevens' theory of measurement scales (a theory widely accepted within psychology) together with the fact that quantitative theories require at least interval scale measurement (a position implicitly endorsed by most quantitative psychologists) entails the conclusion that the relevant attributes must have a structure similar in some respects to that of the real number system. (Strictly speaking, with an interval scale, it is *differences* between magnitudes of the attribute that are quantitative.) However, while the majority of psychologists accept these two premises, something of a general, systemic nature within psychology prevents them from seeing the conclusion entailed.

Now, of course, grasping this implication does not involve recognizing the precise character of quantitative attributes. Quantitative science existed for thousands of years before the structure of the real numbers was finally articulated (Dedekind, 1872), enabling quantitative structure itself to be correctly characterized (Hölder, 1901). However, the nature of the structure of the real number system is now described in the most elementary of university algebra texts (e.g. Birkhoff & MacLane, 1965) and there is an extensive mathematical and philosophical literature available on the structure of quantitative attributes (e.g. Behrend, 1953, 1956; Mundy, 1987; Nagel, 1932; Suppes, 1951; Swoyer, 1987; Whitney, 1968*a*, *b*), the relevant portions of which have been available in specifically psychological literature since Weitzenhoffer (1951) (see also Suppes & Zinnes, 1963). That is, the relevant information has been available in the psychological literature on measurement theory, even if not in the literature on psychometrics, since the time Stevens' definition was first proposed. Thus, other things being equal, the interested psychologist could easily have gleaned the relevant facts and, having done so, recognized the contingent character of the hypothesis that a psychological attribute is quantitative (i.e. at least measurable on an interval scale). Then the interested psychologist who happened to be also a committed empiricist, would have recognized that claims to measure such a psychological attribute in the absence of independent evidence for its quantitative structure have, like resorts to mere postulation in any area of science, all of 'the advantages of theft over honest toil'

(Russell, 1920; as quoted by Stevens, 1951, p. 36). Thus, the interested empirical psychologist would have come to see that Stevens' definition of measurement is nonsense and the neglect of quantitative structure a serious omission by quantitative psychologists. Of course, some psychologists did follow something like this route and periodically critiques of Stevens' definition were published (e.g. Ross, 1964; Rozeboom, 1966). If some psychologists were able to travel this route, why not all? The widespread acceptance of Stevens' definition within quantitative psychology created an intellectual environment in which measurement seemed easily attainable without travelling this route. This prevented the recognition of these otherwise evident facts. That is, we are dealing with a case of thought disorder, rather than one of simple ignorance or error and, in this instance, these states are sustained systemically by the almost universal adherence to Stevens' definition and the almost total neglect of any other in the relevant methodology textbooks and courses offered to students.

The conclusion that follows from this history, especially that of the last five decades, is that systemic structures within psychology prevent the vast majority of quantitative psychologists from seeing the true nature of scientific measurement, in particular the empirical conditions necessary for measurement. As a consequence, number-generating procedures are consistently thought of as measurement procedures in the absence of any evidence that the relevant psychological attributes are quantitative. Hence, within modern psychology a situation exists which is accurately described as systemically sustained methodological thought disorder.

5.3 *Paradigms of measurement*

The traditional view of scientific measurement and the view represented by Stevens' definition are different *paradigms* in Kuhn's (1970) sense (see also Michell, 1986). However, this does not mean that there is no basis for making an informed and rational choice between them. The operationism behind Stevens' definition contradicts the realist view that the subject matter of science is logically independent of the observer. If the subject matter of science (the quantitative attributes studied, for example) is constituted by the operations used to study it, as Stevens clearly thought was the case with his psychophysical scales, then there is no *it* to study. Science simply reduces to the study of our operations and cannot be construed as the study of an independently existing world whose secrets we penetrate via these operations. Only a realist view does justice to the concept of scientific discovery.

On this basis then, the widespread acceptance of Stevens' definition within psychology is an aberration for it does not mesh with an empirical realist view of science. It was accepted within psychology, not because psychologists were converted to the subjectivism of Bridgman's operationism, but because at the time it seemed the only way forward. Stevens' definition seemed to justify the quantitative practices that had developed within psychology. Such a misperception could only have been made in the first place and sustained for half a century because psychologists, generally, remain ignorant about the logic of science and, in particular, about the logic of quantification.

5.4 *The logic of science and the category of quantity*

This ignorance is due, not only to the systemic cause detailed above, but also, it should be said, to a loss of intellectual nerve on the part of philosophers of science. For whatever reason, many have been mesmerized by the relativism deriving from Kuhn (1970), Feyerabend (1975) and others and nervous of asserting a logic of science that does justice to its enormous achievements, a point also noted elsewhere (e.g. Stove, 1982; Theocharis & Psimopoulos, 1987). This vacuum has given psychologists the intellectual space to utilize Stevens' definition as a convenient rationalization without criticism from philosophers of science. A final cause of this ignorance is the fact that methodological education within psychology not only reproduces Stevens' definition but, also, has been maintained at a particularly low level by the social and managerial structures regulating academic psychology (Aiken, West, Sechrest & Reno, 1990).

If psychologists are to be re-educated in the logic of science then the concept of quantity needs to be seen as one of science's fundamental categories and the manner in which both the concept of measurement and its logic unfolds from this concept emphasized. The conceptual and mathematical foundation upon which such re-education could be based has been developed this century, from Hölder (1901) to Luce *et al.* (1990), in ways that are genuinely insightful and revolutionary. This body of mathematical and philosophical knowledge provides an opportunity for quantitative psychologists to remake their discipline as a science rather than a pretence. A narrowly conceived empiricism that ignores relevant conceptual and philosophical issues has here been shown to be intellectually bankrupt. This chapter in the history of psychology shows that

> if the work of inquiry is to be carried on, it must be *at once* scientific and philosophic, that if, in particular, the scientist is not philosophic, he will fall into confusions, he will rebuff philosophic criticism—he will lack a theory of categories, of sorts of problem, of 'method'—especially he will be carried away by practical interests, by interest in producing something or implementing a programme instead of in finding something out (Anderson, 1962, p. 183).

Acknowledgements

Research for this paper was done in 1995 while at Centrum voor Mathematische Psychologie en Psychologische Methodologie, Department Psychologie, Katholieke Universiteit Leuven, Belgium. I am grateful to the Board of Administration of KUL for appointing me Visiting Professor and to Professors Luc Delbeke and Paul De Boeck for arranging my appointment and for generous help while there. I also thank the University of Sydney for granting the period of study leave from normal duties that made this visit possible. This research has been supported financially by ARC Grants from the Federal Government of Australia and I gratefully acknowledge this assistance. Valuable help was received from my research assistants, Fiona Hibberd and Kate Toms. I would also like to thank both the reviewers and editors of this journal for their constructive comments. Other useful comments were given by Dr Agnes Petocz and the members of the Psychology IV, Conceptual Foundations of Quantitative Methods Seminar, University of Sydney.

References

Adams, E. W. & Fagot, R. F. (1959). A model of riskless choice. *Behavioral Science*, **4**, 1–10.

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

Adler, H. E. (1980). Vicissitudes of Fechnerian psychophysics in America. In R. W. Rieber & K. Salzinger (Eds), *Psychology: Theoretical–historical Perspectives*, pp. 11–23. New York: Academic Press.

Aiken, L. S., West, S. G., Sechrest, L. & Reno, R. R. (1990). Graduate training in statistics, methodology, and measurement in psychology. *American Psychologist*, **45**, 721–734.

Anderson, J. (1962). *Studies in Empirical Philosophy*. Sydney: Angus & Robertson.

Bartlett, R. J. (1940). Measurement in psychology. *Advancement of Science*, **1**, 422–441.

Beck, J. & Shaw, W. A. (1967). Ratio-estimations of loudness-intervals. *American Journal of Psychology*, **80**, 59–65.

Beckwith, T. G. & Buck, N. L. (1961). *Mechanical Measurements*. Reading, MA: Addison-Wesley.

Behrend, F. A. (1953). A system of independent axioms for magnitudes. *Journal and Proceedings of the Royal Society of NSW*, **87**, 27–30.

Behrend, F. A. (1956). A contribution to the theory of magnitudes and the foundations of analysis. *Mathematische Zeitschrift*, **63**, 345–362.

Bergmann, G. & Spence, K. W. (1944). The logic of psychophysical measurement. *Psychological Review*, **51**, 1–24.

Birkhoff, G. & MacLane, S. (1965). *A Survey of Modern Algebra*. New York: Macmillan.

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

Boring, E. G. (1921). The stimulus-error. *American Journal of Psychology*, **32**, 449–471.

Bostock, D. (1979). *Logic and Arithmetic*, vol. 2, *Rational and Irrational Numbers*. Oxford: Clarendon Press.

Bridgman, P. W. (1927). *The Logic of Modern Physics*. New York: Macmillan.

Bridgman, P. W. (1950). *Reflections of a Physicist*. New York: Philosophical Library.

Brower, D. (1949). The problem of quantification in psychological science. *Psychological Review*, **56**, 325–333.

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.

Brown, J. F. (1934). A methodological consideration of the problem of psychometrics. *Erkenntnis*, **4**, 46–61.

Brown, W. (1913). Are the intensity differences of sensation quantitative? IV. *British Journal of Psychology*, **6**, 184–189.

Burnet, J. (1955). *Greek Philosophy: Thales to Plato*. London: Macmillan.

Campbell, N. R. (1920). *Physics, The Elements*. Cambridge: Cambridge University Press.

Campbell, N. R. (1928). *An Account of the Principles of Measurement and Calculation*. London: Longmans, Green.

Cattell, J. McK. (1890). Mental tests and measurements. *Mind*, **15**, 373–380.

Cattell, J. McK. (1893). Mental measurement. *Philosophical Review*, **2**, 316–332.

Cattell, J. McK. (1904). The conceptions and methods of psychology. *Popular Science Monthly*, pp. 176–186.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, **40**, 153–193.

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

Clifford, G. J. (1968). *Edward L. Thorndike: The Sane Positivist*. Middletown, CT: Wesleyan University Press.

Clifford, W. K. (1882). *Mathematical Papers*. London: Macmillan.

Comrey, A. L. (1950). An operational approach to some problems in psychological measurement. *Psychological Review*, **57**, 217–228.

Cook, A. (1994). *The Observational Foundations of Physics*. Cambridge: Cambridge University Press.

Coombs, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological Review*, **57**, 145–158.

Coombs, C. H., Raiffa, H. & Thrall, R. M. (1954). Some views on mathematical models and measurement theory. *Psychological Review*, **61**, 132–144.

Crombie, A. C. (1994). *Styles of Scientific Thinking in the European Tradition*. London: Duckworth.

Cureton, E. E. (1946). Quantitative psychology as a rational science. *Psychometrika*, **11**, 191–196.

Dawes Hicks, G. (1913). Are the intensity differences of sensation quantitative? II. *British Journal of Psychology*, **6**, 155–174.

Dedekind, R. (1872). *Stetigkeit und Irrationalzahlen*. Braunschweig: Vieweg.

Debreu, G. (1960). Topological methods in cardinal utility theory. In K. J. Arrow, S. Karlin & P. Suppes (Eds), *Mathematical Methods in the Social Sciences, 1959*, pp. 16–26. Stanford, CA: Stanford University Press.

Ellis, B. (1966). *Basic Concepts of Measurement*. Cambridge: Cambridge University Press.

Fechner, G. T. (1860). *Elemente der psychophysik*. Leipzig: Breitkopf & Hartel. (English translation by H. E. Adler, *Elements of Psychophysics*, vol. 1, D. H. Howes & E. G. Boring (Eds). New York: Holt, Rinehart & Winston.)

Fechner, G. T. (1887). Über die psychischen Massprincipien und das Weber'sche Gesetz. *Philosophische Studien*, **4**, 161–230. (English translation of pp. 178–198 by S. Scheerer (1987). My own viewpoint on mental measurement. *Psychological Research*, **49**, 213–219.)

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). Final report of the committee appointed to consider and report upon the possibility of quantitative estimates of sensory events. *Report of the British Association for the Advancement of Science*, **2**, 331–349.

Feyerabend, P. (1975). *Against Method: Outline of an Anarchistic Theory of Knowledge*. London: New Left Books.

Fraser, C. O. (1980). Measurement in psychology. *British Journal of Psychology*, **71**, 23–34.

Freyd, M. (1926). What is applied psychology? *Psychological Review*, **33**, 308–314.

Gage, F. H. (1934*a*). An experimental investigation of the measurability of auditory sensation. *Proceedings of the Royal Society, Series B, Biological Sciences*, **116**, 103–122.

Gage, F. H. (1934*b*). An experimental investigation of the measurability of visual sensation. *Proceedings of the Royal Society, Series B, Biological Sciences*, **116**, 123–138.

Garner, W. R. (1954). Context effects and the validity of loudness scales. *Journal of Experimental Psychology*, **48**, 218–224.

Gigerenzer, G. & Strube, G. (1983). Are there limits to binaural additivity of loudness? *Journal of Experimental Psychology, Human Perception and Performance*, **9**, 126–130.

Green, B. F. (1954). Attitude measurement. In G. Lindzey (Ed.), *Handbook of Social Psychology*, vol. 1, pp. 335–369. Reading, MA: Addison-Wesley.

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

Gulliksen, H. (1946). Paired comparisons and the logic of measurement. *Psychological Review*, **53**, 199–213.

Hacking, I. (1983). *Representing and Intervening*. Cambridge: Cambridge University Press.

Hardcastle, G. L. (1995). S. S. Stevens and the origins of operationism. *Philosophy of Science*, **62**, 404–424.

Herbart, J. F. (1816). *Lehrbuch zur Psychologie*. (English translation by M. K. Smith, 1897, *A Textbook in Psychology*. New York: Appleton.)

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.

Hornstein, G. A. (1988). Quantifying psychological phenomena: Debates, dilemmas, and implications. In J. G. Morawski (Ed.), *The Rise of Experimentation in American Psychology*. New Haven, CT: Yale University Press.

Hull, C. L. (1943). *Principles of Behavior*. New York: Appleton-Century-Crofts.

James, W. (1890). *The Principles of Psychology*, vol. 1. London: Macmillan.

Jerrard, H. G. & McNeill, D. B. (1992). *Dictionary of Scientific Units*. London: Chapman & Hall.

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

Kant, I. (1786). *Metaphysical Foundations of Natural Science*. (J. Ellington transl., 1970.) Indianapolis, IN: Bobbs-Merrill.

Krantz, D. H. (1972). A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology*, **9**, 168–199.

Krantz, D. H., Luce, R. D., Suppes, P. & Tversky, A. (1971). *Foundations of Measurement*, vol. 1. New York: Academic Press.

Kuhn, T. (1970). *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.

Levelt, W. J. M.,. Riemersma, J. B. & Bunt, A. A. (1972). Binaural additivity in loudness. *British Journal of Mathematical and Statistical Psychology*, **25**, 1–68.

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.

Luce, R. D. (1987). Measurement structures with Archimedean ordered translation groups. *Order*, **4**, 165–189.

Luce, R. D. (1990). 'On the possible psychophysical laws' revisited: Remarks on cross-modality matching. *Psychological Review*, **97**, 66–77.

Luce, R. D., Krantz, D. H., Suppes, P. & Tversky, A. (1990). *Foundations of Measurement*, vol. 3. San Diego, CA: Academic Press.

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

McCormack, T. J. (1922). A critique of mental measurements. *School and Society*, **15**, 686–692.

Massey, B. S. (1986). *Measures in Science and Engineering: Their Expression, Relation and Interpretation*. Chichester: Ellis Horwood.

Maxwell, J. C. (1891). *A Treatise on Electricity and Magnetism*. London: Constable.

Meier, S. T. (1994). *The Chronic Crisis in Psychological Measurement and Assessment: A Historical Survey*. San Diego, CA: Academic Press.

Michell, J. (1986). Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin*, **100**, 398–407.

Michell, J. (1990). *An Introduction to the Logic of Psychological Measurement*. Hillsdale, NJ: Erlbaum.

Michell, J. (1993). The origins of the representational theory of measurement: Helmholtz, Hölder, and Russell. *Studies in History and Philosophy of Science*, **24**, 185–206.

Michell, J. (1994). Numbers as quantitative relations and the traditional theory of measurement. *British Journal for the Philosophy of Science*, **45**, 389–406.

Michell, J. (in press). Bertrand Russell's 1897 critique of the traditional theory of measurement. *Synthese*.

Mundy, B. (1987). The metaphysics of quantity. *Philosophical Studies*, **51**, 29–54.

Myers, C. S. (1913). Are the intensity differences of sensation quantitative? I. *British Journal of Psychology*, **6**, 137–154.

Nafe, J. P. (1942). Toward the quantification of psychology. *Psychological Review*, **49**, 1–18.

Nagel, E. (1932). Measurement. *Erkenntnis*, **2**, 313–333.

Narens, L. (1985). *Abstract Measurement Theory*. Cambridge, MA: MIT Press.

Narens, L. (1996). A theory of ratio magnitude estimation. *Journal of Mathematical Psychology* **40**, 109–129.

Narens, L. & Luce, R. D. (1993). Further comments on the "nonrevolution" arising from axiomatic measurement theory. *Psychological Science*, **4**, 127–130.

Newman, E. B. (1974). On the origin of 'scales of measurement'. In H. R. Moskowitz *et al.* (Eds) *Sensation and Measurement*, pp. 137–145. Dordrecht, Holland: Reidel.

Passmore, J. (1957). *A Hundred Years of Philosophy*. London: Duckworth.

Perloff, R. (1950). A note on Brower's 'the problem of quantification in psychological science'. *Psychological Review*, **57**, 188–192.

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.

Rosnow, R. L. & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, **44**, 1276–1284.

Ross, S. (1964). *Logical Foundations of Psychological Measurement*. Copenhagen: Munksgaard.

Rozeboom, W. W. (1966). Scaling theory and the nature of measurement. *Synthese*, **16**, 170–233.

Russell, B. (1897). On the relations of number and quantity. *Mind*, **6**, 326–341.

Russell, B. (1903). *Principles of Mathematics*. Cambridge: Cambridge University Press.

Russell, B. (1920). *Introduction to Mathematical Philosophy*. New York: Macmillan.

Sena, L. A. (1972). *Units of Physical Quantities and their Dimensions*. Moscow: Mir.

South, E. B. (1937). *An Index of Periodical Literature on Testing, 1921–1936*. New York: The Psychological Corporation.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, **15**, 201–293.

Spearman, C. (1937). *Psychology down the Ages*, vol. 1. London: Macmillan.
Stevens, S. S. (1935*a*). The operational definition of psychological terms. *Psychological Review*, **42**, 517–527.
Stevens, S. S. (1935*b*). The operational basis of psychology. *American Journal of Psychology*, **47**, 323–330.
Stevens, S. S. (1936*a*). Psychology: The propaedeutic science. *Philosophy of Science*, **3**, 90–103.
Stevens, S. S. (1936*b*). A scale for the measurement of a psychological magnitude: Loudness. *Psychological Review*, **43**, 405–416.
Stevens, S. S. (1939). Psychology and the science of science. *Psychological Bulletin*, **36**, 221–263.
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. (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.
Stove, D. C. (1982). *Popper and After: Four Modern Irrationalists*. Oxford: Pergamon.
Suppes, P. (1951). A set of independent axioms for extensive quantities. *Portugaliae Mathematica*, **10**, 163–172.
Suppes, P., Krantz, D. H., Luce, R. D. & Tversky, A. (1989). *Foundations of Measurement*, vol. 2. New York: Academic Press.
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.
Swoyer, C. (1987). The metaphysics of measurement. In J. Forge (Ed.), *Measurement, Realism and Objectivity*, pp. 235–290. Dordrecht: Reidel.
Terman, L. (1916). *The Measurement of Intelligence*. Boston, MA: Houghton Mifflin.
Terman, L. (1921). The status of applied psychology in the United States. *Journal of Applied Psychology*, **5**, 1–4.
Terman, L. (1924). The mental test as a psychological method. *Psychological Review*, **31**, 93–117.
Theocharis, T. & Psimopoulos, M. (1987). Where science has gone wrong. *Nature*, **329**, 595–598.
Thomas, L. G. (1942). Mental tests as instruments of science. *Psychological Monographs*, **54**, 1–87.
Thorndike, E. L. (1904). *An Introduction to the Theory of Mental and Social Measurements*. New York: Science Press.
Thorndike, E. L. (1923). Measurement in education. In G. M. Whipple (Ed.), *Twenty-first Year Book of the National Society for the Study of Education*, part 1, pp. 1–9. Bloomington, IN: Public School Publishing.
Thorndike, R. L. (1982). *Applied Psychometrics*. Boston, MA: Houghton Mifflin.
Thurstone, L. L. (1938). *Primary Mental Abilities*. Chicago, IL: University of Chicago Press.
Titchener, E. B. (1905). *Experimental Psychology: A Manual of Laboratory Practice*, vol. II. London: Macmillan.
Viteles, M. S. (1921). Tests in industry. *Journal of Applied Psychology*, **5**, 57–63.
von Kries, J. (1882). Über die Messung intensiver Grössen und über das sogenannte psychophysische Gesetz. *Vierteljahrsschrift für wissenschaftliche Philosophie*, **6**, 257–294.
Watt, H. J. (1913). Are the intensity differences of sensation quantitative? III. *British Journal of Psychology*, **6**, 175–183.
Weitzenhoffer, A. M. (1951). Mathematical structures and psychological measurements. *Psychometrika*, **16**, 387–406.
Whitney, H. (1968*a*). The mathematics of physical quantities, I: Mathematical models for measurement. *American Mathematical Monthly*, **75**, 115–138.
Whitney, H. (1968*b*). The mathematics of physical quantities, II: Quantity structures and dimensional analysis. *American Mathematical Monthly*, **75**, 227–256.
Zuckerman, M., Hodgins, H. S., Zuckerman, A. & Rosenthal, R. (1993). Contemporary issues in the analysis of data: A survey of 551 psychologists. *Psychological Science*, **4**, 49–53.

Zwislocki, J. J. (1983). Group and individual relations between sensation magnitudes and their numerical estimates. *Perception and Psychophysics*, **33**, 460–468.

Received 16 November 1995; revised version received 21 March 1996

Appendix I

The survey of texts was done in the Psychology Library at the Katholieke Universiteit Leuven, Belgium in June 1995. The books surveyed were:

Allen, M. J. & Yen, W. M. (1979). *Introduction to Measurement Theory*. Monterey, CA: Brooks/Cole.
Andreas, B. G. (1960). *Experimental Psychology*. New York: Wiley.
Black, J. A. & Champion, D. J. (1976). *Methods and Issues in Social Research*. New York: Wiley.
Borgatta, E. F. & Bohrnstedt, G. W. (1981). Levels of measurement once over again. In G. W. Bohrnstedt & E. F. Borgatta (Eds), *Social Measurement*. Beverly Hills, CA: Sage.
Calfee, R. C. (1975). *Human Experimental Psychology*. New York: Holt, Rinehart & Winston.
Cliff, N. (1973). Psychometrics. In B. J. Wolman (Ed.), *Handbook of General Psychology*. Englewood Cliffs, NJ: Prentice-Hall.
Coombs, C. H. (1953). Theory and methods of social measurement. In L. Festinger & D. Katz (Eds), *Research Methods in the Behavioral Sciences*. New York: Holt, Rinehart & Winston.
Crano, W. D. & Brewer, M. B. (1973). *Principles of Research in Social Psychology*. New York: McGraw-Hill.
Dominowski, R. L. (1980). *Research Methods*. Englewood Cliffs, NJ: Prentice-Hall.
Eagly, A. H. & Chaiken, S. (1993). *The Psychology of Attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich.
Eysenck, H. J., Würzburg, W. A. & Berne, R. M. (1972). *Encyclopedia of Psychology*. London: Search Press.
Francis, R. G. (1967). Scaling techniques. In J. T. Doby (Ed.), *An Introduction to Social Research*. New York: Appleton-Century-Crofts.
Galtung, J. (1967). *Theory and Methods of Social Research*. London: George Allen & Unwin.
Green, B. F. (1954). Attitude measurement. In G. Lindzey (Ed.), *Handbook of Social Psychology*. Reading, MA: Addison-Wesley.
Guilford, J. P. (1956). *Fundamental Statistics in Psychology and Education*. New York: McGraw-Hill.
Haimson, B. R. & Elfenbein, M. H. (1985). *Experimental Methods in Psychology*. New York: McGraw-Hill.
Harris, C. W. (1960). *Encyclopedia of Educational Research*. New York: Macmillan.
Hays, W. L. (1967). *Quantification in Psychology*. Belmont: Brooks & Cole.
Hilgard, E. R. (1953). *Introduction to Psychology*. New York: Harcourt, Brace & World.
Kantowitz, B. H. & Roediger, H. L. (1978). *Experimental Psychology*. Chicago, IL: Rand McNally.
Kaplan, A. (1964). *The Conduct of Inquiry*. San Francisco, CA: Chandler.
Kerlinger, F. N. (1966). *Foundations of Behavioral Research*. New York: Holt, Rinehart & Winston.
Kurtz, K. H. (1965). *Foundations of Psychological Research*. Boston, MA: Allyn & Bacon.
Lemke, E. & Wiersma, W. (1976). *Principles of Psychological Measurement*. Chicago, IL: Rand McNally.
Lemon, N. (1973). *Attitudes and their Measurement*. London: Batsford.
Lord, F. M. & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
Manning, S. A. & Rosenstock, E. H. (1968). *Classical Psychophysics and Scaling*. New York: McGraw-Hill.
Marlowe, L. (1971). *Social Psychology*. Boston, MA: Holbrook Press.
Meyers, L. S. & Grossen, N. E. (1974). *Behavioral Research*. San Francisco, CA: Freeman.
Miller, G. A. (1962). *Psychology: The Science of Mental Life*. New York: Harper & Row.
Morgan, C. T. & King, R. A. (1966). *Introduction to Psychology*. New York: McGraw-Hill.
Newcomb, T. M., Turner, R. F. & Converse, P. E. (1965). *Social Psychology*. New York: Holt, Rinehart & Winston.
Nunnally, J. C. (1967). *Psychometric Theory*. New York: McGraw-Hill.

Peck, D. F. & Shapiro, C. M. (1990). *Measuring Human Problems*. New York: Wiley.
Phillips, B. S. (1971). *Social Research*. New York: Macmillan.
Plutchik, R. (1968). *Foundations of Experimental Research*. New York: Harper & Row.
Sartain, A. Q., North, A. J., Strange, J. R. & Chapman, H. M. (1973). *Psychology: Understanding Human Behavior*. New York: McGraw-Hill.
Selltiz, C., Jahoda, M., Deutsch, M. & Cook, S. W. (1959). *Research Methods in Social Relations*. New York: Holt, Rinehart & Winston.
Shaw, M. E. & Wright, J. M. (1967). *Scales for the Measurement of Attitudes*. New York: McGraw-Hill.
Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.
Simon, J. L. (1969). *Basic Research Methods in Social Science*. New York: Random House.
Sjoberg, G. & Nett, R. (1968). *A Methodology for Social Research*. New York: Harper & Row.
Sommer, B. & Sommer, R. (1991). *A Practical Guide to Behavioral Research*. New York: Oxford University Press.
Wiersma, W. & Jurs, S. G. (1985). *Educational Measurement and Testing*. Boston, MA: Allyn & Bacon.

Appendix II

Put as succinctly as possible, *measurement* **is the numerical estimation of the ratio of a magnitude of a quantitative attribute to a unit of the same attribute.**
The concept of measurement is embedded within a matrix of closely related concepts. There is no adequate definition of it that avoids implicating them, as well. The associated concepts used here are briefly explained as follows.

Attribute. An attribute is a range of properties or relations that may vary from instance to instance. E.g. length is an attribute of objects, different objects often having different lengths; so is sex, some creatures being female, others male; nationality is another example, some people being British, some French, etc.

Quantitative attribute. A quantitative attribute (or quantity) is an attribute the instances of which are related to one another both ordinally and additively. One version of (continuous) quantitative structure is given by Hölder's (1901) axioms, another is given in section 1.1 of this paper. Not all attributes are quantitative. E.g. length is quantitative, but neither sex nor nationality is.

Magnitude. A magnitude is a specific instance of a quantity. Thus, each instance of length (say, the length of this page) is a magnitude of the quantity, length.

Ratio. As intended in this context, a ratio is a special kind of relation holding between magnitudes of the same quantity. The ratio of one magnitude of a quantity to another is the size of the first relative to the second. Thus, ratios are relative magnitudes. The most useful way to express a ratio is as a number. E.g. the ratio of the length of a cricket pitch to one yard is 22.

Units. A unit is a more or less arbitrarily selected magnitude of a quantity, singled out to be the instance against which any other is to be compared. If the unit is known, then expressing the magnitude of any other instance of the quantity relative to the unit means that the magnitude of that instance is defined and, so, known via the unit. E.g. one widely used unit of length is the metre.

Numerical estimation. In the first instance, a numerical estimation or number specifies how many? or how much?. E.g. knowing that there is some whole number, say, three, books on the desk is knowing that there is a book, another and one more at that location. The range of possible answers to the question, how many?, is the sequence of whole (or natural) numbers, the meaning of each being definable via the concepts of one and one more than. Using the natural numbers, the real numbers may be specified. They answer the question, how much?. It follows from the axioms for continuous quantity that some multiple (say, n) of an instance of some quantity (call it a) is always less than, equal to, or greater than any other multiple (m) of another instance of that quantity, b. If less than, then the ratio of a to b is less than the ratio of m to n (i.e. $a/b < m/n$); if equal to, then $a/b = m/n$; and if greater than, then $a/b > m/n$. Thus, each ratio of magnitudes has a location relative to each ratio of natural numbers and, so, is given by a real number (i.e. the real number specified by the class of all ratios of natural numbers equal to or less than it). Thus, the ratio of each instance of a quantity to the unit selected is expressed by a unique numerical value. In general, measurement procedures only permit the (approximate) estimation of such unique numerical values.

Copyright of British Journal of Psychology is the property of British Psychological Society and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.