

The Theory of Measurement in Psychology

Before considering in some depth the question of what quantity and measurement really are, I want to consider two conceptions of measurement now widely accepted within psychology: operationalism and representationalism. The preceding chapter traced the emergence of these two theories of measurement from the older, classical theory. It is important to recognize that the history of a theory is logically independent of the issue of its truth. Even granting that these later conceptions of measurement were accepted within psychology *simply* to save the preconception that psychology had to be quantitative, it still may be the case that one or the other of them is true. Other arguments must be presented in order to show them to be false. That is the aim of this chapter.

These two conceptions, operationalism and representationalism, will be treated separately, first, because they are distinct doctrines and, second, because they are motivated by quite different spirits. This is not to say, of course, that they cannot be welded together into a single view of measurement, as in Stevens' case. But even in this case, his theory of measurement was predominantly a representational theory. It was his definition and his practice of measurement (consider for example his method of magnitude estimation), which showed the influence of operationalism.

Operationalism, as a general methodological doctrine, is motivated by the fear of being mistaken. Its aim is to restrict scientific claims as far as possible, to a mere description of scientific observational practice. It is against speculation. Of course, such a spirit is not scientific. Anyone who fears being mistaken and for this reason seeks a "safe" or "certain" scientific method, should never enter upon any scientific enquiry. To do science is to speculate about how things work and to follow up one's speculations with observations.

Representationalism, as a theory of measurement, is motivated by a quite different concern. It is motivated by the belief that numbers are not part of the furniture of the universe. Its aim is to explain how, given this premise, numbers find their way into the practice of measurement. Whatever the philosophical motivations for this view about numbers (and the motives are many and varied), it is mistaken. There is no place in science for entities thought to be outside of space and time or for "convenient" fictions. When we speculate in science, our speculations are about real things located in the same spatio-temporal universe as ourselves. In chapter 3 an attempt will be made to explain just where within the universe the numbers found in measurement are located.

In the first section of this chapter it is argued that operationalism is logically incoherent. It involves the deliberate confusion of *what* is being measured with *how* it is being measured. It attempts to define what is being measured solely in terms of the relations into which it enters (*viz.* the operations involved in its measurement). This is logically impossible because something can only stand in a relation (to other things) if it has an existence independent of that relation. Operationalism denies an independent existence to that which is being measured.

In the second section it is argued that although the representational theory of measurement is logically coherent, it is, as a matter of fact, a false theory of measurement. It is false because it fails to allow for a realist interpretation of measurement statements. Such an interpretation of measurement statements is part and parcel of scientific practice and a theory of measurement that conflicts with that cannot be accepted.

OPERATIONALISM

Bridgman's reasons for putting forward his doctrine of operationalism were quite different to psychologists' reasons for accepting it. As already mentioned, his sprang from the, then recent, revolution in physics. He was disturbed by the fact that the Quantum and Relativity theories involved the revision of concepts (like space and time) fundamental to "Newtonian" theory. This fact conflicted with Bridgman's inductivism. On this view, scientific knowledge accumulates bit by bit, with only minor adjustments being made to laws and theories from time to time. The overthrow of "Newtonian" theory was no minor adjustment and so could have been seen as an indictment of inductivism, but Bridgman concluded, instead, that the fundamental concepts involved in "Newtonian" theory were to blame. They had been improperly defined in the first place, being laced, he thought, with more than a dash of metaphysics.

Bridgman's solution was a far-reaching program of conceptual reform. Scientific concepts were to be purged of their metaphysical content through being reduced completely to the kinds of operations scientists use. A scientist may, for example, say that he takes the word "length" to refer to a property of objects. In his laboratory practice, however, the use of the word is tied to the operations he

performs with certain measuring instruments. It is these operations, says Bridgman, that really show what he means by the word and not any verbal formula he might give. Bridgman goes on:

The concept of length involves as much as and nothing more than the set of operations by which length is determined. 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, Bridgman's italics)

An early reviewer of Bridgman's book, L.J. Russell (1928), noted how ill this view accords with the actual usage of scientific terms. Often one operation is spoken of as being better than others for the measurement of some quantitative variable. No such comparison is allowed by Bridgman, for if "the concept is synonymous with the corresponding set of operations" then each different operation defines a different concept and such comparisons are no longer possible. Quantitative concepts as they are used in science generally transcend the operations used in their measurement. Although length may be measured in different ways, each different procedure is generally held to be measuring the same quantity, namely length. Bridgman's operationalism shatters the conceptual unity of properties such as length, leaving a multiplicity of splinter concepts, one for each different procedure. This only proves, however, that Bridgman's principle is not a description of how things are done in science. He never intended it to be: Bridgman came to prescribe, not to describe.

Many psychologists accepted his principle, using it as a rod with which to beat mentalism, which they saw as a kind of metaphysics. This ignored Bridgman's view that mental operations could be used to define mental concepts. Their operations were invariably of a physical public kind, such as depriving an animal of food for 10 hours or measuring the time it took to run a maze. So they took operationalism to imply that all theoretical concepts used in psychology must be defined in terms of such operations (cf. Bergmann & Spence, 1941). Not only the behaviorists, but also many psychometricians, found this construal of Bridgman's principle agreeable. They were attempting to measure variables like intelligence, attitudes, and the intensity of sensations, variables that had been thought of as mentalistic. The issue of whether or not the procedures they used actually measured such variables lay unresolved in the background, not often addressed, but nagging. Operationalism appeared to solve their problems in one fell swoop. The concept measured could be defined in terms of the procedures used. As Boring was to put it for the concept of intelligence: "Intelligence is what tests test" (1945, p. 244).

It was left to the British operationalist, the physicist H. Dingle, to draw out of Bridgman's operationalism the theory of measurement implicit within it. He defined measurement as "any precisely specified operation that yields a number" (1950, p. 11). Not all concepts so defined will be of equal scientific interest, he notes:

I think the reason why we make some measurements and not others can be understood only in terms of the ultimate object of science, which is to find relations between the elements of our experience. If we passively accept experience as it comes the problem is far too difficult, so we establish artificial conditions that yield results more easily related with one another than natural occurrences are. (pp. 11-12)

Dingle's view captures the operationalist spirit that has pervaded attempts at psychological measurement from the 1930s to the present. On any other view it is a real issue whether or not any number-generating operation is a genuine measurement procedure. Only operationalism makes this a matter of definition. So, if operationalism is refuted, this theory of measurement is refuted as well.

Bridgman's principle, if taken seriously, makes all concepts, and all operations, unknowable. If the meaning of a concept is the set of operations by which it is determined, then knowing the meaning of the concept is dependent on knowing the appropriate set of operations. "Unless one knows the operations one does not know the meaning" (Bridgman, 1938, p. 116). However, recognizing, or knowing operations must in turn depend on knowing concepts. So which comes first, knowing concepts or knowing operations? For example, all cases of length measurement by ruler will have something in common in virtue of which they are just operations of that kind rather than some other. This something must be some set of properties or relations involved in the operation. Identification of the operation requires recognizing them. In this case the relevant properties or relations may be the property of being a ruler, the relation of being placed along side of, and counting, which involves relating the names of the natural numbers (numerals) to things. The operation can only be recognized as being of the right kind if these concepts are recognized. However, according to Bridgman's principle, recognizing these concepts must involve recognizing the operations by which they, in turn, are identified. So, to know a concept is to recognize an operation and to recognize an operation is to know further concepts, which requires knowing further operations and so on *ad infinitum*. In practice this infinite regress does not occur because there are some concepts that are known directly, unmediated by any prior knowledge of operations. However, the regress follows from Bridgman's principle and so proves it unworkable.

Operationalism is not saved by the reply that observation itself is an operation and, hence, that knowledge of concepts by direct observation is knowledge based upon operations (cf. Stevens, 1939, p. 233). It is true that observation is an operation, in the sense that it is something that we do. Stevens calls it science's "fundamental operation." This attempted reply overlooks the distinction between *performing* an operation and *knowing* it. For example, in our seeing that X is red, what we know is that X is red and not that we are seeing that X is red. That is, the seeing operation itself is usually not known while we are performing it. So, clearly, the concept (red) may be known even though the operation by which it is determined (seeing) ordinarily goes unrecognized.

Even if this was not true, operationalism would still fail. For if in seeing that

X is red, the concept of red is reducible to the operation of seeing, then the concept of red would be indistinguishable from that of blue, or any other visible quality. They, likewise, would be completely defined by exactly the same operation. In so linking concepts to the operation of seeing them, the operationalist arrives at a position not unlike Berkeley's subjective idealism. Bridgman was happy with this and in a later publication (1950) made his subjectivism explicit. To psychologists, however, subjectivism was anathema, and few followed him to that destination.

Operationalism served quite a different purpose within psychology. It was not applied in psychology with the same rigour as was used by Bridgman in applying it to physics. It was never applied in psychology to those concepts that are able to be directly perceived and so the regress from concepts to operations to concepts and so on *ad finitum* was avoided. Operationalism was only applied to, so-called, *theoretical concepts* and was seen as a means of purifying them of "surplus meaning." The aim of this more dilute operationalism was to reduce all theoretical concepts to observational concepts by defining them in terms of observationally specifiable operations. Theoretical terms not so reducible were seen as a conceptual extravagance, covering the operational base of science with an unwanted layer of metaphysics.

This is a false view of the scientific enterprise. In as much as the aim of science is to discover nature's ways of working, dilute operationalism is more of a hindrance than a help. Many of nature's ways of working are at first obscure and hidden from our view. An approach to science that does not let us think beyond that which can be currently observed prevents us from understanding these hidden ways. Commenting on this variety of operationalism, W.M. O'Neil says:

It urges the scientist, in effect, not to take risks incurred in moving far from the facts. However, it may properly be asked whether science can be undertaken without taking the risks of skating on the possibly thin ice of supposition. The important thing is to know when one is on the more solid ground of observation and when one is on the ice. (1969, p. 154)

More seriously still, this variety of operationalism incurs a logical error. The events involved in an operation (or its outcome) are at most only causally related to those hidden workings that theoretical concepts are intended to elucidate. To define theoretical concepts in terms of such events is to violate the principle that cause and effect must be logically distinct occurrences (a principle attributed to Hume, 1888). For example, if the strength of hunger drive is defined in terms of hours of food deprivation then it cannot also be partly caused by it, or if spatial ability is defined by performance on some tests then it cannot be invoked to explain such performance. So, quite contrary to dilute operationalism, theoretical concepts *must not* be defined operationally.

Both Bridgman's classical operationalism and the psychologists' dilute operationalism involve a deliberate confusion between concepts and the means (operations) by which they are identified. This confusion is at its most general in

Bridgman's Berkeley-like belief that "it is meaningless to separate 'nature' from 'knowledge of nature' " (1927, p. 60). In the measurement context it takes the form of a confusion of "the act or process of measuring with the object of the act, namely the quantity in question" (Byerley, 1974, p. 376). The error here is simply that of confusing a relation with one of the entities involved in it. Knowing something, identifying it, or measuring it is a relation between the observer and the thing known, measured or identified. The measurement operation, on which Bridgman puts so much emphasis, is simply the specific way the relation of measuring occurs. The logical form of the situation is: person *A* measures (using operation *X*) variable *Y*. *Y* could only be defined in terms of *X* if it was constituted by it. However, relations cannot constitute the entities that they hold between, otherwise there would be no entities to stand in the relation. That is, if person *A* measures variable *Y* then *Y* must exist independently of measurement, otherwise there would be nothing to measure. The confusion of an entity with a relation that it enters, if carried far enough, denies the possibility of nature existing independently of the observer and makes science impossible. Operationalism is the thin end of subjectivism.

The fact that operationalism is false means that the operationalist theory of measurement must be rejected. Useful as it has been to psychologists over the last 50 years in sustaining the image of psychology as a quantitative science, its philosophical confusions mean that that image may be just an illusion. If operationalism is denied then the question of whether or not a given operation really measures a particular variable, or any variable at all, must be a genuine issue. So psychology, which possesses a wide range of number-generating procedures, may really possess no measurement procedures at all. This issue can only be judged relative to a sounder theory of measurement than operationalism.

THE REPRESENTATIONAL THEORY OF MEASUREMENT

This theory is currently the orthodox theory of measurement within the philosophy of science. It is said (Adams 1979, Krantz, Luce, Suppes, & Tversky, 1971) to have had its origins in the work of the German mathematician, Holder (1901). In English, it found early articulation in Russell's *Principles of Mathematics*, when he wrote:

Measurement of magnitudes is, in its most general sense, any method by which a unique and reciprocal correspondence is established between all or some of the magnitudes of a kind and all or some of the numbers, integral, rational or real, as the case may be. (Russell, 1903 p. 176)

Campbell's (1920) was its most influential formulation and there it is applied to physical measurement, at least as it stood early this century. It is his version or versions based on it that have since been presented in many introductory texts in

the philosophy of science (e.g., Carnap, 1966, Cohen & Nagel, 1934, and Pap, 1962).

Stevens' liberalized version of Campbell's theory received a set-theoretical presentation in Suppes and Zinnes (1963). The set-theoretic approach became axiomatic in Krantz et al. (1971). This version (which I call "Suppes' theory") (cf. R.D. Luce, 1979) is the version presented in this chapter. The set-theoretical symbolism is largely dispensed with because it obscures matters for the non-mathematical student. A recent and relatively nontechnical resume of this theory is presented by Narens and Luce (1986). A much more complete, although highly technical presentation is Narens (1985).

The Numerical Representation of Empirical Systems

According to the representational theory of measurement the role of numbers in measurement is to represent. Numbers are assigned to empirical entities in such a way that certain relations between the numbers assigned represent empirical relations between the entities. Representationalists make a distinction between empirical entities and numerical ones. Empirical entities are "identifiable entities, such as weights, persons, attitude statements, or sounds" (Suppes & Zinnes, 1963, p. 7). What is meant by "identifiable" here is not obvious, for if numbers are not also identifiable in some sense it is unclear how they could be used at all. Luce (1979) prefers to define empirical entities as "concrete objects or events" (p. 94), a construal that Adams (1979) further narrows down to "concrete observable things" (p. 208). Here, the adjective "concrete" is obviously intended to distinguish empirical entities from *abstract entities*. In recent American philosophical parlance, abstract entities are those not located in space or time. Although rarely made explicit by these authors, the distinction is only complete if numbers are then taken to be abstract entities. This view of numbers is the conventional wisdom of our time, however, if numbers are abstract in this special sense, then their apparent involvement in "concrete" empirical contexts such as measurement requires explanation. So although it is true that the representational theory does not logically require an "abstract" view of number, it does, as a matter of fact, include it. Indeed, as already indicated, it is precisely because this view of number is so widespread that the representational theory of measurement is so popular. Bring numbers back to earth (as is done in chapter 3) and the representational theory is no longer needed.

One of Suppes' central concepts is that of a *relational system*. A relational system is a class of entities together with one or more relations into which the entities enter. For example, suppose that the entities involved are rods of various lengths and that the relation concerned is that of *being at least as long as*. Then the rods standing in this relation to one another constitute a relational system.

Relational systems possess structure. The structure is constituted by the properties of the relations involved. For example, the relation of being at least as long as has the properties of being *transitive* and *strongly connected*. What these terms

mean can be indicated more clearly if some symbols are used. Let . . . , "x", "y", "z", . . . stand for entities in the class under consideration. Let "R" stand for the relation involved and let expressions such as " $R(x,y)$ " mean that x stands in the relation R to y . Then, a relation R is transitive relative to a class if and only if, for all x , y and z in that class, if $R(x,y)$ and $R(y,z)$ then $R(x,z)$. A relation is strongly connected relative to a certain class if and only if, for all x and y in that class, either $R(x,y)$ or $R(y,x)$. (See Appendix 1 for other properties of binary relations.)

The properties that relations have cannot be taken for granted. They must be discovered by observation and experiment. To continue with the example of the rods, any two rods can be placed side by side in parallel so that the left-hand ends terminate together (as in Fig. 2.1). Then it can be observed that rod x is at least as long as rod y , if y 's right-hand end does not terminate to the right of x 's. Then taking the rods pair by pair the properties of this relation can be discovered. That is, it can be seen that the relation is transitive and strongly connected.

According to the representational theory, measurement depends on finding a numerical relational system that exhibits a structure similar to that of the empirical relational system under consideration. A numerical relational system is one in which the entities involved are numbers and the relations are numerical relations. An example is the set of all positive integers less than 100, and the relation of being at least as great as (conventionally symbolized as " $\geq$ "). Two relational systems are similar in structure when corresponding to each relation in the first system there is a relation in the second system that possesses the same properties. For example, $\geq$ is also transitive and strongly connected. If one relational system has a structure similar to a second relational system, then the first system may be

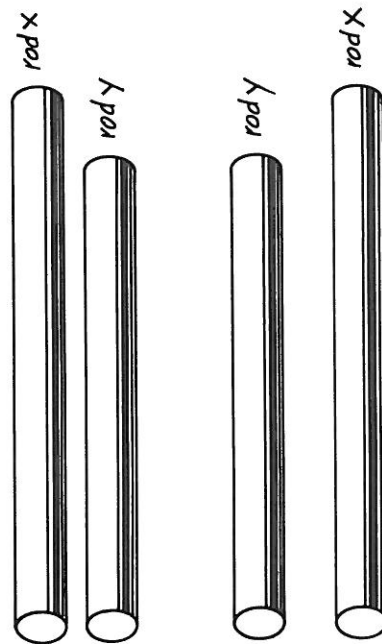


FIG. 2.1 If any two rods, x and y , are side by side in parallel so that their left-hand ends terminate at the same perpendicular line then x is at least as long as y if and only if the right-hand end of y does not terminate at a point to the right of x 's.

used to represent the second, providing, of course, that the first system possesses at least as many entities as the second. Thus the aforementioned numerical relational system may be used to represent the empirical relational system mentioned in the rods example.

The numerical representation would be achieved by matching numbers to objects so that the order of length of the objects is reflected in the order of magnitude of the numbers. That is, each rod would be assigned a single number. Let $n(x)$ be the number assigned to any rod, x . Then the numerical representation is achieved when the numbers are assigned according to the following rule: rod x is at least as long as rod y if and only if $n(x) \geq n(y)$. The rods are said to be mapped into the numbers.

A mapping of objects into numbers is one to one when each object is assigned a single number, that number being different from the number assigned to any other object. A mapping of objects into numbers is many to one when the second of these two conditions is relaxed and the same number may be assigned to more than one object. The representation of one relational system by another is called an isomorphism when the mapping is one to one, and it is called a homomorphism when the mapping is many to one. It is usually the case in measurement that more than one object is assigned the same number. Hence, measurement generally involves a homomorphism between an empirical relational system and a numerical relational system. Such a homomorphism (or, in the cases where it occurs, isomorphism) Suppes and Zinnes (1963) referred to as a scale of measurement.

In the rods example, the set of rods on which the observations are made must be a finite set. It is to this finite set that the numerical assignments are made according to the above rule. However, this finite empirical relational system is embedded within a much larger, indeed possibly infinite, one. This larger system consists of the set of all things having length and the relation of being at least as long as. Obviously, observations cannot be made on each pair of objects in this system. Despite that limitation, the available evidence confirms the hypothesis that this relation is also transitive and strongly connected on that set. So a homomorphism between this larger system and the positive real numbers ordered by $\geq$ may be hypothesized to exist. Thus the finite data structures observed are taken to be representative of the underlying infinite structures of which they are subsystems, in measurement as in all science. The scope of any established measurement scale is then taken to extend to all entities possessing the relevant attribute.

The above rods example is an instance of what Stevens (1946) called an ordinal scale. It is a very simple example of measurement according to the representational theory. A more complex example is measurement on what he called a ratio scale. Length may be measured on a ratio scale but to keep the examples distinct the measurement of weight is chosen to illustrate measurement of this kind.

The example will be kept simple by considering a set of objects (of various weights) small enough to fit into the balance pan of a beam balance like that

illustrated in Fig. 2.2. This time two relations will be considered: that of one object being at least as heavy as another; and that of two objects taken together (concatenated) being at least as heavy as a third. For convenience the first relation will be symbolized as " $x \geq y$ " and the second as " $x \circ y \geq z$ " (following Krantz et al., 1971). The presence of these relations between the objects is detected using the beam balance as follows. $x \geq y$ is taken to be true if, when just x is placed in one pan and just y in the other, the pan containing y fails to descend. $x \circ y \geq z$ is taken to be true if, when just both x and y are placed in one pan and just z in the other, the pan containing, z fails to descend. If neither pan descends in the first case it means that both $x \geq y$ and $y \geq x$ and this is abbreviated as $x \sim y$. The same result in the second case means that both $x \circ y \geq z$ and $z \geq x \circ y$ are true and this is abbreviated as $x \circ y \sim z$. Also, the fact that $x \geq y$ and not $(y \geq x)$ is expressed as $x > y$. Similarly, $x \circ y \geq z$ and not $(z \geq x \circ y)$ as $x \circ y > z$. If $x \sim y$ then y is a replica of x (for weight, that is) and n concatenated replicas of x (where n is any positive whole number) are referred to as " nx ." Finally, $x \circ (y \circ z)$ is the object obtained when x is concatenated with the concatenation of y and z .

Suppose that the behavior of the set of objects is observed on the beam balance and that the following facts are discovered about these two relations.

1. The relation $\geq$ is found to be transitive and strongly connected.
2. For all x, y and z in the set, $x \circ (y \circ z) \sim (x \circ y) \circ z$.
3. For all x, y and z in the set, $x \geq y$ if and only if $x \circ z \geq y \circ z$ if and only if $z \circ x \geq z \circ y$.
4. If for any x and y in the set $x > y$ then for any w and z in the set there exists a positive whole number n such that $nx \circ w \geq ny \circ z$.
5. For all x and y in the set, $x \circ y > x$.

Proposition 4 (which is sometimes referred to as an Archimedean condition) may seem obscure. It simply means that no matter how much heavier z is than w the weight difference between x and y multiplied by some n can compensate for it.

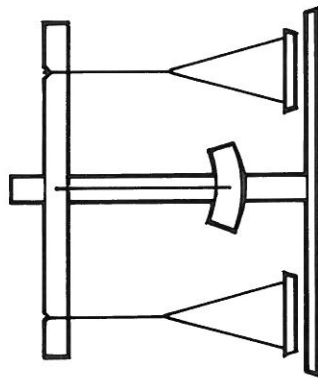


FIG. 2.2 A schematic diagram of a beam balance.

That is, no weight difference is infinitely large relative to another weight difference. This condition might then seem to be partly numerical and thus violate the earlier stipulation that empirical relational systems be nonnumerical. The representationalist has an easy reply to this apparent objection. nx , it is asserted, is an empirical and not a numerical object. It is a collection of n replicas of x . The set of all such collections itself forms an empirical relational system homomorphic to the positive whole numbers, and so numerical assignments can be made to such collections. The number one is assigned to collections like x by itself, two to collections like $x \circ x$, three to collections like $x \circ (x \circ x)$, and so on. Hence, although, numerical expressions like " nx " are used, it is claimed that the objects or facts referred to remain nonnumerical. This interpretation is debatable and an alternative point of view is suggested in the following chapter.

Returning to the five propositions whose truth has been established by observations on the beam balance, Krantz et al. (1971, chapter 3) show that an empirical relational system of just that structure is homomorphic to a numerical relational system containing the relations $n_1 \geq n_2$ and $n_1 + n_2 \geq n_3$ (where, n_1, n_2 , and n_3 are, of course, numbers). That is, numerical assignments may be made to the empirical objects such that for all x, y , and z in the set:

- (1) $x \geq y$ if and only if $n(x) \geq n(y)$, and
- (2) $n(x \circ y) = n(x) + n(y)$.

So in this case the empirical operation of concatenation is represented by the numerical operation of addition.

It is interesting to note at this point the empirical concatenation operations of this kind may also be represented by the numerical operation of multiplication. An example mentioned by Suppes and Zinnes (1963) illustrates this. If to each entity x the number $n_1(x) = e^{n(x)}$ (where e is the base of the natural logarithms and $n(x)$ is an assignment made in accordance with the requirements 1, and 2 above) is assigned, then it is true that:

- (3) $x \geq y$ if and only if $n_1(x) \geq n_1(y)$, and
- (4) $n_1(x \circ y) = n_1(x) \times n_1(y)$
 $= e^{n(x)} \times e^{n(y)}$
 $= e^{n(x) + n(y)}$.

So following (3) and (4) would lead to a multiplicative representation of the same empirical relational system as is additively represented via (1) and (2). Ellis (1966) gives other examples of how an empirical concatenation operation of this kind can be numerically represented in other ways.

The representational theory of measurement must then explain why additive representations are preferred in measurement to any other possible kinds of numerical representations. Ellis' answer is that they are the simplest of possible

representations. In the case of concatenation operations like the one just shown, this may be true. Alternative kinds of representations are also possible with other kinds of empirical relations. For example, because $\leq$ is also transitive and connected it could also be used to represent the empirical relation of being at least as long as mentioned in the rods example. Then longer rods would be assigned smaller numbers. Such a representation seems just as simple as the one conventionally chosen.

Empirical relational systems such as the one described in the weights example are called *extensive* or *additive* systems. N. R. Campbell's view was that extensive systems were the only ones capable of measurement. Russell's (1903) view differed. He considered measurement to be the numerical representation of *magnitude*, by which he effectively meant an empirical relational system involving an ordering relation (such as greater than or less than in some sense). This is also the opinion of both Ellis (1966) and Luce (1979). Stevens (1946) advocated an even wider concept. His famous definition includes *any* instance of the numerical representation of an empirical relational system as measurement, no matter what its structure is. In practice he drew the line with what he called *nominal scales*. These are the numerical representation of empirical systems of a purely classificatory kind. For example, if people are classified by gender with males assigned 1 and females 0, the result is a nominal scale of measurement in Stevens' scheme of things.

As we have seen, Stevens' liberalism enabled psychologists to regard their procedures as measurement even though Campbell's view denied them that status. Where then is the *correct* place to draw the line between measurement and nonmeasurement within the representational theory? Once the shift is made from the classical theory to the representational theory there can be no justification for denying any instance of numerical representation the status of measurement. To do so is to introduce an arbitrary limitation into the application of what is otherwise a clear principle: measurement is the numerical representation of empirical objects and relations. Stevens' thoroughgoing representationalism is the most defensible version.

Of course, not all representations will prove equally useful. But this observation simply reinforces the fact that representationalism is a version of instrumentalism: numbers are a tool of science and it is up to the scientist to use them to their greatest advantage. It is an ironic twist that the way psychologists use them is as a means of presenting their science as if it were a quantitative science, where "quantitative" is intended in the classical sense, not the representational.

Do all instances of measurement recognized by science conform to the representational pattern? Physics deals with over 100 measurable variables. All of them are functionally interrelated, and all of them are ratio scales in Stevens' sense. Although many like weight, length, angle, time, and electrical charge, are based on empirical relational systems involving a concatenation operation able to be represented by numerical addition, many obviously are not. Density is an example. Two objects of different densities, however they are combined, typically

do not give an object whose density is the sum of their densities taken separately. How then can the representational theory explain the measurement of density?

In practice density is measured via its functional relationship to other variables, mass and volume, using the fact that, density = mass/volume. So the measurement of density is based on the prior measurement of mass and volume, and this kind of dependence characterizes what Campbell called *derived measurement*. He thought that derived measurement scales emerged through the discovery of constants in numerical relationships between extensive (or *fundamentally measurable*) variables.

Although this account may be historically plausible, it will not do as an explanation of the logic of derived measurement from the representational point of view. What needs to be exposed is the fact that the numbers assigned via derived measurement actually *represent* something empirical. For, if the representational theory is true then equations like the one above for density are, in the words of Krantz et al. (1971, p. 504), "only a shorthand summary of what is really a complex set of possible qualitative observations." By "qualitative" they mean nonnumerical.

So in cases of derived measurement one is actually dealing with an empirical relational system involving a number of quantitative variables. In the case of density this relational system involves not only density but also mass and volume. The "qualitative" relations between objects within this system must be such that they may be represented numerically by assigning to each object, x , three numbers, $d(x)$, $m(x)$, and $v(x)$ such that:

$$(1) d(x) = m(x)/v(x), \text{ and}$$

$$(2) m(x) \text{ and } v(x) \text{ are proportional to measures of mass and volume obtained fundamentally.}$$

The character of such an empirical relational system is too complex to delve into in this chapter. Suffice it to note at this stage that such a system would be observed if objects were classified according to density (which they would be if classified according to the kind of material they are composed of), ordered according to volume, and "qualitative" observations made on the associated masses. Something of the flavor of such a system is given in discussions of the theory of conjoint measurement (see chapter 4 below and Krantz et al., 1971). So the representational theory is able to give an account of the class of measurements Campbell calls derived.

The Theory of Scale Types

There is always an arbitrariness in the numbers chosen to do the job of measurement, according to the representational theory. Having decided what numerical relations to use in representing the empirical relations, different sets of numbers may be assigned to the empirical entities to achieve the desired homomorphism.

Thus, different scales for the same variable may be obtained. Such scales will not differ in terms of how any given empirical relation is represented numerically, however they will differ in the numerical assignments involved. This point may be clarified by considering an example.

Consider again the rod example treated earlier in this chapter. There an ordinal scale was constructed on the basis of the relation of being at least as long as. Suppose that a , b , c , and d are four rods and the numerical assignments made in scale 1 (see Table 2.1) reflect their length order. Because the numbers assigned under scale 2 are in exactly the same order, they also reflect the same empirical order relationships. Thus, for these four objects, scales 1 and 2 represent the empirical relational system involved equally well. Of course, there are infinitely many other sets of numerical assignments that would also do. On the other hand, scale 3 does not represent length order by numerical magnitude, so it will not do.

Stevens posed the problem of how all ordinal scales for the same variable are related to each other. Suppes called this the *uniqueness problem* for ordinal scales. It is the problem of how unique the numerical assignments are in the case of an ordinal scale. Or, in other words, how may the numerical assignments on any ordinal scale be changed without changing the convention that empirical order is represented by numerical magnitude?

The answer, obviously, is that they can be changed to anything at all providing their order of magnitude is left intact. That is, if $n_i(x)$ is the number assigned to x on scale i and $n_j(x)$ is the number assigned to x on scale j , then the transformation from scale i to scale j leaves the homomorphism intact if and only if, for all pairs of objects x and y ,

$$n_i(x) \geq n_i(y) \text{ if and only if } n_j(x) \geq n_j(y).$$

Such a scale transformation is called an *increasing monotonic* transformation. Stevens calls increasing monotonic transformations the class of admissible scale transformations for ordinal scales of measurement. Ordinal scales are said to be unique up to increasing monotonic transformations.

In contrast to ordinal scales, consider the example of weight measurement given earlier in the section on "The Numerical Representation of Empirical

TABLE 2.1
Hypothetical Ordinal Scale (Scales 1 and 2) and Nominal Scale
(Scale 3) Measures of the Lengths of Four Rods, a , b , c , and d

Rods	Numerical Assignments		
	Scale 1	Scale 2	Scale 3
a	4	10	4
b	3	9	10
c	2	3	8
d	1	2	9

Systems." Let a , b , c , and d be four objects whose weights are measured and suppose that $a \circ a \sim b$, $a \circ b \sim c$, and $a \circ c \sim d$. Table 2.2 lists four sets of numerical assignments that might be made to these four objects. Scales 1 and 2 represent these empirical relations in the manner described in the aforementioned section. That is, $n(a) + n(a) = n(b)$, $n(a) + n(b) = n(c)$, and $n(c) + n(c) = n(d)$. In scale 3 the concatenation relationships are not represented by numerical addition, but the weight order is represented by the order of magnitude of the numbers. In scale 4 not even this is the case, although each different object has been assigned a different number.

Because scale 3 is related to both scales 1 and 2 by an increasing monotonic transformation, and because scale 3 does not represent all of the available empirical information in the manner of a ratio scale, it follows that in ratio scale measurement the class of admissible scale transformations must be narrower than it is in the case of ordinal scales. In fact, as is obvious, scale 1 is simply scale 2 multiplied by 4. This illustrates the fact that if either scale is multiplied by a positive constant, then the transformed scale preserves the homomorphism. A ratio scale is defined by the fact that $x \circ y \sim z$ if and only if $n(x) + n(y) = n(z)$ and if each number is multiplied by a positive real number k then this gives $k \cdot n(x) + k \cdot n(y) = k \cdot n(z)$, which is true whenever $n(x) + n(y) = n(z)$ is true and conversely. Such a transformation is called a *positive similarities* transformation and ratio scales are said to be unique up to such transformations.

Stevens distinguished what he thought were two other important scale types. The *nominal scale* is based on classification and assigns the same number to objects that are equivalent (in the required sense), and different number to objects that are not. For example, scale 4 in Table 2.2 is a nominal scale for weight. The class of admissible transformations for nominal scales is the class of *one to one* transformations. That is, if $n_i(x)$ is the number assigned to x within nominal scale i and $n_j(x)$ is the number assigned within nominal scale j then the transformation from i to j is one to one if and only if $n_i(x) = n_i(y)$ if and only if $n_j(x) = n_j(y)$, for all pairs of objects x and y .

The *interval scale* is one in which the class of admissible scale transformations

TABLE 2.2
Hypothetical Ratio Scale (Scales 1 and 2), Ordinal Scale (Scale 3),
and Nominal Scale (Scale 4) Measures of the Weights
of Four Objects, a , b , c , and d

Objects	Numerical Assignments			
	Scale 1	Scale 2	Scale 3	Scale 4
a	4	1	1	9
b	8	2	4	7
c	12	3	5	8
d	16	4	10	4

is the class of *positive linear* transformations. That is, two interval scales, i and j , for the measurement of the same variable will be related by the following kind of equation: $n_i(x) = a \cdot n_j(x) + b$ (where a is some positive number and b is either a positive number or a negative one). For example, in temperature measurement the Fahrenheit ($^{\circ}F$) and the Celsius ($^{\circ}C$) scales are related by the following linear transformation: $C = (5/9) \cdot F - 160/9$.

An interval scale involves the numerical representation of relations upon differences between entities with respect to some attribute. It requires an ordering upon such differences and a relation of composition upon them. That is, a relation equivalent to the following kind: difference x is equal to difference y plus difference z . In the case of temperature this relation is empirically determined through the hypothesis that temperature differences are directly related to volumes (such as the volume of mercury in a thermometer). Because volume is measured on a ratio scale, this makes temperature differences measurable on a ratio scale. This case illustrates very well the fact that an interval scale is really a ratio scale upon differences.

A staunch critic of Stevens' theory of scale types is W. W. Rozeboom (1966), who has described it as "complete nonsense" (p. 188). His criticism is quite subtle. In chapter 1 the point was made that a relation may be defined as a set of ordered n -tuples. Such a definition is called *extensional*. That is, it is alleged that relations may be defined by reference to the entities standing in the relation rather than by reference to the character of the relation itself. A binary relation then becomes a set of ordered pairs of entities; a ternary relation, a set of ordered triples; and so on. Allowing this point, consider any homomorphic assignment of numbers to a set of empirical entities. The numbers so assigned, simply because they are numbers, stand in a great many relationships to one another. Rozeboom's criticism of Stevens' theory is based on the premise that each such numerical relation represents an empirical relation, even though it may only be an extensionally definable one. For example, suppose that a set of people are classified according to eye-color: persons a , b , and c all have blue eyes; e , f , and g all have brown eyes; and both i and j have green eyes. This classification is an empirical relational system, the relation involved being that of having the same colored eyes as (call it R_1). Let this empirical relational system be numerically represented as follows: assign the number 2 to a , b , and c ; the number 3 to e , f , and g ; and the number 5 to i and j . This creates a nominal scale in which the relation of being equal to represents R_1 . Now $2 + 3 = 5$ is a numerical relation holding between the numbers used in this nominal scale. Does this numerical relation represent any empirical relation holding between the eight people a , b , c , e , f , g , i , and j ? Rozeboom thinks that it must. It represents the following extensionally defined relation: $R_2 = \{ \langle a, e, i \rangle, \langle a, e, j \rangle, \langle a, f, i \rangle, \langle a, f, j \rangle, \langle a, g, i \rangle, \langle b, e, i \rangle, \langle b, e, j \rangle, \langle b, f, i \rangle, \langle b, f, j \rangle, \langle b, g, i \rangle, \langle b, g, j \rangle, \langle c, e, i \rangle, \langle c, e, j \rangle, \langle c, f, i \rangle, \langle c, f, j \rangle, \langle c, g, i \rangle, \langle c, g, j \rangle \}$. It represents this relation because R_2 holds between any ordered set of three people, x , y , and z (from the original set of eight), if and only if $n(x) + n(y) = n(z)$ (where $n(x)$, $n(y)$, and $n(z)$ are the numbers assigned to x , y , and z

respectively). So the empirical relation R_2 has been given an additive representation. Of course, this additive representation is only preserved by a positive similarities transformation of 2, 3, and 5. Hence, what began as a nominal scale has been shown to be a ratio scale as well. The distinction between Stevens' scale types have collapsed.

The argument here is correct up until the last step, providing the initial premise that relations are extensionally definable is granted. Any numerical assignment to empirical objects, be it initially a nominal or ordinal scale, will additively represent some extensionally defined empirical relation and so will also be a ratio scale (providing that the numbers assigned are additively related). But this argument does not entail the collapse of Stevens' theory.

One could save Stevens' theory by denying the principle that relations are extensionally definable or by arguing that relations like R_2 above are not of any scientific significance. Such moves, valid though they may be, are not necessary. Stevens' theory is saved by the fact that empirical relational systems are partly constituted by the relations involved in them, and even though two systems may involve precisely the same objects they are different if they contain different relations. So the eight people together with R_1 is *one* empirical relational system, and the same people together with R_2 constitutes a *different* empirical relational system. The first system, when represented by 2, 3, and 5 together with numerical equivalence ($=$), produces a nominal scale. If the numbers assigned are transformed by anything other than a one-to-one numerical transformation then = no longer represents R_1 . The second empirical relational system, when represented by 2, 3, and 5 together with the relation of numerical addition (N_+), produces a ratio scale. If the numbers assigned are transformed by anything other than a positive similarities transformation then R_2 is no longer represented by N_+ . So the two different empirical relational systems produce two different scales. What they have in common are the empirical objects measured and the numerical values assigned, however the scales involved are no more the same than they would be if the hardness values assigned to a set of minerals on Mohs' scale of hardness fortuitously equalled their weight in grams.

What follows from Rozeboom's argument is not the collapse of Stevens' theory but the following quite different conclusion. Given a scale of any type on a set of empirical objects there also exists a scale of any other type involving exactly the same numerical assignments. In particular, it follows that for each nominal scale on a set of empirical objects, there exists a ratio scale involving exactly the same numerical assignments as the nominal scale, and representing some extensionally defined empirical relation not included in the nominal scale's empirical relational system.

This conclusion raises an interesting question. Other things being equal, ratio scales are always preferred to nominal scales, involving as they do the representation of more empirical information. Why then would anyone ever prefer a nominal scale to the ratio scale it carries along with it? There is no universal valid reason. If a researcher establishes a nominal scale representation and then

extensionally defines the empirical relation additively represented by the numbers employed, he may find this relation to be of enormous scientific interest. He would then be quite justified in preferring this ratio scale to his previously established nominal scale. The matter is purely one of scientific interest. However, let the reader be warned. This is not a recommended method for finding scientifically interesting empirical relationships. The empirical relation extensionally identified by such a process is identified only as an artifact of the numbers used in the nominal scale. To expect such a relation to be of any scientific interest displays a touching faith in serendipity. The *a priori* probability that such a fortunate event will occur must be close to zero.

Rozeboom's criticism is based on a true observation but it fails to take account of the instrumentalist character of the definition of the concept of a measurement scale used within the representational theory of measurement. Any homomorphic assignment of numbers to empirical objects will always represent a large class of, perhaps only extensionally definable, empirical relations. Within this large class the scientist focuses on the small set of empirical relations of interest to him. It is how these relations and these alone are numerically represented that characterizes a scale according to the representational theory of measurement. That other empirical relations are also represented is ignored, and rightly so, given that the numbers are only to be used as an instrument. Stevens' theory of scale types describes how the numbers used may be changed without changing which numerical relations represent the empirical relations of interest. It specifies the degrees of freedom available in the choice of numbers relative to the empirical relations of interest, and the numerical relations chosen to represent them. The fact that, no matter what numbers are chosen within these degrees of freedom, they always represent other (perhaps, only extensionally definable) empirical relations must not be denied, but it is of no relevance to the theory of scale types. However, it is of relevance to another problem encountered by the representational theory of measurement and that is the so-called problem of permissible statistics.

THE PROBLEM OF PERMISSIBLE STATISTICS

Stevens' classification of scale types was no trivial matter. What differentiates scale types is the character of the relations within the empirical and numerical relational systems involved in a measurement scale. He recognized that the purpose of measurement was the inference of empirical conclusions from the numerical assignments made. He further recognized that the character of valid inferences from measurements must be relative to scale type, simply because such inferences depend on the character of the empirical information represented. According to the representational theory a measurement scale is an instrument of deduction. As such it contributes no content to any conclusions obtained: it simply lubricates the inference process.

Suppose that the eight objects listed in Table 2.3 are measured for weight on a ratio scale (e.g., the metric grams scale) and that the measurements are as listed. The objects belong to either set 1 or set 2. The mean of the weight measurements for objects in set 1 is 8. The mean for objects in set 2 is 10. Since 10 is greater than 8 it immediately follows that the mean weight in grams of objects in set 2 is greater than the mean weight in grams of objects in set 1. This conclusion is a proposition about the weight measurements and hence about the numbers. It is tempting to move from this purely *numerical* conclusion to the nonnumerical, *empirical* conclusion that the mean weight of objects in set 2 is greater than the mean weight of objects in set 1. This conclusion is a proposition about the weights (rather than the weight measurements) and so about some of the objects within the empirical relational system. Does this second conclusion follow validly from the first? It does, but this kind of inference is not valid with just any kind of measurement scale whatsoever.

Suppose that the same objects are measured on an ordinal scale of hardness. Such a scale is based upon a single empirical relation. This is the relation of being harder than which is assessed by rubbing two objects together to see which one scratches which. If x scratches y then $n(x) > n(y)$. The hardness measures for these objects on a hypothetical scale A are shown in Table 2.4. The mean of the scale A measures for set 1 is 5.5 and the mean for the objects in set 2, is 4.5. Since 5.5 is greater than 4.5 it immediately follows that the mean of the scale A measures for objects in set 1 is greater than the mean of the scale A measures for objects in set 2. Does it follow validly from this merely numerical conclusion that the mean hardness of objects in set 1 is greater than the mean hardness of objects in set 2? It cannot!

Consider hardness measures of these eight objects on a different hypothetical ordinal scale, scale B (see Table 2.5). Scales A and B are related by an increasing monotonic transformation (i.e., the numbers are in the same order of magnitude). Hence, these two scales do an equally satisfactory job of representing the empirical *harder than* relation. A choice between them would be quite arbitrary.

TABLE 2.3.
Hypothetical Ratio Scale Measure of Weight of the Eight Objects,
a to *h*

Objects	Scale Values	
Set 1	<i>a</i>	2
	<i>b</i>	6
	<i>c</i>	10
	<i>d</i>	14
Set 2	<i>e</i>	4
	<i>f</i>	8
	<i>g</i>	12
	<i>h</i>	16

TABLE 2.4.
Hypothetical Ordinal Scale (Scale A) Measures of the Hardness
of the Eight Objects, *a* to *h*

Objects	Scale A Values
Set 1	
<i>a</i>	1
<i>b</i>	6
<i>c</i>	7
<i>d</i>	8
Set 2	
<i>e</i>	2
<i>f</i>	3
<i>g</i>	4
<i>h</i>	9

However, the mean of the scale *B* hardness measures for objects in set 2 (7.25) is greater than the mean of the scale *B* hardness measures for objects in set 1 (7). So if it is valid to infer from the scale *A* measures that the mean hardness of objects in set 1 is greater than the mean hardness of objects in set 2, then it is also valid to infer the contradictory conclusion from scale *B* measures. This is because in both cases the form of the argument is the same.

In both cases the empirical information forming the premises of the argument (i.e., the set of hardness relationships between the eight objects) is the same and itself noncontradictory. Although different numbers are used to represent this information, the same representation rule is used (i.e., *x* is harder than *y* if and only if $n(x) > n(y)$) and this rule is internally consistent. The arithmetical argument from the measures themselves to the relationship between the means of the measures is of the same valid form in both cases. Therefore, up to this point in the two arguments it is not possible that mutually contradictory conclusions could be entailed. Only when the empirical conclusions about the mean hardnesses are inferred does the contradiction arise. Hence, that last step must be invalid, for the

same set of noncontradictory premises cannot entail contradictory conclusions. So this form of argument is not valid given ordinal scale measurements.

Why then is this form valid given ratio scale measurements? Simply because the argument is to a conclusion that is already entailed by the empirical premises. To see this, the empirical analogue of the otherwise numerical notion of a mean must be specified. For the case of weight this is as follows. If *n* objects (not necessarily all of the same weight) perfectly balance (i.e., weigh the same as) another *n* objects (this time all objects of the same weight), then each of the objects in the second group equals the mean weight of the first group of objects. Therefore, it follows by inspection of Table 2.3 that the four objects in set 2 will outweigh the four objects in set 1 and, will have a greater mean weight. This conclusion follows validly from the empirical information alone, the arithmetic used earlier being merely a deductive convenience.

Even though it is obvious from his writings that Stevens recognized the above facts he failed to set the argument out in this fashion. Furthermore, he misrepresented the issue by laying down prescriptions about what statistics ought not to be computed given measurement scales of the various types. For example, of ordinal scales he said, "In the strictest propriety the ordinary statistics involving means and standard deviations ought not to be used with these scales, for these statistics imply a knowledge of something more than the relative rank-order of the data" (Stevens, 1946; p. 679). However, the matter is not one of what ought or ought not to be done. It is simply a matter of fact: of what kinds of conclusions validly follow from what kinds of premises. Stevens further obscured the issue by attempting to justify his prescriptions on the basis of the invariance of the statistics computed under admissible scale transformations. Although this line is not without merit, Adams, Fagot, and Robinson (1965) show that the sense of invariance intended by Stevens is not clear.

Many psychologists, confident that their freedom to compute was as inalienable as their freedom to speak, took up arms in the cause of computational libertarianism (see for example Burke, 1953; Lord, 1953) and so began the "permissible statistics" controversy. From the start it was misdirected. All computations are permissible, it is just that some are not helpful in the derivation of valid empirical conclusions from measurement data.

That the issue was not one of *permissible* statistics was recognized by Suppes (1959). However, viewing the matter through neopositivist lenses, he also misperceived it, supposing it to be one of the empirical meaningfulness of measurement statements. This view has largely become the accepted wisdom (see for example, Adams, Fagot, and Robinson, 1965; Luce, 1979; and Roberts, 1976, 1979), though Suppes (1979) has recently expressed second thoughts.

As a solution to the problem, Suppes and Zinnes proposed the following criterion of empirical meaningfulness: "A numerical statement is *meaningful* if and only if its truth (or falsity) is constant under admissible scale transformations of any of its numerical assignments, that is, any of its numerical functions expressing the results of measurements" (1963, p. 66).

TABLE 2.5.
Hypothetical Ordinal Scale (Scale B) Measures of the Hardness
of the Eight Objects, *a* to *h*

Objects	Scale B Values
Set 1	
<i>a</i>	1
<i>b</i>	8
<i>c</i>	9
<i>d</i>	10
Set 2	
<i>e</i>	5
<i>f</i>	6
<i>g</i>	7
<i>h</i>	11

This criterion means that statements like the following are not empirically meaningful.

1. The mean of the scale *A* hardness measures for objects in set 1 is greater than the mean of the scale *A* hardness measures for objects in set 2.

If an admissible scale transformation (i.e., an increasing monotonic transformation) is applied to the numerical assignments on which it is based, and the scale *B* values result, then statement 1 goes from being true to being false. Hence, its truth is not constant under admissible scale transformations of its numerical assignments, so it must be classed as empirically meaningless.

On any ordinary understanding of the word "meaningful", statement 1 is meaningful. If it were not it could be neither true nor false. However, the adjective "empirical" is intended to specify a narrower sense of meaningful than that used in ordinary discourse, and Adams, Fagot, and Robinson (1965) have attempted to explain this narrower sense (actually, they preferred the term "scientific significance" to that of "empirical meaningfulness"). A paraphrase of their view is that: A numerical statement is not empirically meaningful if it contains expressions denoting numerical relations, having no empirical analogue within the empirical relational system involved in the scale under discussion. Thus, statement 1 is not empirically meaningful, because the empirical relational system involved in any ordinal scale does not contain an analogue of the numerical relation of *being the mean of*. In contrast to this, the empirical relational systems involved in ratio scales do. So, to put the matter as simply as possible, Adams, Fagot, and Robinson are suggesting that statements like 1 above are not empirically meaningful because they have no empirical interpretation in the relevant empirical relational system.

Is this a good enough account of what could be meant by being empirically meaningless in this context? The answer is no, and for two reasons. In the first place, it concedes far too much to verificationist theories of meaning. Just because a variable (such as hardness) is so far only known to be ordinal, it does not follow that it is empirically meaningless to hypothesize that its structure is really additive (like that of weight or length). Given such a hypothesis, the notion of mean hardness is empirically meaningful, (though speculative and, therefore, evil in the eyes of operationalists and positivists) even though it is not definable in terms of any hardness relations yet known. Of course, nothing about such hypothesized mean hardness follows from statement 1, but that is not relevant to the issue of empirical meaningfulness.

Second, Rozeboom's (1966) observation that any numerical relation between numerical assignments must represent some (perhaps extensionally defined) empirical relation implies that the mean of a set of ordinal hardness measures *does* correspond to *some* empirical relation between the objects measured. Even though this relation will not be involved in the empirical relational system underlying the ordinal scale, it does constitute an empirical interpretation of the mean. Thus,

no well-defined numerical relation upon measurements is ever devoid of an empirical interpretation and so must always be empirically meaningful.

However, even if Adams', Fagot's, and Robinson's sense of empirical meaningfulness had been acceptable it turns out not to be consistent with Suppes' and Zinnes' criterion. If the objects in sets 1 and 2 are rearranged so that all objects in set 1 are harder than all objects in set 2 then the truth of statement 1 remains constant under all admissible scale transformations. Thus according to Suppes' and Zinnes' criterion, it would be empirically meaningful under such circumstances, even though the sole empirical relation involved in the relevant empirical relational system remains unchanged, and incapable of defining an empirical analogue of the mean. So Suppes and Zinnes may well have meant something quite different by "empirically meaningful" from the suggestion made by Adams, Fagot, and Robinson. Whatever Suppes and Zinnes may have meant, it must have been very subtle: simply by rearranging the memberships of sets 1 and 2, statement 1 is transformed from being "empirically meaningless" to being "empirically meaningful."

The problem has never been anything other than simply one of the valid inference (Michell, 1986). Because the role of numbers in measurement, according to the representational theory, is simply to represent empirical facts and through representing them facilitate the deduction of empirical conclusions, the numbers themselves cannot contribute any content to the conclusions arrived at. Campbell was in no doubt that the whole point of applying numbers to reality was, "to enable the powerful weapon of mathematical analysis to be applied to the subject matter of science" (1920, pp. 267-268).

And as Field observed, "the conclusions we arrive at by these means are not genuinely new, they are already derivable in a more long-winded fashion from the premises without recourse to mathematical entities" (1980, pp. 10-11). This, in a nutshell, is the solution to Stevens' problem of permissible statistics within the framework of the representational theory of measurement. Measurement can only validly lead to those empirical conclusions already entailed by the empirical premises. These premises will be propositions about the empirical entities and the empirical relations that they enter into. The empirical relations concerned are those involved in the empirical relational system on which the scale is based. The relations within the numerical relational system will have been chosen because they possess the same formal properties as these empirical relations. Thus for each valid empirical argument there will be a correspondingly valid numerical argument involving just those numerical relations and no others. Thus, if all numerical inferences from measurements are restricted to valid argument forms involving just those numerical relations representing the relevant empirical relations and to no others, then all inferences to the appropriate empirical interpretations of the numerical conclusions obtained will be valid, and no valid inferences of that kind will be excluded. The use of argument forms involving any other numerical relations will then be unnecessary to those whose interest is validity, for the use of such numerical relations is liable to lead to invalidity.

Applying this solution to the above example, it can be seen that the calculation of means of ordinal scale measurements is generally not helpful in scientific research. There is nothing to stop one from doing it and any conclusions arrived at, will be just as empirically meaningful as any other conclusions one arrives at in scientific investigation. It is just that considering only the empirical data on which the ordinal scale is based, no empirical conclusions about means validly follow. To compute the mean is to go beyond the data given, and to infer empirical conclusions about it, is to infer what cannot validly follow from that data. Rozeboom's point implies that there always will be a set of empirical propositions, (about extensionally defined relations) that do entail empirical interpretations of whatever conclusions about means follow from the numbers assigned. However, these propositions are not the premises that the scientist employing the ordinal scale is using. Of course, he always has the option of using them, but if he does he is no longer working with an ordinal scale and it is then incumbent on him to explicate the empirical relational system with which he is working.

So Stevens' prohibition against calculating means of ordinal data does have something to recommend it: ordinal data alone entails nothing about means, so following his prescription excludes no valid consequences of one's data. Likewise, his prescriptions on the permissibility of other statistics relative to types of measurement scales have similar support: they do not exclude the inferences of any valid conclusions from one's data.

A CRITIQUE OF REPRESENTATIONALISM

The literature on measurement theory contains few critiques of the representational theory. This is because in one guise or another it is almost universally accepted amongst measurement theorists. Since 1966, E. W. Adams has been almost its lone critic, and in two papers (1966 & 1979) has expressed his doubts about it. I am not convinced that any of his criticisms is fatal to the theory and, ironically, the greatest problem for the theory is a feature that Adams apparently regards as no disadvantage at all. This is the problem of the meaning of measurement statements.

While discussing the alleged distinction between theoretical and observational statements Adams makes the following remark about the representational theory (which he calls F.M., i.e., fundamental measurement, theory): "And here F.M. Theory, crude as it may be at present, has an important insight to contribute which is too frequently overlooked. This is that numerical measurements are *not* direct observational 'data' and that challenges to the empirical character of science which presupposes this are simply mistaken" (1979, p. 220).

Adams is correct: the view that measurements are not direct observational data does follow from the representational theory of measurement and it is a view that is frequently overlooked. It is frequently overlooked because many scientists and philosophers adopt the opposite view, namely, that at least some statements

expressing the results of measurements are directly observational. Statements such as,

1. The length of x is 2.5 meters, or
2. The weight of y is 3.12 kilograms.

These are commonly taken to be reports of direct observations and free of any nonobservational concepts. That is, the metrical predicates, 2.5 meters and 3.12 kilograms, are like many other metrical predicates, interpreted empirically and realistically. They are taken to refer to particular values of the length and weight variables, values which may be directly observed. For example, 2.5 meters is taken to be a length which is two and a half times that of the standard meter rod in France. This realistic interpretation of metrical predicates depends on the realistic interpretation of numerical terms like "two and a half." However, that is how numerical terms are ordinarily understood. It is this kind of realistic interpretation of metrical predicates that is disallowed by the representational theory, and it is why this theory implies that measurement statements are not observational.

According to the representational theory, statements like 1 and 2 do not describe the empirical objects that they refer to (i.e., objects x and y), neither do they describe relationships between such objects. According to this theory, the meanings of 1 and 2 may be paraphrased as follows:

3. The number assigned to x on the meter length scale is 2.5.
4. The number assigned to y on the kilograms weight scale is 3.12.

These statements merely relate the empirical entities x and y to numbers. Because numbers are held to be external to the empirical situation under study (i.e., x and y and their length and weight relations to other empirical objects) statements 3 and 4 cannot be describing that situation.

This difference between the representational theory and the ordinary realistic interpretation of measurement statements extends to the inferences made from such statements. Ordinarily, statements 1 and 5 would be taken to imply statement 6.

5. The length of y is 5 meters.
6. The length of y is twice the length of x .

However, according to the representational theory, 6 does not follow from 1 and 5. Statement 5 must be paraphrased as:

7. The number assigned to y on the meters length scale is 5.

Obviously, 3 and 7 do not entail 6 and so, according to the representational theory, neither do 1 and 5. What 3 and 7 do entail is:

8. The number assigned to y on the meters length scale is twice the number assigned to x on the meters length scale.

Statement 8 only leads to 6 given certain facts about the meters length scale or, equivalently, given certain facts about the empirical relational system on which it is based. The representationalist must argue that the ordinary habit of inferring statements like 6 from 1 and 5 presumes these facts. This would be news to most people with a practical interest in measurement, including both scientists and tradesmen.

The difference, however, bites more deeply. According to the representational theory, statements like 6 are semantically ill-formed because they mix numerical and empirical modes of discourse. One thing's being twice another is a relationship of a numerical kind. The lengths of x and y are nonnumerical properties. Because representationalism is premised upon the assumption that numbers (and their relations) are external to the empirical relational system involved in measurement, statement 6 cannot be taken literally. On the ordinary realistic view numerical relations are taken to hold between nonnumerical things and the meaning of 6 is obvious. On the representational view statement 6 must be understood metaphorically as meaning something like,

9. The lengthwise concatenation of x with an object of the same length produces an object of the same length as y .

The matter of importance here is not the difference in meaning (or lack of it) between statements 6 and 9 on a realistic interpretation, but the fact that statements like 6 have no literal meaning on the representational view. For statements like 6 (i.e., those asserting numerical relations between empirical entities) are the bread and butter of quantitative science.

The problem is exacerbated by the fact that such "mixed-mode" statements achieve considerable complexity in much of scientific discourse. While it requires little effort to see through a statement like 6 to something like 9 behind it, more complex mixed-mode statements are less transparent. What, according to representationalism, is the literal meaning of such "metaphorical" statements as 10?

10. The product moment correlation coefficient between height and weight amongst school children in Sydney is .65.

A product moment correlation is a numerical relationship and its coefficient is a real number. Yet what is held to stand in this numerical relationship is height and weight, both of which are empirical (nonnumerical) properties. It may be tempting

for a representationalist to suggest that 10 is really a statement about numbers (i.e., the measurements of height and weight), but it is not. Statement 10 may be based on measurements, but it asserts the existence of a relationship between the attributes measured. That some relationship exists between the children's height and weight in the empirical domain, given that 10 is true, cannot be denied. And it will be in virtue of this relationship's existence that 10 is true. Yet, it is a relationship that is remarkably difficult to describe without recourse to numerical terms such as "mean," "variance," "covariance," and "ratio," also it is one that scientists never refer to in anything other than numerical terms.

However, the solution given to Stevens' problem of permissible statistics, required that any valid inferences from measurements to empirical facts must be implications already entailed by the empirical data itself. It now appears that even in cases of moderate complexity we know of no way to make the inference other than via the use of numbers, because our nonnumerical empirical language is too poor. As far as most scientific results and generalizations of a quantitative kind are concerned, the language used to describe them does not penetrate the representation to the reality represented. The representationalist is forced to believe in the existence of relationships that he is in practice unable to describe other than indirectly or metaphorically, through the medium of the language of mathematics. Because this is held to be the language of "abstract" (i.e., nonspatio-temporal) entities, the curious picture that the representationalist paints of the quantitative scientist is of one made to use the language of the nonempirical in order to discourse on the empirical.

The representationalist is led to this puzzling destination through his conviction that numbers are external to the empirical facts involved in measurement. How has such a conviction gained currency in the face of scientific practice, which treats numbers otherwise? The answer is that those who have developed this theory stand within a philosophical tradition that takes it for granted that numbers are not empirical kinds of things. From such a point of view the application of mathematics to reality is a deep problem in need of explanation, and the representational theory was devised to serve that purpose. On the other hand, if numbers are accepted in the same way as quantitative scientists implicitly treat them (i.e., as empirical) then the fact of their use is hardly a problem at all. Then the representational theory becomes a liability, for it prizes apart the numerical concepts used in science from the empirical reality studied. There is nothing logically incoherent about the representational theory, however if the empirical realist theory of measurement described in the next chapter is true, then representationalism is empirically false.