Pergamon

Stud. Hist. Phil. Sci., Vol. 28, No. 1, pp. 167–185, 1997

0039-3681/97 $17.00+0.00

A Hundred Years of Numbers. An Historical Introduction to Measurement Theory 1887–1990

Part I: The Formation Period. Two Lines of Research: Axiomatics and Real Morphisms, Scales and Invariance

*José A. Díez**

The aim of this paper is to reconstruct the historical evolution of the so-called Measurement Theory (MT). MT has two clearly different periods, the *formation period* and the *mature theory*, whose borderline coincides with the publication in 1951 of Suppes' foundational work, 'A set of independent axioms for extensive quantities'. In this paper two previous research traditions on the foundations of measurement, developed during the formation period, come together in the appropriate way. These traditions correspond, on the one hand, to Helmholtz's, Campbell's and Hölder's studies on axiomatics and real morphisms and, on the other, to the work undertaken by Stevens and his school on scale types and transformations. These two lines of research are complementary in the sense that neither of them is enough taken alone, but together they contain all that is necessary to develop the theory, and it is in Suppes (1951) that these complementary approaches converge and all the elements of the theory are appropriately integrated for the first time. With Suppes' work, then, begins what may be called the 'mature' theory, which was to develop rapidly later on, especially during the 1960s. Our historical reconstruction is divided into two parts, each part devoted to one of the periods mentioned. Part I also contains a conceptual introduction which aims to establish the use of some notions, specifically those of *measurement and metrization.* Although the reconstruction is not exhaustive, it intends to be quite complete and up to date compared to what is available in measurement literature; in this sense the aim of this paper is mainly historical but, although secondarily, it also attempts to make some conceptual and metascientific clarifications on the subject of the theory. Copyright © 1997 Elsevier Science Ltd.

*Department of Social Anthropology and Philosophy, Rovira i Virgili University, Plaza Imperial Tàrraco 1, 43005, Tarragona, Spain.
Received 26 October 1995; in revised form 12 February 1996.

S0039-3681(96)00014-3

Introduction: Measurement, Metrization and MT

I have (let us suppose) a diamond in front of me. It is small, sparkling, light, hard, beautiful and expensive. If I am asked to be more precise, I could say that it is very small, quite light, very, very hard and extremely expensive. I could carry on, but however much I refine my adjectives, it seems that I shall always be able to do so a little more (in almost all cases). It is noteworthy that as soon as I give the measures of the diamond for the properties involved I shall no longer be required to be more precise. Nevertheless, as we know, this is not possible for all its properties. I can say that its volume is x, its mass y, and even that its hardness is z, but not (up to now) that its beauty is w. Why not? I also have a piece of chalk in front of me. It is small, dull, light, soft, ugly and cheap. Both pieces are small and light, although the chalk not so much. Now I can also be more precise and even give (when possible) its measures and I may also be interested in comparing them with diamond's measures. So I can say that the mass of the chalk is 100 times that of the diamond while its hardness is only one-tenth. But it is also noteworthy that, while the former means something, the latter does not. Or, being more precise, that both things mean something but only what is meant by the former depends exclusively on facts related to the objects. Both express a numerical fact (the quotient of masses is 100, the quotient of hardness is 0.1) but only what is expressed by the former is exclusively dependent on the objects, not on our conventions. Why?

What we shall call 'fundamental metrization', a theoretical activity in the broad sense of the term, attempts to respond to these questions. This theoretical activity investigates the facts and conditions which make it possible to measure a property and the extent to which we can use the measures obtained to make objective statements about objects. In order to clarify the concept of *metrization* and the object of our historical reconstruction, i.e. the theory which results from this activity, we must carefully distinguish it from that of *measurement*.

Measurement is the assignment of numbers to objects in order to represent their properties, not any property but only those specific properties called *magnitudes* or *quantities*,[1] which are capable of 'more or less' instantiation, i.e. of instantiation in degree. Measurement can be derived or fundamental. In derived measurement, by far the most common in scientific practice, we obtain the desired value of a magnitude for an object with the help of other values we already have which are linked to the first one in a certain way; these known values that we use can be of the same magnitude for other objects or of other magnitudes for the same object. For example, we can measure the mass of a

[1]Terminological note. The words 'magnitude' and 'quantity' are ambiguous. Sometimes they are used to refer to the property we are measuring in the object, at other times to refer to the specific 'quantity' of the property the object has, i.e. the value of the measurement. There is no standard usage in measurement literature to be followed. I use them as synonymous, but when it is necessary to refer to both meanings, I tend to use 'magnitude' for the first, i.e. the property itself, and 'quantity' for the second.

heavenly body using the mass of a rocket, its change in trajectory when it travels near the body and certain mechanical laws which relate all these values. But measurement cannot always be derived, because derived measurement makes use of already known, i.e. *measured*, values and we must begin somewhere. This place is provided by fundamental measurement, not as common in practice but absolutely essential because it is 'where everything begins'. In fundamental, or direct, measurement we obtain the desired values *with no previous values at all*, directly from qualitative empirical data.

Measurement procedures, derived or fundamental, are possible because in the system where we assign a value to an object certain *facts* are obtained, because the system satisfies certain empirical conditions, certain 'laws of nature'. These empirical facts or laws are the possibility conditions of measurement *practice* and they are the object of theoretical investigation. We shall call 'metrization'[2] this theoretical activity which studies the mensurability conditions, i.e. the possibility conditions for measurement.

In derived measurement these conditions are quantitative laws of nature. But because these laws are (some) of the laws studied by common empirical quantitative theories, the theory of derived mensurability, properly speaking, has no autonomous subject; its task is already done by (one part or another of some of) those common quantitative theories. This is the reason why there is no autonomous theory devoted to the foundations of derived measurement.[3]

In fundamental measurement the systems and the conditions or laws are qualitative (remember that numerical values 'begin' only *after* it). But these qualitative conditions are not the subject of any *particular* empirical qualitative theory devoted to the study of (the qualitative aspects of) a *specific* magnitude, for the same set of qualitative mensurability conditions may correspond to qualitative systems of a very different physical nature, i.e. may provide the same foundations for the fundamental measurement of very different quantities. The possibility conditions of fundamental measurement are the subject of a specific (and somehow special) empirical qualitative theory. This theory, which provides the foundations for fundamental measurement, studies the (different

[2]Terminological note. When the term 'metrization' (and 'to metrize') is used in measurement literature, and it is used very rarely, it usually means the *introduction* or *constitution* of a new quantitative, metric concept (cf. Hempel, 1952, §12, Berka, 1983, chap. 6, §3; cf. also Stegmüller, 1970, chap. 1, §4). In the case of *fundamental* metrization, this consists in the specification of a procedure which enables a qualitative order to be represented numerically. This task has two parts. The first one is the theoretical investigation on the conditions that a qualitative empirical system must satisfy for the representation to be possible. The second one is to determine a concrete empirical procedure for qualitative comparison of the specific property involved, and to choose the standard with which the assignment begins. These tasks are essentially different. The way we use '(fundamental) metrization' only corresponds to the first, since the second is part of what we call '(fundamental) measuring procedures'. It is essential to distinguish both things, once we do so, the words we use for each one are not so important.

[3]An autonomous theoretical activity different from a mere 'theory of definitions', which provides the theoretical foundations for a very special kind of derived measurement, that in which the quantity we measure is introduced by *definition* in terms of others. Cf. Suppes and Zinnes (1963).

groups of) qualitative laws that make an empirical system, *of whatever physical nature*, mensurable, i.e. numerically representable. This theory is the theory of fundamental mensurability, or fundamental metrization, and is usually called 'Theory of Fundamental Measurement' or, in short (and because of the absence of a specific theory of derived mensurability), 'Measurement Theory' (MT).[4]

To sum up, Measurement Theory, which is the result not of the practical activity of measuring, but of a theoretical investigation that we call (fundamental) metrization, establishes the conditions that a certain domain of objects exhibiting a magnitude-property must satisfy for it to be possible to assign, without the help of other previous assignments, numbers to the objects in such a way that certain (common) mathematical facts concerning the numbers assigned appropriately represent certain qualitative empirical facts that occur because the objects exhibit the property. Note that if, in this characterization, we make reference only to 'the analysis of the conditions which make it possible to assign numbers to objects which display a property', there would obviously be nothing to analyze, since numbers can be assigned to any domain of objects under any conditions. So, though vague, the addendum 'in such a way that...' is essential.

Not every assignment can be regarded as measurement, and MT should make this additional restriction precise. Objects are involved in facts, some of which *will* be due to the property which is to be measured. The numerical assignment must *represent* these facts expressing them numerically. This is still not enough as a restriction, for the assignment can always be made with the only condition that there are at least as many numbers as objects: I have a particular fact made up of objects, I assign one number to each object and then I arbitrarily *define* mathematical properties and relations between numbers which replicate the ones for the objects. In this way I obtain 'numerical facts' which 'express' the state of affairs between objects. It is obvious that this is not what we want, for we have merely 're-named' the objects. The representative numerical facts must be 'mathematically common' and, in some way, 'natural'; otherwise the so-called measurement has no sense. This restriction, though vague, works, for not every state of affairs among objects can be represented in this way. In order to be represented in such a way there are conditions that systems must satisfy and metrization then becomes an interesting task. Nevertheless, these conditions may be very weak and give rise to 'not very useful' representations. This now raises the other side of the question. Once we have an

[4]Terminological note. Although, to be coherent in the use of the words, i.e. to be coherent with our use of 'metrization' for the theoretical activity which gives rise to the theory, I would prefer 'mensurability' or 'metrization' instead of 'measurement' for labelling the theory, when I do not use the abbreviation 'MT' I shall use the standard 'Measurement Theory' in order to agree with the common practice in measurement literature (the main references for MT are Narens (1985) and the *summa Foundations of Measurement*, Vol. I Krantz *et al.* (1971), Vol. II Suppes *et al.* (1989) and Vol. III Luce *et al.* (1990).

appropriate numerical representation, certain mathematical statements are true about the numbers assigned. But not all of these statements can be regarded as being *meaningful* for our objects, in the sense that not every numerical fact represents a state of affairs dependent only on the objects and the property. And this is essential because, the more meaningful numerical statements there are, the more useful the measurement will be. And this is in fact the other side of the coin since, the stricter the representational possibility conditions are, the greater the meaningfulness is. So, it is certainly possible to find appropriate representations with very weak conditions, but then meaningfulness is very limited and, consequently, so is the utility of the measurement.

Of course these comments would need further clarification, but in the present context they will suffice as a conceptual introduction to our historical task. Let us begin with the beginning of the story.

Helmholtz on alikeness and additivity

Helmoltz's essay 'Zählen und Messen erkenntnistheoretisch betrachtet',[5] published in 1887, is generally regarded as being the first theoretical contribution to questions related to measurement. And indeed, as far as we are concerned, it is in this essay that the question of the conditions which make measurement possible is explicitly formulated for the first time.[6]

Helmholtz calls 'magnitude' the 'attributes of objects which when compared with similar ones allow the distinction greater, alike or smaller'.[7] If we express attributes with numbers, these will be the *values* of the magnitude, and the procedure by which we find the values is the *measurement* of the magnitude. He then goes on to state the question which concerns us here: 'we shall have to investigate in which circumstances we can express magnitudes through (...) numbers'.[8] In my opinion, it is this question which is the starting point of what was later to be known as (Fundamental) Measurement Theory.

Helmholtz states that this investigation should begin with the concept of *alikeness*.[9] Alikeness, '(the) special relationship which may exist between the attributes of two objects', is characterized by two properties, those we know

[5]The references made will be from the English version if it is not otherwise stated (cf. Helmholtz (1921); also Helmholtz (1930)).

[6]The attention we are going to pay to Helmholtz's essay would be a little bit excessive if we only take into account the technical results compared with other authors, but we think it is justified because of the wealth of conceptual questions he raises.

[7]*Op. cit.*, p. 89. English translators render the German original 'vergleichen' and 'Vergleich' by '(to) liken' and 'likening'. I prefer, and I shall use, the more common translations '(to) compare' and 'comparison'.

[8]*Op. cit.*, p. 89. I omit here a qualification that Helmholtz makes (and which is maintained in the English translation) of the numbers in this context. In the German original, these numbers are known as 'benannte Zahlen' translated as 'denominate numbers' although the translator points out that it would be more natural to use 'concrete numbers', even though this does not fully fit Helmholtz's characterization of them (cf. p. 84 of the German original).

[9]'Gleichheit' in the German original.

today as symmetry and transitivity (pp. 88–89). The alikeness between two objects of comparable attributes is reached by observing certain factual results of the interaction of the objects in appropriate conditions and the procedure by which the objects are put into the appropriate conditions so as to be able to observe the result is known as the *comparison method.* So, the magnitudes that the two objects display are alike if: (1) the result observed on applying the comparison method to the two objects does not change when the order of the objects is inverted; and (2) both objects always give the same result when they are compared with the same third object. The concept of *magnitude alikeness* here seems to be a theoretical concept used to explain certain properties observed in the results of a comparison method.[10] Viewed in this way it is obvious that two alike magnitudes (properly speaking, two objects with alike magnitudes) are interchangeable as far as the comparison procedure is concerned, i.e. one can be substituted for the other without modifying the results of the comparison procedure, for it is precisely substitutivity which allows us to determine alikeness (p. 90). More interesting is the fact that they are also interchangeable in other respects, in other phenomena. These other phenomena which are preserved when substituting alike magnitudes are regarded, then, as being effects of the attribute in question: '(we) characterize the further effects in which alikeness is preserved as effects of that attribute, or as empirically dependent upon that attribute alone' (p. 91).

With the concept of alikeness and non-alikeness to hand, the concept of magnitudes of the same type, or homogeneous magnitudes, can be explicated: 'Magnitudes whose alikeness or non-alikeness is to be decided by the same comparison method are termed by us "alike in kind"[11] ' (p. 91). Understood in this way, magnitude kind is what today is known simply as magnitude, the attribute itself which is capable of being measured (e.g. mass). Helmholtz goes on to give some examples of such attributes (weight, length, duration and others) and of well known comparison procedures to determine alikeness for them (equilibrium, congruence, simultaneity...).

The comparison of magnitudes discussed so far only enables us to say if they are alike or not but, if they are not alike, it does not give any measure of their difference. If magnitudes have to be *completely specifiable* by numbers, 'the greater of the two numbers must be portrayable as the sum of the smaller and their difference' (p. 94). For this to be so, there must be some physical conjunction 'between magnitudes alike in kind [expressible] as an addition' (p. 94). It is curious that Helmholtz poses the problem of additivity without examining the question of order first, because if the procedure does not give rise to a certain order on the magnitudes one cannot speak of the highest and the

[10]The properties of symmetry and transitivity 'determine which physical relations we are allowed to recognize as alikeness' (p. 94).

[11]'Gleichartig' in the German original, that is to say, homogeneous.

lowest of two dissimilar ones and pose the problem of additivity in the way he does. The existence of an order does not follow from his conditions,[12] for the conditions he requires for alikeness do not preclude, for example, the same fact being observed for two dissimilar magnitudes when they are exchanged. That is to say, they do not imply the existence of an asymmetrical relationship between dissimilar magnitudes. We shall see how he analyzes 'greater than' after dealing with conjunction.

For the physical conjunction of magnitudes (i.e. of objects with a magnitude) to be similar to addition, three conditions must be satisfied. First, the magnitudes have to be of the same kind (i.e. of the same attribute).[13] Helmholtz thinks that it is obvious that substitutivity follows from this condition: the result of the conjunction does not change (i.e. it is alike) when one magnitude is exchanged for another alike one. But actually it is difficult to see why this follows, unless a previous relationship is established between conjunction and alikeness. We shall consider, then, the intended consequence as the first condition. If we use the standard notations '$\sim$' and '$\bullet$' for alikeness and conjunction, respectively, this condition, called $\sim$-monotonicity of $\bullet$, has the following form: (1) $a \sim b$ iff $(a \bullet c) \sim (b \bullet c)$.[14] The second condition is that the conjunction must be $\sim$-commutative: (2) $(a \bullet b) \sim (b \bullet a)$. The third condition is that $\bullet$ must be what he calls 'associative' (*sic*. p. 95): the result of the conjunction does not change, i.e. it is alike, when some 'compound' magnitude is substituted by another 'undivided' alike one. This use of the term 'associative' differs from normal usage, since the condition he refers to is: (3) if $(a \bullet b) \sim c$ then $((a \bullet b) \bullet d) \sim (c \bullet d)$.[15] Helmholtz points out that (3) follows from the previous ones, and is in fact redundant because it is of course a special case of (1) as it has been understood here.

It is at this point that Helmholtz refers to an order relation: once we have found a 'method of connecting the magnitudes (...) it now also follows which are greater and which are smaller (...), the whole is greater than the parts of which it is composed';[16] $a \bullet b$ is greater than a and than b. Of course this cannot be considered as a definition of 'greater than' for *any* two magnitudes, since it is not defined for non-compound objects. But in the spirit of what he says, it is easy to suggest a proper definition. Although it is not explicitly mentioned by

[12] 'The comparison method tells us only whether the magnitudes are alike or unalike' (p. 96).

[13] Because of the way the kind of a magnitude has been characterized, this condition is only of interest if the conjunction involves three or more objects, since what this condition requires is that the procedure which establishes the alikeness or not between any two of them is the same in all cases, and this is only restrictive if more than two are combined.

[14] This condition is a little bit stronger than mere substitutivity, which is only the 'if' part of this biconditional.

[15] 'The result of connexion should therefore not alter if I introduce, instead of some magnitudes to be connected, others which are alike with the sum of these' (p. 95).

[16] p. 96. And he adds, with respect to the abovementioned examples of magnitudes, that 'we have never doubted about what was greater and what smaller, *because* we have indeed known additive methods of connecting them' (my italics).

Helmholtz, we can take the following as an improvement of his original idea ('M' abbreviates 'greater than'): (D) aMb iff$_{\mathrm{def}}$ there exists c such that $a \sim b{\bullet}c$.[17]

Thus defined, it does follow from previous axioms that relation M is transitive and $\sim$-conservative (i.e. if aMb and $a \sim c$ then cMb). However it does not follow that M is asymmetric, nor that it is $\sim$-connected (i.e. either aMb or $a \sim b$ or bMa). And according to Helmholtz's aims it should follow, for he wants relation M to work in the domain of objects in the same way (with the same properties) as relation $>$ in the domain of numbers assigned to the objects. If we do not add new axioms about $\sim$ and $\bullet$,[18] it will be necessary, contrary to Helmholtz's intentions, to take an asymmetrical, transitive, $\sim$-connected and $\sim$-conservative relation M as primitively determined by the comparison procedure. Then we can add as a new condition the fact which he mentions and from which he attempted to 'infer' the order relation. This new condition is what we know as positivity of $\bullet$: (4) $a{\bullet}bMa$ and $a{\bullet}bMb$. Now we have all the conditions that characterize additive physical conjunction or 'physical addition'.[19] Whether or not a physical conjunction is additive can only be found out empirically; it will be additive if the conjunction satisfies the empirical 'definitional' conditions imposed.

We close this review of Helmholtz with some comments on three additional remarks he makes which are of interest for fundamental metrization. The first is that there are cases in which it is possible to find two different additive conjunctions (which suggests that there are two—types of—magnitudes), and for which, however, the comparison method that determines the alikeness of each of them is the same: 'by exactly the same method of comparison we determine both whether two wires are of like electrical resistance and whether they are of like conductance' (p. 96), but 'we add resistance by joining the wires in succession [i.e. in series] [...] and conductance by placing them side by side and joining them all up together at the one end and at the other [i.e. in parallel]'

[17]Note that this definition only defines '(strictly) greater than' if there is no null element for operation $\bullet$, i.e. if there is no c such that $a{\bullet}c \sim a$. If there were to be such an element, the definition should be '... iff there exists c such that $a \sim b{\bullet}c$ and not $a \sim a{\bullet}c$'.

[18]Even this strategy only works properly for asymmetry. It can be shown that we obtain the asymmetry of M thus defined if we add 'common' (not 'Helmholtzian') associativity of $\bullet$ to previous axioms. But there is no 'natural' condition about $\bullet$ and $\sim$ from which, adding other axioms and the suggested definition of M, $\sim$-connectedness for M follows. Of course there is one condition: either (i) there is c such that $a \sim b{\bullet}c$; or (ii) $a \sim b$; or (iii) there is c such that $b \sim a{\bullet}c$; but of course this is only a trick for it is merely connectedness 'written without M'.

[19]Although I said that *definition* (D) can be viewed as suggested by Helmholtz's work, I do not think we can take (D) as a new *axiom* suggested by Helmholtz if finally, as we did, take relation M as primitive and not defined by (D). As a new condition (taking M, with its properties, as primitive) the 'only if' part of (D) follows from the other conditions, but the 'if' part does not, and it is really a strong new condition, what we now call *solvability*: if aMb then there exists c such that $a \sim b{\bullet}c$. I do not think there is textual support for supposing that Helmholtz has anything like this in mind. On the other hand, although we have interpreted him as explicitly requiring the $\sim$-monotonicity of $\bullet$, I do not find textual support for doing the same with M-monotonicity: aMb iff $a{\bullet}cMb{\bullet}c$, even though this condition is not strong and probably he would have accepted it if he had considered the issue.

(pp. 96–97); and the same happens with condensers for their capacity and voltage. The strange thing is that the characterization that he has made of the type of magnitude seems to imply that they are of the same type (the same attribute), since the alikeness of each of them is determined by the same procedure. If the procedure were required, as we saw it was reasonable, also to determine the 'greater than' order there would be no problem since the orders are inverse.

Secondly, Helmholtz mentions measurement 'by components', or vectorial measurement, as a peculiar type of measurement which assumes that it is possible to additively compose magnitudes of different types (each one of the components) by means of a single physical operation on the objects; he mentions as examples the cases of velocity, acceleration, force and others (among which he includes colour, according to the theory of the three components). Helmholtz entitles this paragraph 'Adding magnitudes of different types' (p. 99), but it is clear that he is not attempting to suggest that magnitudes of one type can be composed with magnitudes of a different type. This section deals only with simultaneously adding different magnitudes, each one with others of the same type, using the same mode of physical combination. The type of representation involved here is what below will be called *multidimensional representation*.[20]

Finally, although at no time does he formally treat the question of whether his conditions are necessary and/or sufficient for the magnitudes to be 'completely specifiable' by numbers, he does explicitly mention the fact that the numbers thus obtained 'only have a proportional value' (p. 89). That is to say, they have no *absolute* representational value, they have representational value only as far as we express proportions or ratios with them. Nevertheless, they can be used absolutely when they are relativized to the value of the magnitude for an arbitrarily chosen standard object (unit).

Hölder on Axionatics and Real Morphisms

Hölder was the first to formally study the conditions that are necessary and/or sufficient for the numerical 'expression', or representation, of the facts occurring in a domain of objects because the objects displays certain property-magnitude.[21] These facts are understood here as being those that correspond to a certain order relation and a certain operation of combination on the elements. The numerical objects assigned to the objects are positive reals. The 'numerical facts' which express or represent the qualitative facts about the objects are those

[20]It is not clear, however, that the examples mentioned by Helmholtz deal with the simultaneous combination of *different* magnitudes.

[21]Cf. Hölder (1901). At about the same time, Huntington proposed a similar task (Huntington, 1902), but it was Hölder's work that was to be the guideline of subsequent research. Much later Wiener (1921) was also heading in the same direction.

which use the relation >and the operation + over the reals. And the 'expression', or representation, consists of a *complete translation*, that is to say, of an isomorphism. Hölder gives seven conditions or axioms that the domain, the relation and the operation must satisfy for there to be an isomorphism *onto* (not only *into*) the positive reals with >and +; this result is what is known as Hölder's Theorem. Among these conditions, he explicitly presents those of *solvability* (if an object is smaller than another, there is a third one that concatenated with the first is equivalent to the second) and *archimedianity*, or the Archimedean axiom (if an object is smaller than another, concatenating it with itself a finite number of times we can exceed the second one, that is to say, no element exceeds another 'infinitely'). Hölder's Theorem is purely mathematical and its empirical importance is slight, for the conditions that it imposes are excessive from an empirical point of view[22] (especially one condition similar to Dedekind completeness). In measurement, it is not essential, quite the contrary, for every real number to correspond to the magnitude of some object. Nor, on the other hand, does it seem reasonable to preclude two objects from having the same magnitude, but this exclusion follows from the fact that the assignment is an isomorphism, i.e. biunivocal. This last point is not so important, however, since the conditions can be regarded as referring to equivalence classes.

Hölder's results, suitably modified so that they can be used in empirical situations, make up the nucleus of most of the subsequent analyses of the conditions that make additive measurement possible. In the spirit of Hölder's work, an essential part of future standard analysis will be the search for axioms that an empirical system must satisfy for there to be a morphism (not necessarily *iso*morphism, in general it will be sufficient for it to be *homo*morphism) of such a system *into* (not necessarily *onto*) the reals. Because additive measurement is the paradigm of measurement, the analysis of its conditions will be the paradigm of metrization. In this sense, the 'spirit' of Hölder's work, the search for conditions for a real morphism, was to inspire the analysis of other types of measurement. It is hardly surprising, then, that research in metrization often seems to be a purely mathematical task.

Campbell on Order and Additivity

It is strange that N. Campbell, universally recognized as the father of MT, does not mention Hölder's results, even in his most important and monumental work, *Foundations of Science*, written almost 20 years later.[23] The reason may be the preferentially philosophical, not mathematical, orientation of his work. Campbell devotes the whole of the second part of his book, which originally should have had four parts but of which only two were finished, to the study of

[22]Its axioms constitute what nowadays is called, in model theory, a categorical theory: all its realizations are isomorphic, and then isomorphic to the additive semigroup of the positive reals.

[23]Campbell (1920), re-edited as Campbell (1957), from which I quote (cf. also Campbell (1928)).

measurement. In his work we find for the first time a systematic study of practically all the questions related to measurement and, among these, the conditions which make fundamental measurement possible. This is what concerns us here.

Campbell characterizes measurement as 'the process of assigning numbers to represent qualities' (1920, p. 267)[24] and the question that concerns us here is explicitly stated: 'Why can and do we measure some properties of bodies while we do not measure others?' (1920, p. 268; cf. also 1921, p. 111). The answer basically is that the measurable properties of bodies must in some special way be similar to the properties of numbers (1921, p. 112). What has to be clearly specified is what this way is.

The first condition for measurement is that the property generates an (1) asymmetrical and (2) transitive relation, i.e. an order relation, between the objects which display it. This relation, adds Campbell, must be such that if it does not connect two objects, these objects must be related with the others in the same way: (using the above conventions) (3) if not aMb and not bMa then aMc iff bMc. The objects which are not connected by the relation are regarded as being 'equal in respect of the property' (1920, p. 273; cf. also 1921, p. 5). A relation that satisfies these three conditions enables us to define another relation $\sim$ of alikeness or indifference: $a \sim b$ iff_{df} not aMb and not bMa. It is easy to see that, thus defined, $\sim$ is an equivalence relation and M is $\sim$-conservative and $\sim$-connected.

The fulfillment of these conditions allows a certain 'empirically informative' numerical representation. Such is the case of hardness and density (if the latter is measured without the aid of other magnitudes). This type of representation is, however, not very informative since the difference between the numbers assigned 'does not represent the physical difference' (1920, p. 274). For this difference to be represented, addition must have a physical interpretation, there must be a way of combining with characteristics 'analogous' to those of numerical addition. Should there be such a way, the property may be measured 'perfectly and definitively'.[25]

Campbell is not always uniform about the conditions that physical combination must satisfy for resembling addition, but we can take the following from

[24]Cf. also Campbell (1921, p. 110): '(measurement) can be defined, in general, as the assignment of numbers to represent properties'.

[25]'The difference between those properties which can be measured perfectly and definitively, like weight, and those which cannot arises in the possibility or impossibility of finding in connection with these properties a physical significance for the process of addition' (1920, pp. 277–278). Whether or not it is appropriate to speak of measurement in nonadditive cases we are not going to discuss here, since it is, partly, a verbal question. As a member of a committee of the *British Association for the Advancement of Science* responsible for analyzing the possibility of measurement in psychology, Campbell rejected such a possibility because of the absence of additivity (Ferguson *et al.*, 1940, p. 340; see the next section below). We have already seen, though, that in other places he speaks of measurement for mere orders (as in the case of hardness).

several places in which he deals with them (I again use '•' for the operation): (1) positivity: $a \bullet b M a$ and $a \bullet b M b$;[26] (2) $\sim$-commutativity and $\sim$-associativity; (3) $\sim$-monotonicity and M-monotonicity; and (4) 'By adding objects successively we must be able to make a standard series (of standard objects, i.e. a consecutive combination of objects which are alike) one member of which will be the same in respect of the property as any other object we want to measure' (1921, p. 117).[27] This last condition, which is extremely strong, implies the weaker ones of solvability and archimedianity, which are empirically more reasonable and enough to carry out the function for which Campbell introduces (4).

Again, as in Helmholtz, whether or not a physical combination • and a comparison relation M fulfil these conditions is a question that only experience can decide. One could think, says Campbell, that some of these conditions (for instance in the case of the measurement of weight using balances) can be deduced without experiment from known laws or principles (e.g. the laws of statics). But, he continues, research shows that our belief in the truth of these laws is based on our knowledge that the measurement of weight is possible, and so assumes that these conditions are fulfilled (1920, p. 286 and note). Campbell also raises the question of how arbitrary or univocal numerical assignment is, for a property which satisfies the conditions seen. If we consider a single mode of combination, he (informally) demonstrates that, given two different assignments such that in both cases the numbers assigned satisfy with respect to $>$and $+$ the conditions satisfied by the objects with respect to M and •, one is a multiple of, proportional to, the other. So, the values assigned are only arbitrary in the choice of the unit, i.e. the ratio of values is constant for any assignment.

It may happen, however, that there is more than one mode of combination that satisfies the conditions. The additional arbitrariness involved here, concerning which procedure is chosen, may only be apparent, since the order M, in relation to which the modes of combination satisfy the conditions, may be different. Campbell mentions the case already seen in Helmholtz of the combination of wires in series and in parallel. Both fulfil the conditions, but while one does so with respect to an order M, the other does so with respect to another M', the converse of the former. The properties that both procedures make it possible to measure are different, resistance and conductance as we saw. If two different combination procedures make it possible to measure the *same* property, it must happen that the order in relation to which they satisfy the conditions is the same in both cases. Although Campbell does not mention any cases, from what he says it follows that if there were to be one such a case, we would find ourselves faced with a new and uneliminable element of

[26]Another condition which he mentions is a special case of this one: if $a \sim b$ then not $a \bullet b \sim a$.
[27]In Campbell (1920) this condition is not *explicitly* required, but see p. 280.

arbitrariness: by which method of combination is 'the' property measured? *If* the question is well-posed (and it) will be depending on whether it makes sense to speak of 'the' property independently of *one* particular mode of combination), it is also crucial, since different procedures may give rise to non-proportional assignments. We shall see below that Ellis (1966) presents such a case.

Stevens on Scales, Transformations and Invariance

Between the works of Campbell and the publication in 1946 (over 20 years later, although most of the material is from the end of the 1930s) of 'On the Theory of Scales of Measurement', by S. S. Stevens, there is nothing of special relevance on the subject of metrization.[28] This article and later ones by the same author and his collaborators are a turning point in research on metrization and have a decisive influence on subsequent studies.

Stevens came from the field of psychology and his article is, in principle, a reply to the rejection by a committee of the *British Association for the Advancement of Science*, of which Campbell was a member, of the validity of certain scales for psychological magnitudes, such as the intensity of hearing sensations, on which Stevens himself had been working.[29] The rejection was based on the non-existence of an additive operation of concatenation for sensations. In the final report, Campbell states that, for measurement to be meaningful, the number assigned to the object must be able to be seen as the number of standards that when concatenated are alike to the object with respect to the property in question (cf. Ferguson *et al.*, p. 140), and another member concludes that 'any law purporting to express a quantitative relation between sensation intensity and stimulus intensity is not merely false but is in fact meaningless unless and until a meaning can be given to the concept of addition as applied to sensation' (Ferguson *et al.*, p. 145). It is not clear whether Stevens' work on sensorial scales should be interpreted as derived measurement of certain psychological magnitudes or as the establishment of psychophysical laws between independently measured magnitudes.[30] In any case, and even though Stevens' theory about scales is applied both to fundamental and derived measurement, it is essential to our discussion of the former.

Stevens, who does not want to argue about names, proposes to consider as measurement, in the broad sense, any assignment of numbers to objects or events following a rule. A scale is one of these assignments. Although he considers that scales are possible 'only because there is a certain isomorphism between what we can do with the aspects of the objects and the properties of the

[28]The work of Nagel in the 1930s, alone and in collaboration with Cohen, does not present anything new worthy of mention (cf. Nagel (1932), the second chapter of his doctoral dissertation Nagel (1930); and chapter XV, 'Measurement', of Nagel and Cohen (1934)).

[29]Cf. Stevens and Davis (1938).

[30]See, for instance, Roberts (1979, p. 142).

numerical series' (Stevens, 1946, p. 142), and that empirical operations that determine certain relations between the aspects of the objects are involved in the scales, he does not analyze either these empirical operations or the conditions they must satisfy; he only mentions 'what' the empirical operations should determine (alikeness, order, compared differences, ...). If despite all this his work is crucial for metrization, it is because what most concerns him and what he deals with, scale types, is essential for establishing the extent to which assignments (made possible by certain empirical conditions such as the ones studied by Helmholtz and Campbell) are unique or arbitrary.

The type of a scale is characterized by its transformation group, that is, by the transformations admissible for it. For each type of scale there are: (1) associated empirical operations which should determine certain facts, facts which must be preserved under the transformations; and (2) a permissible statistics (function), a measure of location. The well known classification is the following. Variables denote the values of the scale, the numbers assigned to the objects; $f(x)$ is the admissible transformation, a function of the numerical set, which contains the counterdomain of the scale, into itself;[31] and $\varphi(x_1, ..., x_n)$ is the numerical fact that empirical operations must determine, that is, the strongest formula $\varphi(x_1, ..., x_n)$ for which it is true that '$\varphi(x_1, ..., x_n)$ iff $\varphi(f(x_1), ..., f(x_n))$'.[32] The classification is cumulative, progressively stronger conditions being expressed.[33]

Nominal scale

$f(x)$ is any one-to-one function. φ is '$x_1=x_2$'. Statistical measure: mode. Example: any numeration. e.g. 'numbering' football players. Given the value of an object, the value of another is absolutely arbitrary. In this case we are not really measuring, we are merely renaming the objects.

Ordinal Scale

$f(x)$ is any increasing monotonous function. φ is '$x_1>x_2$'. Statistical measure: median. Example: hardness. Given the assignment of one object, the assignment of another is arbitrary as long as the order is preserved.

[31]The term 'transformation' will be used here with a certain ambiguity. Sometimes, as in this case, it will denote a function, with numerical domain and counterdomain, which is applied to another function with an empirical domain and a numerical counterdomain (i.e. to a scale); in this sense a transformation is the transformer function. Other times, when we state that a particular assignment of numbers to objects is a transformation of a previous assignment, it will denote the result of the composition, i.e. denote a new scale, with empirical domain and numerical counterdomain; in this sense a transformation is the result of applying the transformer function to a given scale. The context will make the intended sense clear.

[32]Coombs (1950) gives a classification for scales similar to Stevens' and takes as primitives, not the admissible transformations, but certain equations that roughly coincide with these φs. So, for example, an ordinal scale is one such that the equations for which it can be used are of the type $x \geq y$.

[33]Cf., for example, Stevens (1946, 1951, 1968, 1968b), Stevens (1959, p. 24 ff).

Intervals, or Differences, Scale

$f(x)=ax+b$ ($a>0$, i.e. positive linear transformations). φ is '$x_1-x_2=x_3-x_4$' (or 'x_1-x_2 is constant'). Statistical measure: arithmetic mean. Example: thermometric temperature, calendar time.[34] Assignment to one object determines the assignment of any other, once the origin or zero and the unit are arbitrarily chosen. The ratio of intervals of values does not change: $(x_1-x_2)/(x_3-x_4)=(f(x_1)-f(x_2))/(f(x_3)-f(x_4))$.

Proportional, or Ratio, Scale

$f(x)=ax$ ($a>0$, i.e. positive similar transformations). φ is '$x_1/x_2=x_3/x_4$' (or 'x_1/x_2 is constant'). Statistical measure: geometric mean. Example: length, mass, duration, thermodynamic temperature.[35] Assignment to one object determines the assignation of any other, once the unit is arbitrarily chosen (zero is absolute). The ratio of values does not change: $x_1/x_2=f(x_1)/f(x_2)$.

In some places Stevens adds another scale type which he locates, like the intervals scale, between the ordinal and the proportional:

Logarithmic Interval Scales

$f(x)=ax^n$ ($a>0$, $n>0$, i.e. exponential transformations). φ is '$\log x_1-\log x_2=\log x_3-\log x_4$' (or '$\log x_1-\log x_2$ is constant'). He says that he does not know any statistical measure specific to it and that there are no cases of the same in physics, although there are in psychology.[36] It is easy to see that here the ratios of logarithmic intervals do not change: $(\log x_1-\log x_2)/(\log x_3-\log x_4)=(f(\log x_1)-f(\log x_2))/(f(\log x_3)-f(\log x_4))$.

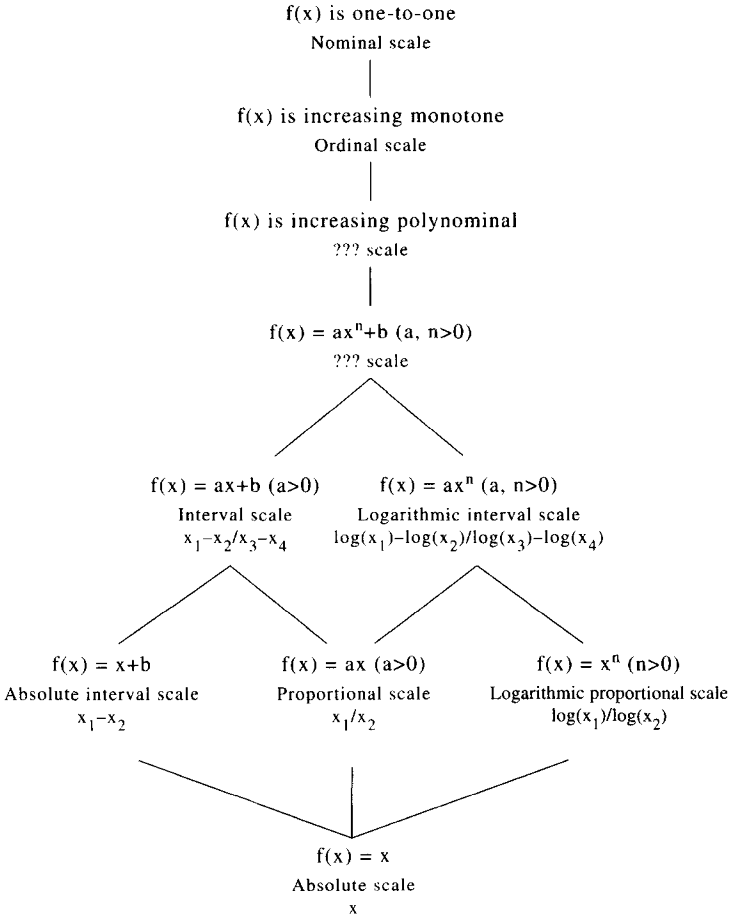
This classification can be completed in a natural way by introducing other types of transformation close to the ones seen. Each type of transformation $f(x)$ characterizes a scale type. Figure 1 summarizes the situation. The function, if any, that is beneath the name of the scale does not change; it is invariant up to the transformation $f(x)$, i.e. it is a function $g(x_1, ..., x_n)$ such that $g(x_1, ..., x_n)=g(f(x_1), ..., f(x_n))$. Lines connect transformations; lower transformations are special cases of higher ones with which they are connected, i.e. if a function g is invariant up to transformation $f(x)$ it will also be so under lower transformations connected to $f(x)$. When the sign '??? scale' appears in the place of the name of the scale, it is because the type added has no standard name.

This is Stevens' classification, with its natural extensions. To conclude this first part, we shall make some final remarks about the significance of Stevens' work for the previous research in metrization that we have already surveyed.

[34]Calendars are, strictly speaking, scales of a subtype of this one, absolute intervals scale (see Fig. 1).

[35]Now it does not concern us whether thermodynamic temperature is sensitive to fundamental metrization.

[36]It is not clear if what he wants to say is that there are psychological magnitudes for which exponential transformations are admissible or that there are laws of the form $s=a\mu^n$ where s is a psychological magnitude and μ the magnitude of the physical stimulus (see, for example, Stevens (1959 p. 36)).



And the best way to show this significance is to point out, first, that Stevens' general approach is unsatisfactory in an essential way or, less crudely, it is in essential need of supplementation. For, although Stevens does a great deal of work on transformations and scale types, in the end the precise sense in which a specific scale is of a certain type remains unclear. A scale, he says, is of type, say, *A* if its *admissible transformations* are of type, say, *a*. But, when (and why) is a transformation admissible for a scale? When it leaves invariant something relative to the scale. But, what is *it*? Saying that it is the scale form or structure, with no further clarification, is to say nothing. It cannot be what he calls its permissible statistics measure, for they are not always invariant.[37] What is it

[37]For instance, for arithmetic mean and positive linear transformations the following is obtained: $((ax_1+b)+(ax_2+b))/2=a((x_1+x_2)/2)+b$, that is, $g(f(x_1), \ldots, f(x_n))=f(g(x_1, \ldots, x_n))$. So Stevens distinguishes two senses of invariance for *a* statistics: (a) that it does not change its numerical value; and (b) that it changes the value but not the designated item (cf. Stevens, 1959, pp. 27–28).

then? We have seen that most transformations leave invariant some relatively simple function. But even if this were so for all transformation types, it would not provide a satisfactory answer. For, what do these functions have to do with the scale *before knowing* that the scale is of a certain type? That some functions are invariant under some transformations is a purely mathematical fact. So, if we characterize a particular scale type by associating to it a particular group of transformations, it follows by purely (and relatively simple) mathematical considerations that certain numerical functions are invariant. But what we are then doing is simply *defining* the scale type by what we *call* (with no further explanation) 'the type of admissible transformation' for the scale, and it remains entirely obscure what makes a transformation *admissible* for a scale, i.e. why some transformations of a certain scale of magnitude *m are also scales of the very same magnitude m* and some other transformations are not. If we proceed in this way, and this seems to be the way in which Stevens proceeds, we cannot know, for example, why a non-positive linear transformation of the Celsius scale does not measure temperature, or why a (non-similar) monotone transformation of the MKS scale does not measure mass.[38]

To sum up, if we *define* the scale type by a type of transformation (or what is tantamount, by a representative invariant function), then we do not know, in the epistemologically relevant sense, what makes a transformation admissible for a scale, why some transformations of a scale are measures of the magnitude and others are not. But this is just what we want of a clarification of a scale type. And if we try to clarify the notion of admissible transformation by appealing to the invariance of certain numerical functions, we still want to know what these numerical functions have to do with scale types. And this is just what we require of a clarification of an admissible transformation. To get out of this circle it is necessary, in the definition of admissible transformation, to get away from purely mathematical invariances and appeal to the properties of the objects, to their empirical relations. For this reason, perhaps, Stevens states that in each scale type *empirical* operations must determine certain *numerical* facts that remain invariant under transformations of the type corresponding to the scale, but he says nothing about these empirical operations and the *empirical* facts which determine those other numerical facts that remain invariant. Once this task, which is precisely the task begun by Helmholtz, Hölder and Campbell, has been done, Stevens' results will be essential to see the extent to which the numerical representation is unique and, therefore, what is the use we can make of it.

Helmholtz, Hölder and Campbell analyzed the qualitative conditions that an empirical system must satisfy in order to be numerically representable, but they

[38]Mundy (1986, p. 403 ff.) characterizes Stevens' approach as dealing with what he calls 'unstructured representation'. As far as I understand Mundy's criticisms of unstructured representation, my criticisms to Stevens are along the same line (cf. esp. *op. cit.*, p. 406).

did not say anything about the relations between the different possible representations of the same empirical system. Stevens formally studies the different formal relations there are between different representation-scales of a magnitude (i.e. of an empirical system), but he does not say anything about why the representations which show this relation are representations of the same magnitude. The answer cannot be that certain functions are invariant, since this is simply another way of characterizing the relation between transformations. To give an appropriate answer to this question it is necessary to make reference to the empirical conditions the system satisfies. If a transformation of a scale for magnitude *m* is admissible, it is because the function which results from the transformation is also a representation-morphism of the empirical system. This is the link that is lacking between the two approaches, the bridge which unifies these two lines of research of the formation period of our theory.

Acknowledgements—Research for this work has been supported by the DGICYT Research Projects PB92-0846-06 and PB93-0655, Spanish Ministry of Education.

References

Berka, K. (1983) *Measurement. Its Concepts, Theories and Problems* (Dordrecht: Reidel).

Campbell, N. (1921) *What is Science?* (New York: Dover edn 1953).

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

Campbell, N. (1920/57) *Foundations of Science: The Philosophy of Theory and Experiment* (New York: Dover).

Coombs, C. H. (1950) 'Psychological Scaling Without a Unit of Measurement', *Psychological Review* **57,** 145–158.

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

Ferguson, A. *et al.* (1940) 'Quantitative Estimates of Sensory Events', *Report of the British Association for the Advancement of Science* **2,** 331–349.

Helmholtz, H. V. (1887) 'Zahlen und Messen, erkenntnistheoretish betrachtet', in Helmholtz, *Schriften zur Erkenntnistheorie*, pp. 70–108. Engl. trans., 'Numbering and Measuring from an Epistemological Viewpoint', in Helmholtz, *Epistemological Writings*, pp. 72–114.

Helmholtz, H. V. (1921) *Schriften zur Erkenntnistheorie*, Julius Springer, Berlin. Engl. trans., *Epistemological Writings* (Dordrecht: Reidel, 1977).

Helmholtz, H. V. (1930) *Counting and Measuring* (Princeton: Van Nostrand).

Hempel, C. G. (1952) *Fundamentals of Concept Formation in Empirical Sciences* (University of Chicago Press).

Hölder, O. (1901) 'Die Axiome der Quantität und die Lehre vom Mass', *Berichte über die Verhhandlungen der königliche sächsischen Akademie der Wissenschaften zu Leipzig Math.-Phys. Classe* **53,** 1–64.

Huntington, E. V. (1902) 'A Complete Set of Postulates for the Theory of Absolute Continuous Magnitude', *Transactions of the American Mathematical Society* **3,** 264–284.

Krantz, D., Luce, D., Suppes, P. and Wersky, A. (1971) *Foundations of Measurement*, Vol. I (New York: Academic Press).

Luce, D., Krantz, D., Suppes, P. and Wersky, A. (1990) *Foundations of Measurement*, Vol. 3 (New York: Academic Press).

Mundy, B. (1986) 'On the General Theory of Meaningful Representation', Synthese **67,** 391–437.

Nagel, E. (1930) *On the Logic of Measurement* (Stanford University Press).

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

Nagel, E. and Cohen, M. R. (1934) *An Introduction to Logic and Scientific Method* (New York: Harcourt).

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

Roberts, F. (1979) *Measurement Theory* (MA: Addison).

Stegmüller, W. (1970) *Theorie und Erfahrung* (Heidelberg: Springer).

Stevens, S. S. (1946) 'On the Theory of Scales of Measurement', *Science* **103,** 667–680.

Stevens, S. S. (1951) 'Mathematics, Measurement and Psychophysics', in *Handbook of Experimental Psychology*, p. 149.

Stevens, S. S. (1959) 'Measurement, Psychophysics and Utility', in Churchman and Ratoosh (eds), *Measurement: Definitions and Theories*, pp. 18–63.

Stevens, S. S. (1968a) 'Measurement, Statistics and the Schemampiric View', *Science* **161,** 849–856.

Stevens, S. S. (1968b) 'Ratio Scales of Opinion', in Whitla (ed.), *Handbook of Measurement and Assessment in Behavioral Sciences*, pp. 171–199.

Stevens, S. S. and Davis, H. (1938) *Hearing: Its Psychology and Physiology* (New York: Wiley).

Suppes, P. (1951) 'A Set of Independent Axioms for Extensive Quantities', *Portugaliae Mathematica* **10,** 163–172.

Suppes, P. and Zinnes, S. (1963) 'Basic Measurement Theory', in Luce, Bush and Gallanter (eds), *Handbook of Mathematical Psychology 1*, pp. 3–76.

Suppes, P., Krantz, D., Luce, D. and Wersky, A. (1989) *Foundations of Measurement*, Vol. 2 (New York: Academic Press).

Wiener, N. (1921) 'A New Theory of Measurement: A Study in the Logic of Mathematics', *Proceedings of the London Mathematical Society* **19,** 181–205.