2. OBJECTIVITY

This chapter introduces the essentials of objectivity (also known as monotonicity, composite transitivity, conjoint additivity and fundamental measurement), and to deduce the measurement model that objectivity requires.

The progress of science depends on the invention, construction and maintenance of useful measures. Science lives on measurement. Measurement exists on objectivity. An everyday term for objectivity is generality. Objectivity is the expectation and, hence, requirement that the amount and meaning of a measure has been well enough separated from the measuring instrument and the occasion of measurement so that the measure can be used as a quantity without qualification as to which was the particular instrument or what was the specific occasion.

Although a measuring occasion is necessary for a measure to result, the utility of the measure, depends on the specifics of the occasion disappearing from consideration. It must be possible to take the occasion for granted and, for a time being, to forget about it. Were such a separation of meaning from the circumstances of its occasion not possible, not only science but also commerce, and even communication, would become impossible.

The essentials of measurement can be brought out by reviewing the characteristics of the archetypical variable, length. When you ask a person's height and it is reported 70 inches, you do not demand the yardstick or to know who made the measure, when or where. You expect, and hence require as a precondition for continuing communication concerning height, that 70 inches was obtained in the usual way. Even though you know what the circumstances necessary to produce the measure were necessarily fraught with unique particulars, you use the 70 inch quantification of length as though it were entirely independent of those circumstances of its construction. In other words, you take 70 inches to be objective. Were you unable to do that, the quantity 70 inches would become meaningless not only to you but also to anyone, for nobody would know to what, if any, enduring state it referred.

Measuring length is so familiar and commonplace that the way we do it seems obvious. We are tempted to think of length as an explicit, manifest variable that can be seen directly. But there are essential details which the measurement of length requires. Although these details are taken for granted, they cannot be neglected, if length is actually to be measured. In fact, that they can be taken for granted signifies that we have made a solid habit of not neglecting them.

In spite of its "looks," length is not, by itself, manifest. Nor, in fact, is there any variable at all which is manifest on its own. Variables are inventions and measurements are constructions. An agent of measurement, a ruler of some kind, is necessary to make length "visible." Length cannot be "seen" on its own, let alone measured, without the deployment of some kind of ruler. This requires the measurement of length to be a conjoint operation. The calibrated ruler and the thing to be measured must be brought into a disciplined conjunction. The ruler, through its calibration, recapitulates the founding definition of the variable "length." The ruler's calibrations are the criterion definition of this variable. The ruler, while necessarily concrete in its realization of "length," depends for its utility on the extent to which it implements an abstract fiction. It must not matter at all which particular concrete realization of a "ruler" is actually used to make the abstract measurement. It must only need be any "ruler" in good standing.

All measurements made by all calibrated rulers must be quantitatively comparable without any reference to the physical details or work histories of the particular rulers used or who used them.

SOCIAL SCIENCE MEASUREMENT

These ideas are not new to social science. To be generally useful, the individual measure must not depend on which particular test items are used.

> It should be possible to omit several test questions at different levels of the scale without affecting the individual score.
>
> It should not be required to submit every subject to the whole range of the scale. The starting point and the terminal point, being selected by the examiner, should not directly affect the individual score (Thurstone, 1926, p. 446).

Nor should the measuring function of a test, that is, the calibrations of the test items, depend on which particular persons are being measured.

The scale must transcend the group measured. One crucial experimental test must be applied to our method of measuring attitudes before it can be accepted as valid. A measuring instrument must not be seriously affected in its measuring function by the object of measurement. To the extent that its measuring function is so affected, the validity of the instrument is impaired or limited.

If a yardstick measured differently because of the fact that it was a rug, a picture, or a piece of paper that was being measured, then to that extent the trustworthiness of that yardstick as a measuring device would be impaired.

> Within the range of objects for which the measuring instrument is intended, its function must be independent of the objects of measurement (Thurstone, 1928, p. 547).

Indeed, Thurstone's eloquent and detailed 1931 specification of the essentials of measurement meets and resolves most of the "big" measurement misgivings that social scientists continue to fret about.

Measurement is Necessarily One-Dimensional:

> One of the most frequent questions (concerning the possibility of social measurement) is that a score on an attitude scale, let us say the scale of attitude toward God, does not truly describe the person's attitude.
>
> There are so many complex factors involved in a person's attitude on any social issue that it cannot be adequately described by a simple number such as a score on some sort of test or scale. This is quite true, but it is also equally true of all measurement.
>
> The measurement of any object or entity describes only one attribute of the object

measured. This is a universal characteristic of all measurement. When the height of a table is measured, the whole table has not been described but only that attribute which has been measured.

Similarly, in the measurement of attitudes, only one characteristic of the attitude is described by a measurement of it.

Measurement is Necessarily Linear:

Only those characteristics can be described by measurement which can be thought of as linear magnitudes. In this context, linear magnitudes are weight, length, volume, temperature, amount of education, intelligence, and strength of feeling favorable to an object. Another way of saying the same thing is to note that the measurement of an object is, in effect, to allocate the object to a point on an abstract continuum. If the continuum is weight, then individuals may be allocated to an abstract continuum of weight, one direction represents small weight while the opposite direction represents large weight.

Measurement is Necessarily Abstract:

The linear continuum which is implied in all measurement is always an abstraction. For example, when several people are described as to their weight, each person is in effect allocated to a point on an abstract continuum of weight. All measurement implies the reduction or restatement of the attribute measured to an abstract linear form. There is a popular fallacy that a unit of measurement is a thing such as a piece of yardstick. This is not so. A unit of measurement is always a process of some kind which can be repeated without modification in the different parts of the measurement continuum (Thurstone. 1931, p.257).

But no ruler in its concrete embodiment of the abstract idea of length does its job without further specification. There are rules concerning how rulers must be employed to produce acceptable measures. The ruler and the object to be measured must be carefully aligned so that they lie parallel to one another. A starting point, or origin, and units to count must be installed. The line of sight along which the viewer reads the object against the ruler must be determined and maintained. The procedure by which coincidence is identified and interpolation accomplished must be specified. Without care for these rules, the results of ruler measurements become too disorderly to be useful.

AXIOMATIC MEASUREMENT THEORY

The axiomatic theory of measurement has made great strides in the past 30 years. There are detailed and scholarly discussions of these accomplishments in print. Unfortunately these discussions are too esoteric for most social scientists. It is hard for practitioners to see how to put axiomatic measurement theory to work.

The heart of axiomatic measurement theory, however, can be simply put. The crucial axiom which all measurement theorists agree is necessary for the construction of measurement is the one they call "monotonicity" or "conjoint additivity."

This axiom can be useful to social scientists because it marks out exactly the condition which both scientist and layman expect of numbers which are intended to serve as measures, namely generality or objectivity.

The joint ordering of conjoint additivity is also not new to social science. Monotonicity under the name of "conformity" and later "objectivity" appears in the practical work of Georg Rasch in 1953 and is defined, developed and implemented in detail in his seminal book of 1960 (Rasch, 1960/1980) and article of 1961 (Rasch, 1961).

> A person having a greater ability than another should have the greater probability of solving any item of the type in question, and similarly, one item being more difficult than another one means that for any person the probability of solving the second item correctly is the greater one (Rasch, 1960, p. 117).

Rasch "objectivity" is a stochastic conjoint additivity. Even earlier in 1944, Louis Guttman (1944, 1950) formulated what must be the best known, but least followed, requirement for social science measurement. Guttman deduced that a score could not be unequivocally on a "scale," unless the particular data from which the score was accumulated were completely specified by the value of the score.

> If a person endorses a more extreme statement, he should endorse all less extreme statements if the statements are to be considered a scale.
>
> We shall call a set of items of common content a scale if a person with a higher rank than another person is just as high or higher on every item than the other person (Guttman, 1950, p. 62).

Guttman "scalability," a deterministic conjoint additivity, is impractical when applied deterministically. But its stochastic version is identical to Rasch's objectivity and entirely practical, as Rasch demonstrated in the 1950's and as has been shown so many times since for hundreds of tests and questionnaires (Wright and Bell, 1984).

What may not be quite as obvious is that the stochastic version of Guttman's requirement is equivalent to Ronald Fisher's seminal definition of a sufficient statistic (1958/1922). Fisher's "sufficient" statistic is the one and only statistic that exhausts the information modeled in the data with respect to the parameter to be estimated.

What this definition means is that a Fisher "sufficient" statistic is the statistic that provides the best stochastic reconstruction of the data. This is exactly Guttman's scalability criteria, expressed stochastically. The realization that Fisher "sufficiency" is a necessary concomitant of stochastic monotonicity may prove, in the end, to be the decisive reason for preferring sufficient statistics over all others.

The common sense of this, so often reiterated, foundation for measurement is plain enough. It would seem that no sane researcher could argue or act otherwise. Yet, and strangely, few social scientists require or even hope for conjoint additivity in the numbers they use as "measures."

The consequence of this innocent carelessness is a plethora of ill-defined and unstable pseudo-quantifications and a great deal of confusion and disappointment over ambiguous and irreproducible results.

This unhappy situation is completely unnecessary. A derivation of a practical stochastic measurement model from the requirement of monotonicity, conjoint additivity or objectivity is easy to follow and the resulting model for measurement is easy to apply.

Here is a simple derivation of the model necessary to meet Thurston's 1928 requirement that a scale be independent of the objects of measurement.

THURSTONE INVARIANCE

The construction of a scale depends on the relative calibrations of the items used to define the scale. These calibrations must be established in a way that can be made independent of which persons happen to provide the calibration data. We begin by asking what is required so that the comparison of any two items i and j will be independent of whatever persons are used to elicit evidence of the relative scale standing of these two items?

Items *i* and *j* can be observed to differ only when they are answered differently. Realizing a comparison of *i* and *j*, then, requires counting how often *i* is answered 'yes' by persons when *j* is simultaneously answered 'no' and comparing this "$i > j$" count with the reciprocal "$j > i$" count of how often the reverse occurs among other persons.

The estimation of a quantitative comparison of items *i* and *j* from this pair of reciprocal counts requires a probability model for the occurrence of the counts which can implement an objective, i.e. sample-free, person-invariant, comparison of their probabilities.

The pair of probabilities can be represented by

$$\Pr[(i = yes),\ (j = no)]$$

and

$$\Pr[(i = no),\ (j = yes)]$$

and their comparison specified by the ratio,

$$\frac{\Pr[(i = yes),\ (j = no)]}{\Pr[(i = no),\ (j = yes)]} \qquad 2.1$$

Let $P_{ni} = f(n,i)$ be the, as yet undefined, probability that person n succeeds on item i.

What we seek is the particular function $f(n,i)$ which maintains Thurstone (1928) invariance and hence Rasch (1960/1980) objectivity.

To obtain invariance the comparison of probabilities in Equation 1 must stay the same regardless of which persons are involved. That is, Equation 1 must hold for any suitable persons n or m as in,

$$\frac{\Pr[(i=yes),(j=no)]}{\Pr[(i=no),(j=yes)]}=\frac{P_{ni}(1-P_{nj})}{(1-P_{ni})P_{nj}}\equiv\frac{P_{mi}(1-P_{mj})}{(1-P_{mi})P_{mj}} \qquad 2.2$$

for all n and m

where n is some person, m is any other person and the symbol $\equiv$ specifies that the comparison of item i with j is "defined" to remain the same whoever the persons, n or m.

To simplify our appreciation of the implications of Equation 2 for $P_{ni}=f(n,i)$, we can choose $j=o$ and $m=o$ as origins for the item and person scales so that the calibration of item i becomes its comparison with a reference item $j=o$ and the measure of person n becomes their comparison with a reference person $m=o$.

We can also align these scale origins so that the reference person has a fifty-fifty chance to succeed on the reference item. This makes

$$P_{mj}=P_{oo}=1/2 \text{ and } (1-P_{oo})/P_{oo}=1$$

When we insert $j=o$ and $m=o$ into Equation 2-2 and solve the middle and right side for the odds of person n succeeding on item i we get

$$\frac{P_{ni}}{(1-P_{ni})}=\frac{P_{no}}{(1-P_{no})}\frac{P_{oi}}{(1-P_{oi})}=g(n)*h(i) \qquad 2.3$$

$(P_{no})/(1-P_{no})$ has a value between 0 and infinity depending only on person n, and $[P_{oi}/(1-P_{oi})]$ has a value between 0 and infinity depending only on item i.

The measurement scale defined by Equation 2-3 is a ratio scale. Zero corresponds to the measure for a person having no chance of success on any item and also to the calibration of an item on which there is no chance of success by any person.

The ratio scale defined by $P_{ni}/(1-P_{ni})$ can be transformed into an equal-interval linear difference scale by taking logorithms.

$$\log[P_{ni}/(1-P_{ni})]=\log[P_{no}/(1-P_{no})]+\log[P_{oi}/(1-P_{oi})] \qquad 2.4$$

$$=G(n)+H(i) \text{ for an interval scale}$$

$$=B_n-D_i \text{ for convenience}$$

or

$$P_{ni} \equiv \exp(B_n - D_i) / [1 + \exp(B_n - D_i)$$ 2.5

where the item calibration D_i depends only on the attributes of item i, which we can call its difficulty, and the measure B_n depends only on the attribute of person n, which we can call his ability.

This model relating the ability of person n and the difficulty of item i to the performance of person n on item i is the objective model of measurement known as the Rasch model.

This deduction arrives at the only *f(n, i)* which can support the construction of Thurstone invariant or Rasch objective scales.

Equation 2-2 can be rewritten to address Thurstone's concomitant 1926 requirement that the individual measure not depend on which particular items are used so that it becomes "possible to omit several test questions at different levels of the scale without affecting the individual score." This requires that the comparison of any pair of persons n and m be invariant with respect to the particular items employed as in

$$\frac{\Pr[(i = yes),\ (j = no)]}{\Pr[(i = no),\ (j = yes)]} = \frac{P_{ni}(1 - P_{mi})}{(1 - P_{ni})P_{mi}} \equiv \frac{P_{nj}(1 - P_{mj})}{(1 - P_{nj})P_{mj}}$$ 2.6

for all i and j

which is equivalent to Equation 2-2 and so leads to Equation 2-5.

MEASUREMENT ESSENTIALS

2nd Edition

BENJAMIN WRIGHT

MARK STONE

WIDE RANGE, INC.
Wilmington, Delaware