

Making the Myth of Mental Measurement

THE QUANTITATIVE IMPERATIVE

Quantitative concepts possess power, simplicity, and beauty. So the temptation to suppose that the fundamental structure of all natural processes is quantitative is strong. It has been thus since science began. Pythagoras was the first to yield: "All things are made of numbers," he is said to have taught. Plato ingeniously developed the same theme in the *Timaeus*: all things were composed out of the four basic elements, earth, fire, air, and water; they, in turn, were thought to be made up of certain regular polyhedra; the polyhedra, he said, were made of triangles; and, as is well known, triangles can be reduced to lines and angles and these to numbers. Such flights of mathematical fancy were to shape the course of science.

Aristotle disagreed with this reduction of all things to quantity. Naturally, he recognized that there were quantities (numbers, lengths, areas, volumes, angles, weights, etc.). But alongside these he thought that there were *qualities* as well. These were the nonquantitative properties of things, like color or temperature, so he said. The distinction he took to be an observational one. Properties that were quantitative possessed an additive structure. Those that were qualities did not. Imposed upon this distinction was his prejudice that the *essential* properties of things were qualities, not quantities. This affected his theorizing and he developed a fundamentally qualitative physics.

In medieval Europe interest in science reawakened about the 11th century. This coincided with the rediscovery of Aristotle's scientific works, which had been preserved by the Arabs after the barbarian invasions. The struggle between quantitative and qualitative conceptions of nature resumed. From the 12th to the 17th centuries Aristotle's nonquantitative physics dominated scientific thinking.

But the Platonic-Pythagorean tradition lived on. A portion of the *Timaeus* had been preserved in the Augustinean monasteries. Its influence sparked the beginning of modern quantitative science. The 13th-century English bishop, Robert Grosseteste, not only rediscovered the experimental method; he proclaimed the fundamentally quantitative structure of the universe: "All causes of natural effects can be discovered by lines, angles and figures, and in no other way can the reason for this action possibly be known" (*De Lineis, Angulis et Figuris*, cf. J. McEvoy (1982), p. 168).

Galileo echoed these sentiments four centuries later when he wrote that the "book of nature" (by which he meant the universe itself) "... is written in mathematical language, and the letters are triangles, circles and other geometrical figures, without which means it is humanly impossible to comprehend a single word" (*Il Saggiatore*, cf. S. Drake (1957) pp. 237-238). These Platonic sentiments were shared by his great contemporaries, Kepler and Descartes. But, more than anything else, it was the victory of Galileo's physics over its Aristotelean rival that caused this theme to dominate the future of European science. Quantitative science had triumphed over nonquantitative.

Out of this success was born the quantitative imperative, encapsulated in the quote from Galileo that heads this section. Its impetus was strengthened by the work of Newton and it culminated in the 19th century with the successful extension of quantitative theories to the study of most physical phenomena, including electricity, magnetism, and heat. Again the philosophers taught it. Kant wrote that "... a doctrine of nature will contain only so much science proper as there is applied mathematics in it" (1786, p. 7). The 19th-century British philosopher of science, W. S. Jevons, concluded that there was "... no apparent limit to the ultimate extension of quantitative science" (1873, p. 274). Even a philosopher as unlikely as Nietzsche considered that "... knowledge has become scientific to the extent that it is able to employ number and measure" (1968, p. 378).

However, it was that great 19th-century physicist, Lord Kelvin, who gave it its most forthright proclamation:

I often say that when you can measure what you are speaking about and express it in numbers you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts advanced to the stage of *science*, whatever the matter may be. (W. Thomson, 1891, pp. 80-81)

This saying became a favorite of the quantitative movement and versions of it abound (cf. Eysenck, 1973; Hacking, 1983; Kuhn, 1961; Merton, Sills, & Stigler, 1984; Pearson, 1978; Yule & Kendall, 1911).

Into this intellectual milieu modern psychology emerged in 1860 with the work of G. Fechner. No discipline can free itself from the methodological presuppositions of its time. Modern psychology was no exception. As a 19th

century invention it conformed to 19th century conceptions of the nature of science. It conformed to Kelvin's dictum that science had to be quantitative. Indeed, the mere fact that modern psychology is taken to have begun with Fechner's contribution confirms this judgment. He began his scientific career as a physicist and established his reputation measuring direct currents. Later, he became preoccupied with issues of a psychological kind, and sought to establish the reality of mental phenomena, attempting to measure the intensity of sensations. Doing so he developed psychology's first quantitative methodology. Without this contribution, Boring (1929) wrote, "... there would ... have been little of the breath of science in the experimental body, for we hardly recognize a subject as scientific if measurement is not one of its tools" (p. 286). So it was "... as a consequence of Fechner's contribution that psychology advanced from prescience to science" (H. E. Adler, 1980, p. 13).

Fechner was not the first to think of measuring psychological variables. That honor belongs, I think, to the 14th century scholar, Nicole Oresme. Later, J. F. Herbart (1816) was to propose a quantitative psychological theory. But Fechner proposed *methods of measurement* and that is why modern psychologists date their science from his work.

In the minds of the founding fathers the issue was paramount. Francis Galton wrote, "... until the phenomena of any branch of knowledge have been submitted to measurement and number, it cannot assume the status and dignity of a science" (1879, p. 147). His one-time assistant and one of America's first professors of psychology, James McKeen Cattell, wrote, "Psychology cannot attain the certainty and exactness of the physical sciences, unless it rests on a foundation of experiment and measurement" (1890, p. 373).

The founder of the factor analytic method, Charles Spearman, wrote that

... great as may be the potency of this [the experimental method], or of the preceding methods, there is yet another one so vital that, if lacking it, any study is thought by many authorities not to be scientific in the full sense of the word. This further and crucial method is that of measurement. (1937, p. 89).

The early American structuralist, E. B. Titchener, thought that there was "... no question but that, in some way or other, mental processes are measurable" (1905, pp. xxi-xxii). Kulpe, the founder of the Wurzburg school was equally confident that "... the universal validity of psychological results is guaranteed by the measurability of mental phenomena" (1895, p. 11). And Hermann Ebbinghaus, the first to attempt to measure the "higher mental processes" wrote, "The brilliant results produced in natural science by measurement and calculation suggested the idea that something similar might be done for psychology" (1908, p. 13).

Wilhelm Wundt, said by some to be the father of modern experimental psychology, felt it necessary to refute Kant's claim that psychology could not be a science because it could not be quantitative. Wundt concluded that, "Our inner life ...

has at least two dimensions which implies the general possibility of expressing them in mathematical form" (1873, p. 6). E.L. Thorndike, the American functionalist, attempted to justify the quantitative imperative with Pythagorean dogma (in an oft-quoted saying, cf. Eysenck, 1973; Johnson, 1936; & McCall, 1923): "Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quantity" (quoted in Eysenck, 1973, p. x). Freud, also, was victim to the same doctrine: he wrote that "... science recognizes only *quantities*" (1895, p. 309).

The early years of modern psychology's development are largely the history of psychophysical measurement (in the work of Fechner, Wundt, Titchener, and others); of mental chronometry (Donders); of the attempt to measure learning and memory (Ebbinghaus); and of the attempts to measure mental abilities (Galton, Spearman, and their followers).

Few who took seriously the claim that psychology was a science questioned the idea that it could be quantitative. Too much seemed to be at stake. Psychology's credentials as a science, both in the eyes of its founders and of the wider scientific community, depended on its development as a quantitative discipline.

One who did have a more relaxed attitude about this matter was Franz Brentano, the founder of Act Psychology. While going along with the prevailing view that psychology was, as a matter of fact, a quantitative science, he recognized the contingent nature of this situation. The general issue he saw as follows: "Mathematics appears to me necessary for the exact treatment of all sciences only because we now in fact find magnitudes in every scientific field. It there were a field in which we encountered nothing of the sort, exact description would be possible even without mathematics" (Brentano, 1874, p. 65).

Uncharacteristic as it is of its age, Brentano's view has an Aristotelian ring to it, as does much of his work. If, as Aristotle argued, the distinction between quantity and quality resides in the structure of the variable involved (quantitative variables being additive, qualitative ones being nonadditive) then it is possible that some sciences may be purely qualitative, precisely because the variables studied within them are qualitative variables. Measurement need not be a universal characteristic of science. Instead, what characterizes science is the method of observation. And it is only through observation that the structure of any variable can be known. The structure cannot be determined *a priori*. On this matter at least, Aristotle and Brentano were correct. The only way to decide whether or not the variables studied in any particular science are quantitative is to put that hypothesis to the test. This essential step is missing in the development of modern psychology and this fact has put the pursuit of psychological measurement on the defensive. In its turn this defensiveness has helped shape psychologists' thinking about psychological measurement. As will be shown, they have been led to adopt some curious views.

Until the 1930s the development of psychological measurement was relatively untroubled. Certain variables were taken to be quantitative and various methods were hypothesized to measure them. The presumption that these speculations

delivered the substance of measurement was not seriously challenged. Then in 1932 a committee was established by the British Association for the Advancement of Science, its purpose being to enquire into the status of psychophysical measurement. Some serious criticisms were made of these supposed methods of measurement and this, by implication, raised questions about the entire enterprise of psychological measurement. The impact on psychology was catastrophic. Psychologists responded by redefining the concept of measurement, in the end accepting a definition so inflated as to rule out none of their methods. That this devalued the concept was ignored, for at least it gave the appearance of conformity to the quantitative imperative.

Of course, such a response inhibited serious interest in the issue of psychological measurement. The irony is that such redefinition was not the only way to justify attempts at psychological measurement. The psychologists of that period were not only in the grip of the quantitative imperative, they were also deeply confused about the nature of measurement. Had this concept been properly understood the crisis could have been resolved satisfactorily. The source of confusion about the concept of measurement was ignorance about the history of the concept prior to the 20th century.

HISTORY OF THE CONCEPT OF MEASUREMENT

Psychologists are not alone in ignoring the classical treatments of the concepts of measurement and quantity. Modern treatments of the topic read as if serious thinking about measurement began with Campbell (1920), Helmholtz (1887), Holder (1901), and Russell (1903). Occasionally reference is made to an earlier, "traditional" view (e.g., Ellis, 1966, Fraser, 1980) but only to dismiss it, never to understand it. Such 20th century arrogance, with its ahistorical insinuation that no clear thinking was ever done prior to this century, is not compatible with a serious interest in a topic as ancient as measurement. Its origins cannot be dated but the ancient Greek mathematicians and philosophers thought deeply about the subject. Their views sustained the practice of measurement from their time until the late 19th century. The theories of measurement which then developed, the representational theories of Helmholtz, Russell, and Campbell, need to be understood as the outcomes of interactions between this classical theory and other influences. They should not be taken as either the first or last word on any aspect of measurement or quantity.

A coherent theory of measurement and quantity is to be found in the writings of Aristotle and Euclid. The following summary takes ideas from both writers, and I take it to be a fair appraisal of what these concepts meant then and, with some modification, continued to mean for more than 2,000 years. Indeed, the view of the modern layman and scientist uninfluenced by philosophy is still remarkably close to this ancient view.

Aristotle took quantity to be one of the fundamental categories of reality. He

divided it into discrete and continuous. The discrete quantities were the natural numbers (i.e., the numbers 1, 2, 3, . . .) by which he meant, not the elements of some formal system, but the properties common to certain classes or collections. The *number 2*, for example, was thought to be that which is common to all pairs of things. The continuous quantities Aristotle called *magnitudes*, by which he intended to designate variables like length, time, weight, volume, and so forth. What he thought distinguished quantities from qualities was the fact that any particular quantity (i.e., any natural number or any value of a magnitude, say, a particular length) is additively composed of parts. In the case of natural numbers the parts are *ones*. In the case of magnitudes the parts are smaller magnitudes of the same kind. For example, every length is additively composed of smaller lengths. Hence, magnitudes must be infinitely divisible and this is part of what Aristotle meant by calling them *continuous*. Pairs of quantities of the same kind (homogeneous quantities) stand in a relation of *ratio* (i.e., relative magnitude) to one another. For example, the magnitude of a yard relative to that of a foot is 3. More generally, let *A* and *B* be any two homogeneous magnitudes (e.g., 2 lengths) then the ratio of *A* to *B* (A/B) $\leq n/m$ (where *n* and *m* are natural numbers) if and only if $mA \leq nB$ (where mA , for example, is that magnitude entirely composed of *m* discrete parts each identical to *A*). Then two ratios of magnitudes, A/B and C/D , are equal if and only if for all natural numbers, *n* and *m*,

$$A/B < n/m \text{ if and only if } C/D < n/m,$$

$$A/B = n/m \text{ if and only if } C/D = n/m \text{ and}$$

$$A/B > n/m \text{ if and only if } C/D > n/m.$$

So not only was the concept of relative magnitude well understood but so was the idea of a ratio of nonnumerical quantities equalling or not equalling the ratio of two natural numbers. As was also well understood by the time of Aristotle and Euclid, some ratios of magnitudes equal ratios of numbers and some do not. Those that do not are incommensurable pairs of magnitudes (like the lengths of the side and the diagonal of a square). If $A/B = n/m$ then n/m is the *measure* of *A* in units of *B*. If *A* is incommensurate with *B* then the measure of *A* in units of *B* is only *approximated* by a numerical ratio, though as was known by Euclid, in principle it may be approximated to any degree of precision.

As it stands this theory copes admirably with all of the then recognized quantities. These are some of those that are now called *extensive quantities*. Length may be taken as the paradigm. The fact that lengths are additive, and therefore quantitative, is readily seen in the behavior of some objects possessing lengths. A pair of rigid, straight rods may be combined end to end in a straight line, the length of the joined rods equalling the sum of their separate lengths. In such a case the operation of combination reflects the additive structure of the length variable in a simple and direct fashion. This is the way with all extensive quantities. But it needs to be emphasized that the classical theory is about the

structure of quantitative variables and not about the behavior of objects possessing values of such variables. This aspect of the classical theory was not emphasized by the ancient philosophers and mathematicians. The only magnitudes they knew of were extensive and so there was little need to make the point.

However, modern science, in contrast to ancient, deals with many quantitative variables that are not extensive. Temperature is a good example. Combining two liquids does not generally produce a liquid whose temperature is the sum of their separate temperatures. In this case there is no operation that simply and directly reflects the additive structure of the quantity. Such quantities are called *intensive*. They differ from extensive quantities not in their inner, additive structure but only in their external relations to the behavior of objects manifesting them. These relations are less direct and more complex, hence, so is the evidence for these variables being quantitative. Despite this, the classical theory still describes the structure of the variable involved.

This capacity of the classical theory to cover both extensive and intensive quantities was first made explicit by the 14th century French scholar, Nicole Oresme. His insight sprang from what began as a theological problem. For some time prior to Oresme, theologians had argued over how they should conceptualize the fact that the virtue of charity could increase or decrease within the same person. Eventually the problem was generalized to that of whether *any* qualities admit of quantitative increase or decrease in intensity and if so, how this fact might be conceptualized. Oresme's solution was as follows:

Every measurable thing except number is imagined in the manner of continuous quantity. Therefore, for the mensuration of such a thing, it is necessary that points, lines and surfaces, or their properties be imagined. For in them (i.e., the geometrical entities), as the Philosopher has it, measure or ratio is initially found, while in other things it is recognized by similarity as they are being referred by the intellect to them (i.e. to geometrical entities). . . . Therefore, every intensity which can be acquired successively ought to be imagined by a straight line . . . For whatever ratio is found to exist between intensity and intensity, in relating intensities of the same kind, a similar ratio is found to exist between line and line and vice versa Therefore, the measure of intensities can be fittingly imagined as the measure of lines. (*De Configurationibus* I, i, pp. 165-167)

In this passage it is Oresme's suggestion that the quantitative structure of the length variable (lines and their properties) is the paradigm or model of all quantity, extensive and intensive alike. Indeed, to give it authority he attributes this suggestion to Aristotle ("the Philosopher"). According to this view, length is the model for measurement and to describe a variable as quantitative is to attribute to it a structure similar to that of length. Therefore, measures of any such variable may be conceptualized by analogy to lines and their ratios one to another. Oresme was thus enabled to speculate about the measurement of all manner of variables, many hitherto considered nonquantitative. These included not only physical

variables like velocity, but psychological ones as well, such as pleasure and pain. In so doing he became, in theory at least, the first psychometrician. His solution, obvious to us because of our facility with graphical methods, freed the classical concepts of magnitude and measurement from their ancient exclusive attachment to extensive magnitudes and broadened them to include the intensive ones as well. This proved important in the ensuing scientific revolution and its aftermath.

Considered in the light of this extension of the classical theory, the attempts at quantification made by the founding fathers of modern psychology were perfectly proper. To suggest that variables like mental ability or the intensity of sensations are quantitative is to speculate that they possess an additive structure akin to that of length. It is perfectly meaningful to speculate thus, even though the evidence supporting the speculation is thin. Furthermore, if procedures are suggested for the measurement of such variables then the hypothesis is that they *measure* them in the classical sense of that word. Of course, validation of these measurement procedures may be a long way off. The next step is to develop the means of testing these speculations and hypotheses. One cannot properly boast of measurement at this early stage. However, it is a necessary stage and progress comes in science through finding ways to test new hypotheses.

Now there is every reason to believe that the founders of modern psychology accepted the classical theory of measurement. Not that many of them explicitly defined the concept, the classical theory being taken too much for granted then. But those who did, did so in a way consistent with the classical theory. For example, Fechner (1860) wrote, "Generally the measurement of a quantity consists of ascertaining how often a unit quantity of the same kind is contained in it" (p. 38). E. B. Titchener (1905) defined it similarly: "When we measure in any department of natural science, we compare a given magnitude with some conventional unit of the same kind, and determine how many times the unit is contained in the magnitude" (p. xix). And the entry under "measurement" in Baldwin's (1902) *Dictionary of Philosophy and Psychology* states: "In order that a concept may be measured as a mathematical quantity without ambiguity, it is essential that the special attribute by which we measure it should be conceivable as made up of discrete parts, admitting of unambiguous comparison *inter se*, with respect to their quality or difference of magnitude" (p. 58).

So in their acceptance of the quantitative imperative, and in their speculations that their own special variables were measurable they presumed the classical theory of measurement and quantity. In doing so they courted quantity at the end of an era. The classical theory was soon to go.

The classical theory has it that ratios of magnitudes may equal (or be approximated by) ratios of numbers. Now magnitudes, like for example length, are *empirical* properties or relations. That is, they belong to the spatio-temporal world of experience. On the other hand, the empirical character of numbers has always been a matter of debate, even in ancient times. However the classical theory, in asserting the existence of relationships between ratios of magnitudes and ratios of numbers, implies that numbers must have an empirical character.

This could have been Aristotle's view (cf. J. Lear, 1982). Nevertheless, it was a view lacking wide support at the dawn of the 20th century. One factor causing its eclipse was the critique given by the German philosopher, G. Frege, of the empirical view. In his *System of Logic*, John Stuart Mill had proposed an empirical theory of number. Frege (1884) mercilessly attacked Mill's thesis. Since then few have followed Mill's or any other empirical theory of number. A second factor was the development during the 19th century of the theory of noneuclidean geometries. Their appearance was taken to imply that the science of mathematics generally possesses a nonempirical subject matter. Then late in the 19th and early in the 20th centuries nonempirical theories of number were proposed. The time was ripe for a new theory of measurement, one compatible with nonempiricism about numbers.

The second intellectual development weakening the classical theory was *verificationism*. This is the tendency to confuse the meaning of a proposition with the observational conditions that would verify it. Within the classical theory the fact that a variable is quantitative is distinguished from the observational evidence supporting that fact. For example, length and density are both taken to be quantitative variables (i.e., to possess additive structure), although the kinds of observations supporting each are quite different. In the case of length, the behavior of rigid rods directly reflects the hypothesized additive structure. In the case of density the evidence is less direct, being via the relation of density to mass and volume. There are, as will be shown in the following chapters, a number of different ways in which the hypothesis that a variable is quantitative may be tested. The laws that characterize the structure of a quantitative variable are what might be described as *uniformities of coexistence* (to borrow a term from J. S. Mill, 1843). In themselves they do not license any predictions, they simply state how values of the variable are related to one another. For example, one such law is the following proposition: $A + (B + C) = (A + B) + C$ (where A , B , and C are any three values of the variable and the relation $A + B = D$ means that value D is entirely composed of the two discrete parts A and B). Laws like this only license predictions in conjunction with other laws relating not only to the character of the objects involved but also to our capacity to operate on them and to make observations. Because these other laws may vary from one situation to another, even though the laws of quantity are invariant, there may be no limit to the number of different ways in which quantitative structure may be manifest in the behavior of objects.

Verificationists have little patience with such subtleties. In the context of measurement they attempt to reduce quantitative structure to a specific set of observations, most often to those that directly reflect it. That is, they tend to reduce the concept of quantity to that of extensive quantity, making the existence of the latter a necessary condition for all measurement.

The theory of measurement put forward by N. R. Campbell (1920) satisfied both of these tendencies. He distinguished two kinds of measurement: *fundamental* and *derived*. Fundamental measurement requires the existence of a physical

analogue of numerical addition, such as the combining of two rods end to end in a straight line. Given that this physical operation satisfies properties similar to those of numerical addition, numbers (or numerals) may be assigned to the objects so as to reflect their physical addition. For example, if rod c is the same length as rod a combined with rod b then the number assigned to c should equal the sum of the numbers assigned to a and b . In this way numbers and the relations between them may be used to represent empirical, non-numerical relations of order and addition. Thus Campbell's theory involves no commitment to numbers as empirical entities, as does the classical theory. It is indifferent to their ontological character. On this view numbers are a mere tool enabling "... the powerful weapon of mathematical analysis to be applied to the subject matter of science" (Campbell, 1920, pp. 267-268). Campbell's theory of fundamental measurement also satisfies the requirements of verificationism, for it ties quantity to fundamental measurability. To be a fundamental quantity a variable must be related to an operation of physical addition formally similar to the operation of numerical addition.

However, according to Campbell this is not the only way that quantity is manifest. In derived measurement the fact that a variable is quantitative is manifest through its relationship to fundamental quantities, thought Campbell. When a set of fundamental measures reveals a system of constants, then the constants may be taken as derived measures of another quantity. For example, relative to each different kind of material the ratio of mass to volume is a constant and this constant is taken to be a derived measure of that kind of material's density. Because derived measurement depends on fundamental measures (a derived measure being a numerical function of fundamental measures), for Campbell the concept of quantity is still tied to that of fundamental measurement and fundamental measurement is the first, necessary step along the road to quantification.

Campbell's theory had enormous influence, particularly on psychologists. From 1935 onwards a series of articles and books appeared within psychology (and associated disciplines) relating Campbell's theory to attempts at psychological measurement (c.f. Bergmann & Spence, 1944; Comrey, 1950; Johnson, 1936; McGregor, 1935; Reese, 1943; Smith, 1938; and Thomas, 1942). Initially, attempts were made to reconcile psychological procedures with Campbell's theory, but this was not possible. Already, in 1934, E.G. Boring had correctly followed the logic of Campbell's theory to the conclusion that "Psychology lacks true measurement ... because there are not additive units" (Newman, 1974, p. 138).

Eventually, all agreed that psychology lacked fundamental measurement in Campbell's sense because it possessed no additive operations for its supposedly quantitative variables. This conclusion was made quite public in the Final Report delivered by the committee of the British Association for the Advancement of Science. One of the psychologists on the committee wrote: "If all measurement must conform to the Laws of Measurement enunciated by Dr. Campbell, and, in particular, if the second law can only be satisfied by the physical juxtaposition

of equal entities, then sensation intensity cannot be measured" (Bartlett, 1940, p. 423). Of course, what was here concluded of psychophysical measurement applied equally to all psychological measurement. Without fundamental measurement there could be no derived measurement and so, according to Campbell's theory, psychology was without measurement. What is more, the task of locating analogues of numerical addition pertinent to psychological variables seemed hopeless and thus the prospects of psychological measurement appeared grim.

This dismal situation, combined with acceptance of the quantitative imperative appeared to place psychology in an intolerable position. To be a science in the full sense of the word psychological measurement was thought to be necessary, however, the requirements held to be necessary for measurement were absent from psychology and seemed unattainable. The obvious conclusion, nevertheless, was totally unthinkable. The tension was resolved by the American psychophysicist, S. S. Stevens. He had a special interest in the report of the British committee, because his *sone scale*, for the alleged measurement of sensations of loudness, had been singled out for examination. However, for a long time he had been considering the plight of attempts at psychological measurement. He was at Harvard University with E.G. Boring and so he was, no doubt, aware of the latter's critical stance based on Campbell's theory. In an autobiographical article Stevens wrote, "My own central problem throughout the 1930s was measurement, because the quantification of the sensory attributes seemed impossible unless the nature of measurement could be properly understood" (1974, p. 409). The theory of measurement that he came to propose owed much to Campbell's, but it also owed a lot to what was then a fledgling philosophical movement, *operationalism*.

Operationalism was a homespun, American philosophy of science, presented by the Nobel prize-winning physicist, P. W. Bridgman (1927). His concerns centered around the then recent revolution in physical theory, the replacement of 19th century physical theory (what is sometimes called "Newtonian" theory) by Einstein's theory of relativity. Bridgman hoped to show that the fault with "Newtonian" theory was the inadequate definition of its terms. His proposal was for a program of conceptual reform based on a system of what he called *operational definitions*. It was really a blend of verificationism and pragmatism. According to Bridgman, "... the true meaning of a term is to be found by observing what a man does with it, not by what he says about it" (1927, p. 7). Thus, the meaning of supposedly theoretical terms had to be reduced to observations (verificationism) and the relevant observations were those relating to scientific practice (pragmatism). His pivotal slogan was that, "In general we mean by any concept nothing more than a set of operations; the concept is synonymous with the corresponding set of operations" (1927, p. 5).

Operationalism's iconoclastic rejection of traditional concepts and methods and its promise of a fresh-start appealed to the generation of American psychologists who had already broken with the older mentalism and taken to behaviorism. In particular, S. S. Stevens was a leading spokesman for the psychological arm

of the movement throughout the 1930s (cf. Stevens, 1935a, 1935b, and 1939), the same period during which he was struggling to find a new theory of measurement.

Operationalism implies a theory of measurement. This was made explicit by the British operationalist, H. Dingle, who defined measurement as "... any precisely specified operation that yields a number" (1950, p. 11). Obviously, according to operationalism, what such a number measures is "synonymous with" the "precisely specified operation." The issue of whether the number relates to a quantitative variable is, on this view, a nonissue. One has a procedure and the procedure generates a number, so one must have "measurement." Stevens' starting place was Campbell's theory and not the operational theory of measurement. However, he developed those aspects of Campbell's theory that led in Bridgman's direction. He finished up with a theory that, whether he realized it or not, was indistinguishable from Dingle's operational theory of measurement. He did this by developing Campbell's *representationalism* at the expense of his emphasis on additivity.

The concept of additivity is an essential part of the classical theory of measurement, for without additivity one cannot have ratios of magnitudes and without ratios of magnitudes there is no measurement. Campbell, however, logically detached the use of numbers in measurement from the existence of additivity. He still insisted on additivity, although not because this was thought to be logically necessary for the use of numbers. He replaced additivity as the basis for measurement by the concept of representation: numbers are used in measurement to *represent*. Reading this shift of emphasis quite accurately, Stevens recognized that other relations apart from additivity could be represented numerically and given this new emphasis such representations had an equal right to be included within the concept of measurement.

Numbers could be used to represent identity or difference with respect to some attribute. For example, if people are classified by sex and all males assigned 1 and all females 0 then one has the numerical representation of a nonadditive empirical relation. The relation is that of *being the same sex as* and *A* is the same sex as *B* if and only if *A* and *B* are assigned the same number. This assignment rule gives what Stevens thought of as the simplest variety of measurement, the *nominal scale*.

Or, numbers could be used to represent order relations between things. Students' essays, for example, fall into the set of ordered categories: high distinction, distinction, credit, pass, fail. Numbers may be assigned to essays to represent this ordinal information. Then the empirical relation represented is that of *being at least the same quality as*, and essay *X* is at least the same quality as essay *Y*, if and only if the number assigned to *X* is at least as great as the number assigned to *Y*. This assignment rule gives what Stevens called an *ordinal scale*.

Thirdly, the numbers assigned may be used to represent equality and order of differences between things with respect to some attribute. Suppose, for example, that the volume of a certain liquid expands linearly with increases in temperature. Let *X* and *Y* be two samples of the liquid, equal in volume but at different

temperatures (temperatures *A* and *C* respectively). Suppose that they are heated to new temperatures, *B* and *D* respectively. Then a judgment can be made via the change in volume as to what temperature difference is the greater, *B* - *A* or *D* - *C*. In this case the relation is that of one temperature difference being at least as great as another temperature difference and numbers may be assigned so that temperature *B* - temperature *A* is at least as great as temperature *D* - temperature *C* if and only if $n_B - n_A \geq n_D - n_C$ (where n_A, n_B, n_C and n_D are the numbers assigned to the temperatures *A, B, C* and *D* respectively). Such an assignment rule produces what Stevens called an *interval scale*. Finally, Stevens recognized, that empirical, additive relations could also be represented numerically, producing what he called a *ratio scale*.

These are Stevens (1946, 1951, 1959) four general types of measurement scales and although he sometimes mentioned others (e.g., the log-interval scale), he seems to have believed that these four were the most important. The completed theory distinguished them one from another by classes of admissible scale transformations and prescribed for each scale-type a class of permissible statistics. This latter aspect of his theory always remained controversial within psychology, though it would seem to be a trivial point that numbers used to represent different kinds of empirical information could not all be treated in the same way. In particular, rules would have to be devised to ensure that only conclusions already entailed by the empirical information alone, could be drawn from the numbers assigned. Stevens' "measurement" scales are really instances of numerical coding, the purpose of the coding being to allow inferences to be made by numerical means.

Of course, these four types of scales are not the only ones possible. Every empirical relation, no matter what its character, can be represented numerically. Again, this is a fairly trivial observation. As is well known, any relation can be defined set-theoretically as a class of ordered *n*-tuples. That is, each binary relation is a set of ordered pairs, each ternary relation a set of ordered triples, and so on. Now, let *R* be any binary empirical relation which we wish to represent numerically. Let each object in the field of *R* be assigned a different number, any number. If *x* and *y* are any such pair of objects then there exists a binary numerical relation, *S*, such that

$$<n_x, n_y> \in S \text{ if and only if } R(x, y)$$

(where n_x and n_y are the numbers assigned to *x* and *y* respectively, a set described by "<" and ">" is an ordered set and $R(x, y)$ means that *x* stands in the relation *R* to *y*). *S*, being a set of ordered pairs of numbers, is a numerical relation and it represents the empirical relation, *R*. A similar relation exists for any ternary relation or for any relation of any degree. So it follows that if measurement is nothing more nor less than numerical representation then *all* variables can be measured and any consistent assignment of numbers to objects measures *something*. Of course, such measures may not generally be very useful, but the important point is that they exist.

Whether Stevens grasped this point or not is unclear. However, from 1946 onwards he always defined measurement as, "... the assignment of numerals to objects or events according to rules" (1946, p. 667, 1951, p. 1, and 1959, p. 19). In 1959 he added the comment that "... provided a consistent rule is followed, some form of measurement is achieved" (p. 19). So it seems he may have. If so, then he took the representational theory as far as it can go, making it equivalent in practice to the operational theory. On both views, any "precisely specified" operation for making "consistent" numerical assignments to things is "measurement."

In being brought to this destination the concept of measurement is virtually emptied of all content. Not only are all variables measurable and all consistent numerical assignments measurement, but also the ancient distinction between quantity and quality is dissolved. Whether a variable is quantitative or not simply becomes a pragmatic issue, one determined not by the character of the variable itself, but by the decision of the researcher. This point was acknowledged recently in the context of multidimensional scaling by F. W. Young, who noted that "... the distinction does not reside in the observation circumstances, but is in the mind of the observer" (1984, p. 56). Thus to say of a variable that it is measurable by a certain procedure is not to say anything about it nor about the procedure (other than that it is number generating). It is simply to say that some one wants to treat the variable numerically.

Precisely because it was so emptied at the hands of Stevens, the concept of measurement came to encompass all of the quantitative and pseudo-quantitative techniques existing within psychology. And it was because it so obviously promised to do this that Stevens' definition of measurement was rapidly absorbed into psychology's storehouse of conventional wisdom. Not only that, it was also made to appear as if this definition was all that measurement had ever meant. By 1954 Coombs, Raiffa, and Thrall were referring to it as the "classical measurement theory." Even Stevens was surprised by that (cf. his 1959, p. 21). Also in 1954, J. P. Guilford, in the second edition of his famous text, *Psychometric Methods*, quoted it (p. 5). He remarked with some satisfaction that now the concept of measurement was wide enough to include psychological procedures. Then he endorsed Stevens' definition strangely attributing it, not to Stevens, but to Campbell. A similar kind of error was also recently made by the historian of psychology, H. E. Adler, who wrote that "Fechner understood the essential nature of measurement as 'the assignment of numerals to objects or events according to rule'" (1980, p. 14). As already indicated, Fechner did no such thing: he endorsed the classical theory of measurement! These mistakes suggest a tendency to bury all other views on measurement beneath Stevens' definition.

Guilford's textbook was not the only influential authority to endorse Stevens' definition. It was also endorsed by most of the major reference books of the time within psychology. Stevens' own *Handbook of Experimental Psychology*, of course, contained his (1951) treatise on measurement, which, for many psychologists over the next 30 years, "stood like the Decalogue" (Newman, 1974). B. F.

Green (1954), in his article on attitude measurement in Lindzey's *Handbook of Social Psychology*, also quoted it, attributing it not only to Stevens but, to add to the weight of authority, also to "Lorge (1951) and others." Whoever the "others" were, it is not difficult to guess from whence they obtained Stevens' definition. In the case of Lorge a guess is not necessary. His article had appeared in another widely used reference book, E. F. Lindquist's *Educational Measurement*, and he explicitly attributed the definition to Stevens.

Thus, there was within the psychology of the 1950s an almost immediate consensus about the meaning of measurement. Stevens' definition produced a collective sigh of relief. Since then, more than a generation of psychology students have been trained in the belief that this definition states what measurement really is. The misconception is now firmly entrenched. As a consequence, it is now accepted within psychology that mental tests measure abilities, that multidimensional scaling measures perceptual attributes, that magnitude estimation measures the intensity of sensations, and that summated rating scales measure attitudes and personality traits. In each case the justification for these beliefs resides in Stevens' definition. To illustrate this, consider Norman Cliff's justification of rating scales as a method of psychological measurement: "Crude as they are, rating scales constitute a workable measurement technology because there has been repeated observation that numbers assigned in this way display the appropriate kinds of consistency" (1982, p. 31).

Prior to the acceptance of Stevens' definition few psychologists would have been so bold on behalf of their quantitative methods and least of all on behalf of this "crude" method. As already indicated, Boring had concluded in 1934 that psychology lacked measurement and he was referring to psychophysics, which had until then been regarded as the jewel in psychology's quantitative crown. From then until the acceptance of Stevens' definition, treatments of psychological measurement were full of doubt. Stevens' definition allowed the illusion of quantification to persist in the face of Campbell's critique.

Stevens' definition is mistaken because it confuses two quite different practices: measurement (in the classical sense) and numerical coding. As will be made clear in chapter 3, measurement involves the discovery of empirical facts of an intrinsically numerical kind. Numerical coding on the other hand is simply a useful cosmetic in the analysis and presentation of what is otherwise nonnumerical information. To confuse these two is to confuse how things are, with the use of a symbolism for representing them. For measurement involves an ontological commitment to quantity and number as part of the furniture of the universe, whereas numerical coding is simply a notational device for the symbolic representation of facts. Any theory that confuses these two is a hindrance to scientific progress for it obscures important factual issues that need to be addressed.

This chapter has shown why psychologists were prepared to accept this confusion. They were obsessed with the implementation of a quantitative program within their science. Such an obsession would permit scientific progress only if the variables studied actually were quantitative. Thus, the logically prior task

was that of finding out whether or not these variables were quantitative. The mistake of the psychologists was to be more interested in the pursuit of their quantitative program than in the pursuit of the underlying facts. They steered this course because they believed that all science must be quantitative. Campbell's theory of measurement, inadequate as it was, challenged psychologists to examine the underlying factual issues. They declined to meet this challenge and took instead the apparently easier way offered by the acceptance of Stevens' definition. However, this definition gives nothing more than the illusion of measurement. It has not served psychology well, for the underlying factual issues still need to be addressed. What is needed to replace Stevens' definition is an understanding of how these issues can be tackled. Clearly, Campbell's approach via his concept of fundamental measurement is too narrow. A wider range of methods is needed in order to test whether or not any given variable is quantitative. Then those methods that can be applied to psychological variables may be used to settle the matter.

Section 2

And Philosophy

what has philosophy got to do with measuring anything?
—Galileo