On Objectivity and Models for Measuring

By

G. Rasch

Lecture notes edited by Jon Stene.

On Objectivity and Models for Measuring

By

G. Rasch

Lectures notes edited by Jon Stene.

---

1. The Basic Problem.

Among the empirical sciences physics enjoys an exceptional status as regards exactness and objectivity. Unquestionably the fact that the observations in most fields of physics are quantitative and can be measured with great precision has contributed immensely for physics to acquiring this status, so it is no wonder if students of other fields of knowledge look forward to measuring as an indispensable instrument for raising their sciences to a state comparable to that of physics.

Particularly from psychological and sociological quarters the concept of measurement has been held to be necessary prerequisite for mathematical formalizations, as the hallmark of "real science". On the other hand many students of the social sciences refute the wish for mathematization as an unwarranted imitation of a natural science which largely works quantitatively, while most data in psychology and sociology, not to speak of humanities, are qualitative - and remain so, even if numbers are assigned to them - to which situation the methods within these fields should be adapted.

The latter point is certainly not without justification, and it may even be added that general structures- descriptive relations, causal explanations, deductive systems, etc. - in humanities more often than not are also qualitative, while in physis most structures are expressed in terms of differential equations and other equations in quantitative variables.

The question may, therefore, be raised whether the kinds of knowledge acquired in Natural Sciences on one hand and in Social Sciences and Humanities on the other hand are on the whole fundamentally different - if really not only the observational material, but also ends and means have almost nothing in common.

In an attempt at answering this question in the negative the present lectures suggest a conceptual framework that - notwithstanding the vast differences in types of problems and of data - may be taken as a common basis for at any rate large areas within the said fields of Human Knowledge.

In consequences of the contrasts pointed out it seems obvious that a framework as comprehensive as intended could not count measurement among its fundamental concepts; if at all evolving it would appear as a derived concept.

Looking then for concepts that could possibly be taken as primary it seems worth-while to concentrate upon two essential characteristics of "scientific statements":

1. they are concerned with "comparisons";
2. the statements are claimed to be "objective";

both terms of course calling for precise qualifications.

2. Systematic Comparisons.

Generally speaking a comparison is a statement about two or more objects, and so comparing seems to be a most fundamental activity in acquiring any sort of knowledge, ranging from identifications and classifications that are necessary for collecting and presenting data, to formulating more or less broad syntheses and theories.

A characteristic feature of "scientific statements" is that they do not deal with isolated observations, but enter in a comprehensive set of comparisons in which the objects belong to a certain class of objects, defined by a criterion according to which it can be decided whether any given object does or does not belong to the class.

This class, $\mathcal{O}$, may comprise only a finite number of elements or potentially it may be infinite, although only a finite number of objects enter into any actual investigation.

Having delimited a class of objects our next question is which sort of comparisons to make, depending, of course, upon which properties of the objects should be compared. One way of bringing forth such properties is to expose the objects to various conditions, the results of which may then be used for formulating comparative statements. However, it quite often happens that under some conditions two objects seem to react in much the same way, while under different conditions they diverge clearly from each other. Therefore a reliable comparison requires a wide variation of the conditions in question. On the other hand all the conditions employed should bear upon the same group of properties of the objects considered.

Both concerns may be taken care of if we confine the conditions to a certain class, $\mathcal{A}$, of elements which we shall call "agents". This class which, like $\mathcal{O}$, should be defined by an unambiguous criterion may have a finite or an infinite number of elements, but of course in any actual investigation only a finite number of objects and a finite number of elements are brought into contact with each other. However, for the concepts we are going to develop it is essential that any object belonging to $\mathcal{O}$ could contact any agent belonging to $\mathcal{A}$.

Now, each contact usually results in a lot of different reactions, among which only one (possibly multidimensional, however,) is recorded as a basis for the comparisons. Of course any type of result may be used for a comparative purpose, but then the comparisons will be concerned with different qualities of the objects. Thus, specifying which kind of results should be recorded is just as important for characterizing the comparisons to be achieved as is the choice of agents.

This point may be illustrated by way of an example from educational testing. The objects may be school-children at various ages and grades to be compared as regards their abilities in reading aloud, say. The agents may be ranging from very easy ones to most difficult ones - both qualifications supposed to be readily accepted by teachers in mother's language.

Usually a test result is recorded as the number of erroneously read words, i.e. a natural number not exceeding a certain maximum. On this basis the children may be compared as regards their abilities of reading correctly.

For special purposes the kinds of errors may have been somehow classified and the number of errors of each type may be counted, an observation then being a vector with natural numbers as elements. In this case the testing has a more diagnostic aspect.

Often also the time used for reading the whole text or the number of words read within a fixed time period is recorded, yielding the reading speed as a real positive number.

Thus three distinct modes of observation are readily available, but that means that the children could be compared with regard to three quite different aspects of reading, and it also means that ideally the choice among the modes should be directed by the kind of comparisons aimed at. In practice, however, our limited knowledge may often twist this direction the other way round.

Having already restricted the observational situation to contacts between elements of a well defined class of objects - thus "similar" in some sense - and elements of a well defined class of agents - also somehow "similar" - we add a further "similarity" limitation by requiring that the result of a potential contact of any object in $\mathcal{O}$ and any agent in $\mathcal{A}$ must belong to a fixed set, $\mathcal{R}$, of elements.

In physics $\mathcal{R}$ often is the set of real, positive numbers - as opposed to intelligence testing where the answer to each question often is just "right" or "wrong".

In the following we shall, for the sake of simplicity, confine ourselves to cases where only a finite number of results can occur:

$$\mathcal{R} : \{ R^{(1)}, R^{(2)}, \ldots, R^{(m)} \} \quad .$$

The extension to enumerable sets being immediate, these sets are well suited for recording the qualitative data of Social Sciences and Humanities in numerous cases.

In order to cover quantitative data a theory for $\mathcal{R}$ being a Euclidean space of finite dimension would be needed, but this problem we shall defer to some other occasion.

Meanwhile it may be noticed that, by stretching the argument a bit, we may formally consider quantitative observations as qualitative, in the last analysis depending upon a purely positional statement: location of the interval between two marks (visible or imagined) on a measuring scale to which a point on an observed object is decided to belong.

3. Specific Objectivity.

Through the preceding discussion an external framework for carrying out comparisons has been established, namely a set of objects, $\mathcal{O}$, and a set of agents, $\mathcal{A}$, which together form a set of contacts, $\mathcal{C}$, leading to a set of results, $\mathcal{R}$:

$$(3.5) \qquad \mathcal{O} \times \mathcal{A} = \mathcal{C} \rightarrow \mathcal{R}$$

$\mathcal{O}$, $\mathcal{A}$ and $\mathcal{C}$ may and usually do contain an infinity of elements, while $\mathcal{R}$ may be a finite set and even if it is infinite the same element of $\mathcal{R}$ may occur in an infinity of contacts.

A practical investigation, of course, only implies a finite numer of objects, agents and contacts which for a simple design may be arranged in a rectangular scheme:

$$(3.2) \qquad \begin{array}{c|c} & \text{Agents} \\ & A_1, A_2, \ldots, A_i, \ldots, A_k \\ \hline \text{Objects} \quad \begin{array}{c} O_1 \\ O_2 \\ \vdots \\ O_\nu \\ \vdots \\ O_n \end{array} & C_{\nu i} \rightarrow R_{\nu i} \end{array}$$

where each $O_\nu \in \mathcal{O}$, each $A_i \in \mathcal{A}$, $C_{\nu i} = (O_\nu, A_i) \in \mathcal{C}$ and each $R_{\nu i} \in \mathcal{R}$.

An essential point in the present approach is that it is the objects O_ν, we compare, not the observations $R_{\nu i}$, these being only instrumental for comparing the objects with respect to how they react upon contacts with the agents.

In order to illustrate the sort of problems that is implied in this point we may return to the above example from educational testing.

In ordinary test situationsonly one text, suitably chosen, will be presented to each child. Taken by itself the value of comparisons based upon a single text is very limited, the risk being obvious that the result would be quite different if another text had been used. In order to make clear if children can at all be meaningfully compared with respect to reading ability we need an extensive experiment in which every child of a group $O_1, O_2, \ldots, O_n$ belonging to $\mathcal{O}$ was exposed to the whole series of texts representing $\mathcal{A}$*. On examining the re-

---

* In practice, of course, there would be no point in presenting a difficult text to kids that have barely learned to read,neither in presenting texts suitable for say, second grade to pupils in secondary school, but avoiding such malpractices requires a certain sophistication of the theory here developed (cf. "Chaining of tests", [1], p.29).

sults it should then become evident whether the comparison between the children yields the same result for all of the texts; it not, the value of such comparisons would seem questionable. Furthermore, for a pure comparison of some children prompting or other unwarranted influences from the other children must of course be excluded.

Transferring these considerations to the general framework and formalizing them we reach at the following rules for valid comparisons:

The comparison within (3.2) of, say, two objects O_1 and O_2 is a statement about them to be derived from the whole set of records $R_{\nu i}$. This means that a certain procedure has been prescribed which from any matrix of answers

(3.3) $\quad R = ((R_{\nu i})), \quad \nu = 1,\dots,n, \; i = 1, \dots,k$

leads to a statement

(3.4) $\quad \mathcal{U}\,\{O_1, O_2 | R\}$

about O_1 and O_2 in so far as these objects are included in (3.2).

This statement should then be independent of which objects other than O_1 and O_2 enter into the observational scheme (3.2), and also of which agents are used. This means that if $O_3,\dots,O_n$ were replaced by any other set of elements $O_3',\dots,O_n'$, of $\mathcal{O}$, giving rise to another matrix of results R' to which the same procedure was applied then we should have

(3.5) $\quad \mathcal{U}\,\{O_1, O_2 | R'\} \; = \; \mathcal{U}\,\{O_1, O_2 | R\} \quad .$

Furthermore, if $A_1,\dots,A_k$ were replaced by any other set of elements $A_1',\dots,A_k'$, of $\mathcal{A}$, giving rise to still another matrix of results R" then we should have

(3.6) $\quad \mathcal{U}\,\{O_1, O_2 | R''\} \; = \; \mathcal{U}\,\{O_1, O_2 | R\} \quad .$

A comparison of two (or more) objects satisfying these exacting conditions is <u>objective in the sense of being uninfluenced by factors within the framework of $[\mathcal{O}, \mathcal{A}, \mathcal{R}]$ which are extraneous to the objects compared</u>.

To the term "objective", however, a number of different meanings have been attached. To distinguish the kind of objectivity just described we shall therefore add the qualification "specific" - thereby indicating that it has been specified in a particular way, viz. through the $[\mathcal{O}, \mathcal{A}, \mathcal{R}]$ - framework.

In an empirical science specific objectivity can never be fully ascertained if $\mathcal{O}$ and/or $\mathcal{A}$ is an infinite set; it can only be set up as a working hypothesis which has got to be carefully tested, e.g. by exposing an extensive body of objects to a wide range of agents and analyzing the matrix of records. And whenever additional data are collected we must be ready to do it over again - possibly having to revise previous optimistic conclusions.

The concept may also be applied to deductive sciences like Mathematics and Mathematical Physics, the presence of specific objectivity then being an analytical question. In these fields our concept is closely related to that of Hermann Weyl* who - following the lead of Felix Klein - defines objectivity in terms of invariance under a chosen group of transformations. In Social Sciences and Humanities, where transformation groups rarely are readily available, their role may often be taken over by $[\mathcal{O}, \mathcal{A}, \mathcal{R}]$ - frameworks.

In passing we may notice that in both senses objectivity is a relative concept. However true a statement may be, it is never objective in an absolute sense, the objectivity always being conditioned by a frame of referance, whether this be a specified group of transformations or a specified $[\mathcal{O}, \mathcal{A}, \mathcal{R}]$- framework.

In the definition of specific objectivity the distinetion between objects and agents is a purely formal one, we might just as well have considered comparisons of some agents as defined through data obtained by applying a set of agents to a set of objects. Now it may very well happen

---

* [p. 77.]

- and the case is in fact of importance in some applications to social psychology - that we may compare objects, but not agents (or the reverse) with specific objectivity. This situation complicates matters and the present analysis will be restricted to the symmetrical case where both objects and agents may be mutually compared.

4. Motivating Parametrization.

Unquestionably the analysis of the concept of specific objectivity may be pushed further along the lines of the previous sections without any attempt at bridging the gap between quality and quantity. However, in view of our aspiration of finding a common ground for Humanities and Physics, and in view of the strong impact of the theory of continuous and differentiable functions upon the latter science we shall consider the possibility of imposing a quantitative viewpoint upon fields where the observations are qualitative.

As a point of departure we may turn to one of the fields within Physics where the basic observations are definitely qualitative, e.g. the study of radioactive substances. The determination of the "radioactivity" of a substance is based upon the observation of discrete critical events, e.g. scintillations on a screen, produced by the emission of α-particles. The intensity of this radiation is "measured" by dividing the number of scintillations by the length of the observational period, a figure that of course easily varies from one interval to another one, but by taking a sufficiently long period the intensity may be measued to any desired accuracy, thus being just as good as a measurement of a length, say.

Such experiments may be described in terms of $[\mathcal{O},\mathcal{A},\mathcal{R}]$-framework, $O_1, O_2, \ldots$ being a number of radioactive substances, $A_1, A_2, \ldots$ being observational periods, $R_{\nu i}$ the number of scintillations recorded. The theory behind the method of measuring the intensity tells that the probability of $R_{\nu i} = a$ is given by a Poisson law

$$(4.1) \qquad p\{R_{\nu i} = a\} = e^{-\xi_{\nu i}} \cdot \frac{\xi_{\nu i}^a}{a!}$$

in which the parameter of the "contact" (or situation) is the product of two parameters

$$(4.2) \qquad \xi_{\nu i} = \theta_\nu \sigma_i ,$$

θ_ν pertaining to the substance (in fact mass × its α-radioactivity), σ_i to the observational period (in fact its length).

In this context the main point is that although the observations are qualitative, parameters - in the present case real positive numbers - are assigned to both observations, object and agents, a structure that is, I think, typical for any physical measurement. And formally, at least, nothing seems to tell against using the same way of describing any other observational situation. How successful such attempts turn out to be is of course an empirical question, and how meaningful the results be is a matter that belongs under the particular field of knowledge and its philosophy - just as in the case of Physics. In consequence of this point of view we shall investigate the conditions implied in such parametrization, understood to be a tool for carrying out specifically objective comparisons.