

Chapter 4

On Scientific Data

It is therefore correct to say that the senses do not err – not because they always judge rightly, but because they do not judge at all.

Immanuel Kant: The Critique of Pure Reason

4.1 Measurement and Scales

Galilei, Kepler, and some other scholars around 1600 began making systematic measurements of the motion of bodies on earth and in the sky. This was a crucial aspect of the scientific revolution. At start such measurements was not consciously aimed at testing hypotheses, but the first results could be generalised to simple regularities, which could be treated as hypotheses and tested in new experiments in which kinematic quantities are measured.

A measurement is a comparison with a standard of some kind, and the result of repeated measurements is an ordering of objects in terms of how they compare to the standard. Another way of saying the same thing is to say that a particular type of measurement determines a scale, where each scale imposes an order among a set of objects. One says that the pike *weighs* 4 kg, that the A4-paper is 210 mm *wide*, or that Ann got an *A in English*. A statement about a measurement is an indication that a certain object is assigned a *measure* and a *quantity*. One can have various units for measurements of the same object. For example, length can be measured in meters, yards or feet. For it to be meaningful to say that one and the same property can be measured in various units, there must exist a well-defined transformation that converts one unit into another. The characteristics of these transformations can be used to give precise definitions of the different types of scales, as we shall soon see. First, however, I will give an informal characterization of four scale types.

A **nominal scale** (Latin *nomen* = name) is simply a classification of observations into categories. For such scales, no numerical comparison between values is possible, since no numbers can be assigned. One cannot say that some object has more or less of some property than another.

Example Consider the blood types A, B, AB, and O. One cannot say of one of these types that it has more or less of some property than any other. Blood type A has the A-factor but not the B-factor, and blood type B has the B-factor but not the

A-factor. The only thing we can say is that two people with different blood types have blood with different properties.

An **ordinal scale** is a ranking of objects with respect to some property where objects have more or less of the measured property. However, there is no universal standard as regards to the relative sizes of the increments in the scale: all that can be said is that one object has more or less of some property than another, not how much more or less of that property.

Example Many grading systems are ordinal scales, such as the European ECTS grading system A, B, C, D, E, FX and F. In this system one can say that a student who was given a C is more knowledgeable (according to the teacher) than a student who received a D, and that a student who received a B is more knowledgeable than both. Yet, we cannot say that the difference between B and C is the same as the difference between C and D.

An **interval scale** is a scale where the measurement data is ordered into a hierarchy with equal steps between increments. This means that one can make quantitative comparisons of measurements.

Example The most common scale for measuring temperature is Celsius. Using this scale, we can show that the differences between 30 °C and 20 °C, and 20 °C and 10 °C, are the same; since a difference of 1° is equal for any two consecutive points on this scale. However, we cannot make sense of ratios of measurements in this scale, as this would presume a non-trivially chosen zero-point representing the absence of the property being measured. 0 °C is not the absence of temperature, but rather a zero chosen for practical reasons, viz., the freezing point of water.

Finally, the **quotient scale** is an interval scale with a fixed zero-point representing complete absence of the measured property. This allows meaningful talk of quotients between measurements of different sizes.

Example Measurements of length. An object's length is measured using a variety of instruments: meter sticks, callipers, lasers, satellites, etc. We have an internationally accepted scale of measurement (1 m = the distance light travels in 1/299,792,458 s). If an object has a length of 0 m, then it has no length. Thus we can talk about a certain object being twice as long as another. If the zero were arbitrarily chosen, then the ratios between measurement values would change depending on where one chooses to place the zero-point. By comparison, we can hardly say that 20 °C is twice as warm as 10 °C, as becomes obvious when one expresses the same temperature in the Fahrenheit scale: 10 °C = 50 °F and 20 °C = 68 °F; hence this ratio of 2:1 in Celsius is a ratio of 6.8:5 in Fahrenheit.

Thus a scale is an assignment of numbers to objects (one can even assign numbers to categories in a nominal scale). Such an assignment is called a *measure*. With this terminology in hand, we can now give the following formal definitions of the various types of scale:

Def. 1: A measure f for a quantity q is a *quotient scale* if and only if for any other measure g of the same quantity there exists a number $k > 0$ such that $f = kg$

Def. 2: A measure f for a quantity q is an *interval scale* if and only if for any other measure g of the same quantity there exists a number $k > 0$ and a real number r such that $f = kg + r$

Def. 3: A measure f for a quantity q is an *ordinal scale* if and only if for any other measure g of the same quantity there exists a strictly increasing function j such that $g = j(f)$

Notice that the nominal scale is not given a formal definition above. There is perhaps reason to claim that the nominal scale is not actually a scale in the full sense of the word, but merely a classificatory scheme. However, the reason why it appeared first in the informal characterization above is that, in the transition from the nominal scale to the quotient scale, via the interval and ordinal scales, one sees an increase in the structuring of one's set of observations. I shall discuss the different types of classifications, especially classifications in the humanities and social sciences, in more detail in the next chapter.

In order to facilitate automatic recording and statistical processing one often assigns reference numerals to recorded observations, but it is a mistake to treat these labels as *numbers*, which are the normal reference of *numerals*.¹ This situation often arises with surveys. Many different types of surveys are comprised of various claims and the one taking the survey is required to mark one of the following alternative responses to each claim: *Strongly Agree, Somewhat Agree, No Opinion, Somewhat Disagree, Strongly Disagree*. This is a typical ordinal scale. It is commonplace that one assigns to these alternatives the following numerals: 5, 4, 3, 2 and 1. But the use of numerals to designate outcomes does not imply that one is using an interval or quotient scale; indeed, in this case the scale is plainly ordinal. In such situations, where numerals serve merely to rank or label, calculating means and standard deviations over them is meaningless. For it to be meaningful to calculate means and standard deviations over the results of opinion surveys the difference between, e.g., 'strongly agree' and 'somewhat agree' would have to be the same as the difference between 'no opinion' and 'somewhat disagree', but as standardly conceived this is just not the case.

4.2 Statistical Relations

It is quite often the case that we want to know if there is some connection between two variables. In an ideal experiment, one variable is varied in order to study the effects on another while all other variables are kept constant. But in most cases this is not possible. In many disciplines, such as medicine or climate research, we just do not know all of the relevant variables involved in a certain situation. Furthermore, even if we knew exactly which variables should be held constant, it is often

¹A numeral is a symbol, a linguistic sign that usually denotes an abstract object, a number. However, in some cases the numeral denotes a category of responses.

practically impossible to do so; how is one to construct an experiment to study the effect of a volcanic eruption on the world's average temperature? There may also be ethical restrictions, such as that one shouldn't manipulate people, animals or the environment without due regard to their wellbeing.

The conclusion is that actively controlled experiments are often infeasible, and that one must often be satisfied with passive observations instead. In order to determine whether two variables are in some way connected, one has to study many cases. The idea is that if one studies sufficiently many, randomly selected cases, then one has reason enough to believe that all other factors effecting the variables one wants to study are haphazardly distributed among the studied cases. Using statistical methods, one can then estimate the probability that an eventual connection between two studied variables is the result of the influence of uncontrolled variables; the larger the number of cases, the smaller the probability that the eventual connection does not obtain.

The first step is to conduct a sample test of a population and measure the two properties. One can then study the graph one gets if one places one of the variable properties on the x-axis and the other on the y-axis. Consider the following example:

A laboratory studied the relation between how long it takes for laboratory assistants to perform a certain procedure and the number of mistakes that were made. The results are presented in the following table.

Lab assistants	Time in minute	Number of mistakes
A	9	1
B	8	4
C	6	6
D	5	2
E	7	2
F	6	5
G	5	3
H	6	2
I	4	6
J	7	3

It is difficult to infer anything from this table, but let us look at the so-called scatter plot (Fig. 4.1).

In this diagram one can see a certain tendency that the longer a procedure takes, the fewer mistakes are made. A measure for this tendency is the so-called *coefficient of correlation* (also called *product-moment coefficient of correlation*) which is given by the formula

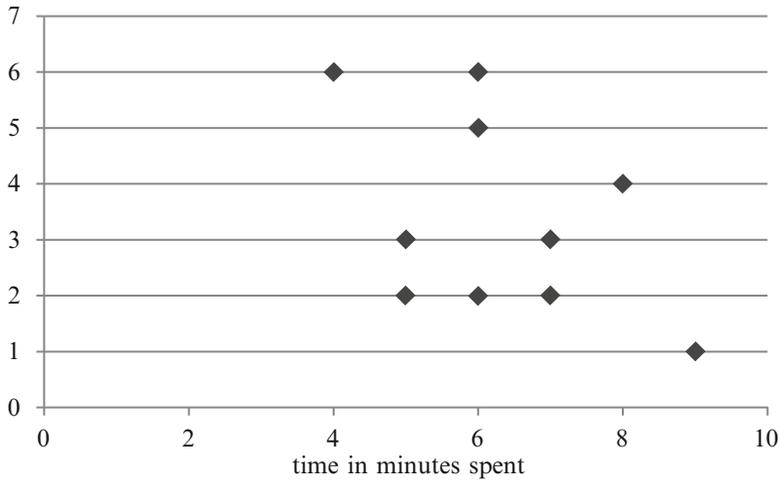


Fig. 4.1 Time spent versus number of mistakes in laboratory work

$$r_{xy} = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2) \cdot (n\sum y^2 - (\sum y)^2)}}$$

where n is the number of observations, and x and y are the two variables. In this case the correlation is -0.47 , which is a moderately negative connection. Of course, this was just a sample test, and without more information no definite conclusion can be drawn; it may very well happen that a larger sample test will show no correlation. But let us assume, for now, that this negative correlation is close to the actual: that it is the correlation that would result in an unlimited test. Given this assumption, our guess that the tendency for fewer mistakes to be made when time is increased is correct, since the correlation coefficient is negative.

Correlation coefficients range between $+1$ and -1 . If the correlation is $+1$ we have a fully positive correlation, and if it is -1 we have a fully negative correlation. The minus sign shows that an increase in the value of one variable tends to decrease the value of the other variable. The following are two scatter plots with calculated correlations for different sets of data (Figs. 4.2 and 4.3).

The second scatter plot shows a ‘thicker’ figure than the previous, which implies a lower correlation. Roughly speaking, the thinner and longer the area covered by accumulated points is, the higher the correlation. In the next graph we see a correlation between data from two secondary school physics tests, of which one is called the ‘practical test’, and the other ‘paper-and-pen’ (Fig. 4.4).

Here we see that the plot is quite narrow, which implies a high correlation. The calculation $r_{xy} = 0.81$ confirms this. (I once conducted these two tests myself in order to see if there were any significant differences between the practical and

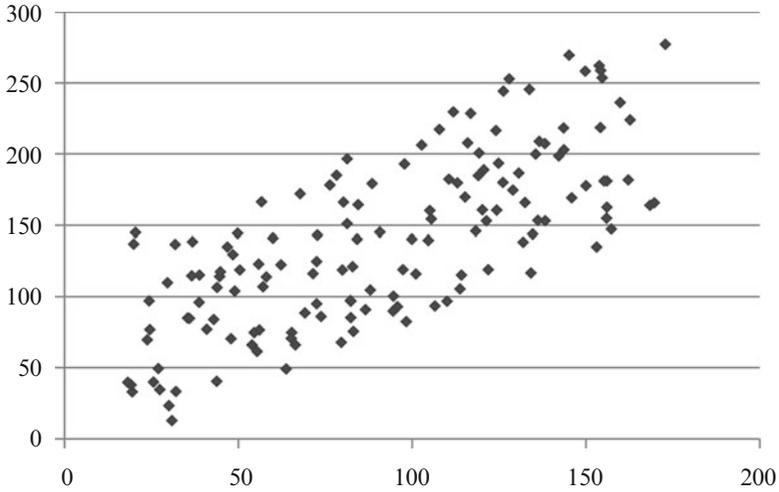


Fig. 4.2 $r_{xy} = 0.74$

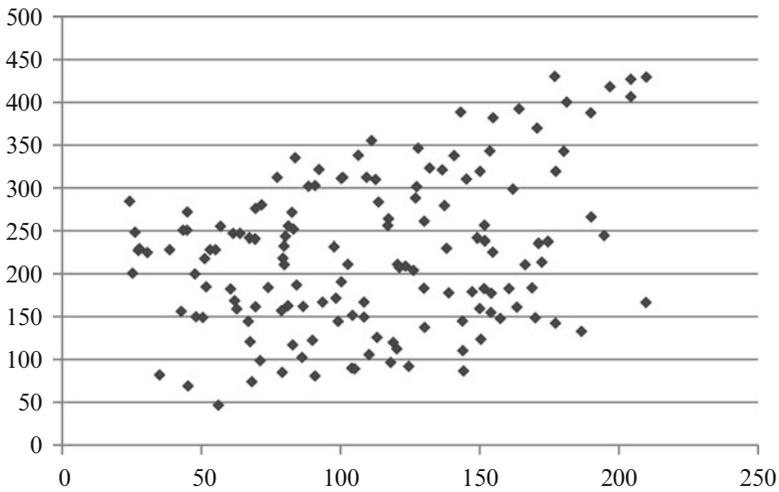


Fig. 4.3 $r_{xy} = 0.33$

theoretical capabilities of my physics students. The high degree of correlation shows that there was not.)

In the extreme case, when all points are located on a straight line, $r_{xy} = 1$ (or if negative, then $r_{xy} = -1$). The exception is when the line is vertical or horizontal, which means that the coefficient of correlation is 0.

However, the coefficient of correlation is not always useful as a measure of correlation. A moment's reflection tells us that this mathematical measure presupposes that both variables form interval or quotient scales, i.e. scales with

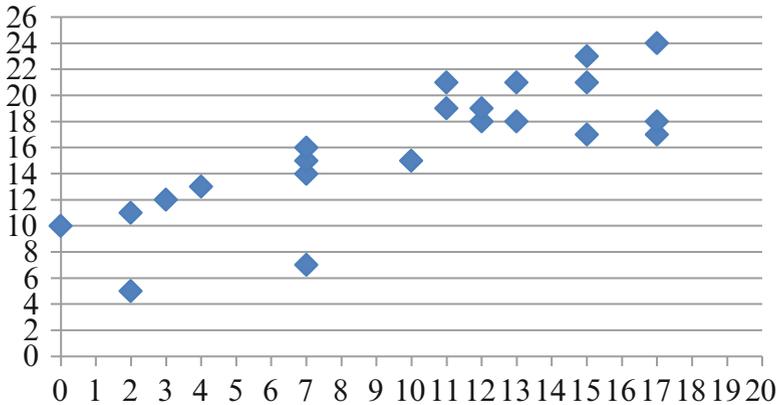


Fig. 4.4 The results of a paper-and-pen-test versus the results of an experimental test in a physics class

equidistant steps. If one of the scales is an ordinal scale, one must use other means for studying correlations.

In Chap. 7 we will discuss the conclusions one can draw from the existence of a correlation between two variables.

Exercises

For each of the following examples, state the type of scale the collected data will likely form.

1. An interview questionnaire containing the following question: How are you going to vote in the upcoming referendum?
2. A questionnaire containing the following question: How would you rate the healthcare system in Stockholm county: Very Good, Good, No Opinion, Bad, or Very Bad?
3. Classification of the colour of 20 particular species of forest mushrooms according to the international colour code.
4. pH measurements in the urine of a group of patients.
5. Measurements of temperature, in degrees Kelvin, of a number of stars in the milky way.

4.3 Data, Observation, Observational Statement

One of the central elements of science is the critical discussion of results; in any scientific paper all observations, conclusions, theories and hypotheses should be openly stated so that any competent reader is able to judge the extent to which the

conclusions are supported by the evidence. Of course, there are practical limits to what it is possible to make explicit for the reader to inspect. The reader will naturally have to take some things for granted, such as the raw data, which is often automatically recorded by a computer. While certain scientific journals require that one send in original recordings of one's measurements in order to check the reported data, sometimes this is impossible in principle as the raw data maybe comprised of a researcher's reports of their private observations, which are not open to independent scrutiny. Here is the limit of the science's objectivity; a space for both conscious cheating and unconscious mistakes. The following two examples from the history of science should illustrate this fact.

Example 1 The American psychologist Robert Rosenthal performed the following famous experiment. He told a number of students to conduct an experiment on mice in order to investigate the mice's learning ability. Sixty mice were distributed among 12 students. Half of the mice were said to be of a particularly 'gifted' kind, and half were supposedly of a 'slow' kind. The students were to place the mice in a labyrinth in which one route led to a food bowl. The students would then study the mice's ability to learn by measuring the time it took for them to find the food. The result was as expected: the group of mice that were more 'intelligent' was quicker to learn than the not so gifted mice. In addition, the students noted that the quick thinking mice refused to move 11 % of the time, whereas the less gifted refused to move 20 % of the time.

The point of this experiment was not to investigate the intelligence of the mice, but to discover something about the students! In fact, there was no intrinsic difference between either of group of mice! The students had merely let their expectations direct their observations of how long it took the mice to find food. It may strike one as strange that there could be a subjective element in the measuring of time, but such is the case according to Rosenthal. His general conclusion, based on this experiment, was that there is a measure of unconscious judgment, not a simple recording, in each observational statement. Note, however, that this concept of *judgment* is different from the *interpretation* of intentional entities like human actions, artefacts and texts, to be discussed in the next chapter.

Example 2 For a long time surgical transplantation was plagued by the body's rejection of foreign objects. Then a young American scientist claimed to have succeeded in controlling the mechanisms of the immune system to prevent rejection of foreign objects and tissue. He presented a number of white mice with large black fur patches on their backs, which he claimed to have been transplanted onto the mice. Other researchers then attempted to duplicate this experiment, but all such attempts failed. Not until an assistant to the American scientist happened to get black paint on his fingers upon returning the mice to their cages did people begin question the validity of his results. In fact, it turned out that the scientist has simply painted black spots on the backs of the mice!

The blatant fraud committed by the scientist in the story of the mice with black spots triggers the question of how common such deceit really is? No one knows for sure; however, in their book *Betrayers of the Truth*, Broad and Wade discuss a large number of fraudulent cases, and come to the pessimistic conclusion that deceit in science is widespread. There are two explanations for this. Firstly, since academic careers are extremely competitive, and largely depend on the significance and quantity of the results one has published, there is a great incentive to cheat. As they say in USA, 'Publish or perish'. Secondly, little academic merit is accrued by repeating previously conducted experiments; so few experiments are ever repeated. Not until there is reason to suspect fraudulent or sloppy work does anyone take the trouble to repeat an experiment. Hence according to Broad and Wade there is both powerful motive, and ample opportunity, for fraud in the sciences.

One should perhaps not over-generalize Broad and Wade's conclusions. As they themselves point out, the possibility and risk for fraud is different in different scientific disciplines. In some disciplines, a great mass of empirical research is conducted in rather routine fashion. This is what Kuhn calls *normal science* (see Chap. 6). An individual investigation often aims at testing a fairly well established hypothesis, theory or method in a new situation, e.g. the measurement of physical or chemical properties of new substances. For the individual researcher, an investigation can often be motivated by the prospect of greater academic accreditation, while the relevance to the scientific community is often restricted to the results strengthening an already successful theory. Thus, scientific results can often glossed as 'we have confirmed that our methods, or theories, hold in yet another case'. In these kinds of situations there is a significant risk of fraud, as the chance of detection is minimal. However, it would be too hasty to draw the conclusion that fraud and deception are common practice in the scientific community, or that experimental results are entirely untrustworthy. My own view is quite the opposite; namely, that fraud and deception belong to the exception and not the rule. After all, just because scientists have motive and plenty of opportunity to commit fraud, this does not mean that most will do so.

Returning to the central theme of this section, we shall now differentiate *observations* from *observational statements*. Observations are mental processes; and as such, they are private and exempt from external verification. Observational statements – which include the statements made by scientists and researchers – are verbal or written and so externally verifiable in principle. Observational statements are what scientific theories are tested against, not the observations reported by such statements. To make an observational statement – to report what one has observed – is a more or less conscious human act, to judge, and not a passive reflection of incoming stimuli. The senses, in so far as we regard them as distinct from the mind, merely transmit signals to the mind. This is what Kant implied by the citation at the beginning of this chapter.

4.4 On the Theory-Dependence of Observational Statements

Intuitively, it seems an easy task to distinguish between observation and theory. A theory is something abstract, often comprised of a number of formulas or general sentences. (Recall the definition of a theory presented in Chap. 3; a theory is a set of propositions whose inherent connection is explicitly stated). An observation, on the other hand, is a concrete action; something one does. However, actions must be described using words if they are to be of any value in a scientific community; hence, observations are described by observational statements.

The question then arises: is it just as easy to distinguish between theoretical statements and observational statements? The answer is no, as can be seen from the previous examples regarding how unconscious background beliefs can affect what is observed and reported even in a very simple task such as time measurements.

Many have drawn from such considerations the conclusion that all observation is theory-dependent. Their argument is as follows: Observation is an active process. Our sensory organs are exposed to physical stimuli, which spark a number of mental activities including processing and sorting of those stimuli in various ways. This processing is shaped by one's past experience, attentiveness, attitudes and expectations. The result is an observation expressed as an observational statement. Formulation of this statement requires a language containing a number of general terms by which one describes one's observations. This language and its categorizations are not theory-neutral; rather, they are mostly formulated in accordance with our preconceptions of the world. In short, our use of language is not theory-neutral in the strict sense of the word; neither the act of observation, nor the observational statements expressing those observations, are theory-neutral. N. R. Hanson argued this point in his (1965).

W. V. Quine disagrees² to a certain extent. He concedes that many of the observational statements in science are highly dependent on theory, but, he claims, there does exist some basic observational statements that everyone would agree to, independent of observers' culture, perspective or beliefs. A couple of obvious examples are 'the sky is now blue', 'there are seven apples in the basket', and 'the meter presently shows no deflection'. These statements describe here-and-now situations; and thus they are called *occasion sentences*. All competent speakers of the language in which these statements are expressed are able to immediately determine whether they are true or false simply by observing what is the case here and now. Thus according to Quine there is a solid foundation for our theorizing, though only some of the sentences we usually call observational statements are, in fact, theory-independent *occasion sentences*.

This opposing argument is based upon an important premise; namely, that two observers have access to the same language. This is certainly not always the case.

²This argument can be found in several of his writings, e.g. in the first chapter of his *Pursuit of Truth*, Harvard University Press, 1990.

There would be no issue if one could claim that, in principle, there is no problem in translating sentences of one language into sentences of another. Unfortunately, as any translator knows, this is far from the truth. Quine has discussed this question in connection with an investigation of the concept of *linguistic meaning*, and his conclusion is that there is a degree of indeterminacy, far from negligible, in any translation from one language to another.³ Yet, no translation is entirely arbitrary: for, if a translation is to be acceptable, it must translate occasion sentences in one language held true by persons talking that language into occasion sentences of the other language held true by persons talking the second language.

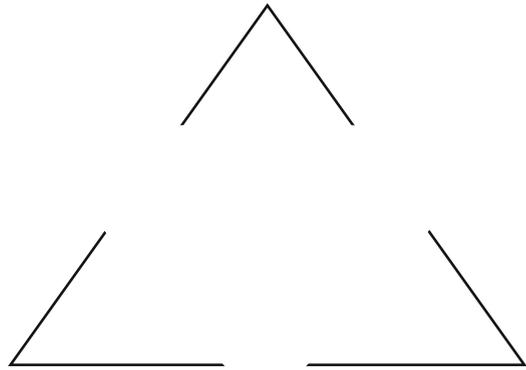
In my opinion, Quine is correct. This foundation of occasion sentences whose truth-value people agree upon is what is needed in order to claim that science has any measure of objectivity. It should be observed that if two or more persons agree to assent or to dissent to a sentence, they need not interpret it the same way. One can for example imagine two persons with different world-views and different languages seeing a horse in the meadow. One has adopted Plato's world-view, and says, in his language, something like 'Lo, the Horse-form is instantiated', the other thinks in terms of nature populated with individual objects belonging to different categories and says, in his language 'Lo, a horse.' (I have here given a translation of both expressions to English, of course.) If the speakers agree on what they are looking at, and are able to recognize outward expressions for assent and dissent, they can agree to each others utterances, without having any translation manual at hand, although their ontologies differ. So Quine's argument is that we may differ radically in our beliefs, theories etc., but it is nevertheless possible to find a common ground, viz., agreement on truth and falsity of occasion sentences.

Petra Stoerig provides good justification for this view in her interesting analysis⁴ of visual observation processes based on the presence of various types of brain damage and the associated loss of different aspects of one's vision. She describes three steps in the process of visual observation, or seeing. Firstly, we can distinguish a purely *phenomenal encounter* in which an *object* is not yet recognized, but where merely a number of qualities, such as light-dark contrasts and blotches of colour, are perceived. One is aware *that* one is seeing, but not *what* one is seeing. Some people who are born blind has been surgically treated for their blindness so that signals from their eyes enter the brain. But they have had great difficulties learning to interpret visual perceptions, like seeing an object, even where the object is familiar to them. During this period of learning, the patients experience the visual world as a unordered chaos of colours, forms, light and dark, and only after a long time do they learn how to order impressions into figures and objects (many patients fail to do so).

³ We may carefully observe that Quine claims that there is no fact of the matter which translation is correct, it is not merely that we are uncertain about which is correct.

⁴ Stoerig, P. (1996). Varieties of vision: From blind responses to conscious recognition. *Trends in Neurosciences*, 19(9), 401–406.

Fig. 4.5 We perceive this figure, without exception, to be a partially hidden triangle, that is, we construct a figure of the given visual elements

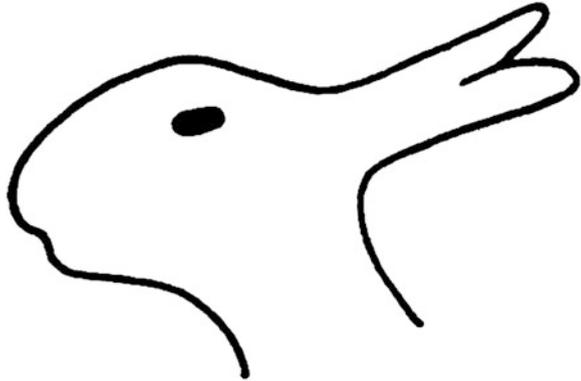


The second step in the process of seeing is to *perceive objects* as constituted by separate colours, shapes and contrasts. This results in distinguishing visual objects from their background. However, this perception is still not recognition (Fig. 4.5). There are people who perceive an object, and yet cannot say *what kind* of object it is, even though they have the linguistic comprehension required to do so. For example, one can know what a cat is, and be able to describe what a cat looks like, and yet still be unable to recognize that a certain visual perception of a physically extended object is of a cat. This inability to recognize objects can be found in dementia patients. Oliver Sachs has colourfully described such a patient in his book *The Man who Mistook his Wife for a Hat*. (It regards an elderly man who could no longer recognize his wife, even though he was still able to distinguish various objects in his surroundings).

The third, and last, step in this process is to *recognize* a given visual object as an object of a certain kind. Only when this step is completed does one have a visual observation in the full sense of the word. Hence, the act of perception/observation can be analysed, both conceptually and physically, as a three-step process. When we talk about visual perception – seeing – what we usually mean is the complete process culminating in recognition of an object as an object of a certain kind. Furthermore, it seems reasonable to assume that ‘perceiving’ by the other senses could also be analysed into stepwise processes, but Stoerig does not discuss this possibility.

The transition from step two to step three in the process of observation given above can be illustrated by the following example. Suppose the object of our perception is a figure that looks like a duck or a hare, depending on how you look at it (see Fig. 4.6). We agree *that* we observe a certain figure, but we do not agree on *what kind* of figure it is. According to the definition of an occasion sentence, in this situation the sentence ‘This figure depicts a duck.’, is no occasion sentence, if it is the case that some people see the figure as a hare. However, if we reformulate the observational statement so that it is neutral with respect to the kind of object being observed, then it will count as an occasion sentence. In the case of Fig. 4.6, the

Fig. 4.6 Duck or hare?



sentence ‘I see a contour that can be interpreted as either a duck or a hare’ is a theory-neutral occasion sentence, since everyone would agree with it. The point is that one can formulate observational statements without definite recognition of the object. However, it should be noted that what would elicit agreement from all humans irrespective of culture, and what would do so only within particular cultural contexts, is an empirical question. All that is required for the possibility of theory-neutral observations is the existence of *some* universal agreement among all humans: i.e. that all people have certain features of their cognitive equipment in common.

In medicine, we come across yet another problem regarding testing the effectiveness of medical treatments; namely, the placebo effect. If a patient believes that the medicine he is told to take is effective, then the patient’s expectations of being cured are increased. This changes the prognosis of the disease by raising the probability of a positive outcome regardless of whether the medicine has any pharmacological effect! In order to eliminate the placebo effect in the evaluation of treatments, one needs two groups whose members are ignorant as to which group they belong to: a test group and a control group. Furthermore, to minimize the effects of expectations on the data, such as those discussed previously in the timing of mice (expectations being, or resulting from, a sort of primitive theorising) one must make it so that even those administering the treatments are ignorant as to which group is which. This procedure – known as *double blind testing* – is standard for all trials of medical treatments. Since neither the participants nor the people administering the treatment know who is receiving the placebo and who is receiving the effective treatment, the members do not know which outcome to expect. Thus, the purely psychological effects of the treatment can be assumed to be the same in both groups and thus eliminated as a cause of different outcomes. Eventual differences between the test group and the control group can then be attributed to the administered medicine.

It follows from the above discussion that if we want to talk about objective scientific facts, we should restrict ourselves to only those statements that can be expressed by *occasion sentences* in order to limit the extent to which theory colours

our judgments. Similarly, we should be careful in saying that a hypothesis/theory is proven, as it is always possible to reinterpret the empirical data hitherto used to support that theory. This possibility is sometimes utilised during scientific revolutions, where one begins to 'see reality' from a wholly new perspective. A classic example is the Copernican revolution.

Before Copernicus, it was generally believed that the Earth lay motionless at the centre of the universe. The thought that the Earth could rotate around its axis was rejected by the following simple observations: If you throw an object straight up into the air, it will fall in the same place from where it was thrown. If the Earth rotated, then the object would fall some distance away from where it was thrown, since the Earth would have moved while the object was in the air. However, this is not the case, as this simple observation is supposed to show. What Galilei did, in his campaign against the traditional view that the Earth is fixed in the centre of the universe, was to interpret this experiment in a different way. He introduced the concept of relative motion and a different definition of the concept of force. Since the days of Aristotle, it was believed that force is required to maintain motion, and that motion ceases in the absence of force. Galilei claimed that motion with constant speed does not require a preserving force. From this it follows that when we throw an object into the air, the object follows the Earth's rotation even when it has left the hand of the thrower. The Earth and the thrown object rotate with the same speed; and so the object falls in the same place from which it was thrown, given that it was thrown vertically.

This example shows that we sometimes discard a theory or hypothesis without making any new observations. As in the previous example, we sometimes 'merely' interpret old observations in a new way. Such reinterpretation is nothing other than a change of one's auxiliary assumptions. In the case of the thrown object above, the essential point is what one means by *force*. Do forces cause motion, or do they cause changes in motion? In the argument against the Earth's rotation, the sentence 'a force is required to maintain motion' is an auxiliary assumption. This view is so natural that one is easily led to claim that one can 'see' that the Earth does not move when an object thrown upward falls precisely from where it was thrown. (The most profound change from Aristotelian to Galilean theory of motion is the conceptual change from the idea 'force causes motion' to the idea 'force causes change of motion'.)

Again, a theory or hypothesis can be strong, believable and very probable, and yet the cautious would refrain from saying that it is verified. Our knowledge is always provisional, more or less.

A sceptic would draw a different conclusion; namely, that we can never know anything with certainty and that all propositions can be doubted. The latter is correct, but one cannot doubt everything simultaneously. For, every time we doubt a proposition, that doubt is based upon another proposition (or more) that we take to be true. Our epistemological standpoint should thus be described using sentences of the form 'I do not believe in p because I hold that q is true, and q 's truth supports the belief that *not- p* is true'. It is certainly possible to replace p with any statement whatsoever, but this does not imply that one can doubt everything at the

same time. Even if there is no secure foundation for the empirical sciences, it is reasonable to assume that most of what we believe is true, while some of the propositions we take to be true are false.

4.5 Observations and History: On Source Criticism

Questions regarding the reliability and theory-dependence of observations are given a different character when applied to history. Students of history do not observe historical events, in the strict sense of the word. However, they do observe the various *remains* (e.g. bones, tools, and ruins) as well as *narrative sources* (e.g. letters, ledgers, and other writings). Only after interpreting and evaluating these historical sources do they discern the historical facts. In turn, these historical facts become the basis of various historical hypotheses.

History as an academic discipline underwent a profound change at the end of nineteenth century, first in Germany. The method of *source criticism* was introduced by Leopold von Ranke, (1795–1886), who was professor at Humboldt University in Berlin for a long time. Here is a brief exposition of the main principles.

It is fairly obvious that one cannot accept the contents of letters, reports, or minutes as historical facts without proper analysis. Consider, for example, a letter in which the author describes a historical event. One can pose a number of questions as to the truth of this letter. These questions are usually ordered into three groups; *proximity*, *bias* and *dependence*.

Proximity The first question regards the sources' proximity to the described event; did the author of the source see the event take place? How much time has passed between the acts of observation and their recording? We all know that the memory is notoriously untrustworthy; when describing an event that took place many years ago, it is doubtful that the resulting description will be accurate. Even worse are those descriptions written by people who did not observe the event themselves, but merely relate another person's observations. We are naturally led to two rules of thumb: (i) if possible, use sources that are contemporary with the events in question, and (ii) that are first-hand. Of course, these two principles can conflict. A contemporary source can be a second-hand source, and a first-hand source can fail to be contemporary. However, experience has shown that if these two principles conflict, then the contemporary second-hand source is usually more reliable than the later primary source.

Bias The next point in the critical evaluation of sources regards the author's intentions with respect to what he is writing. It is possible that the purpose of the text is to impress a particular view upon the reader, rather than give an objective description of the events concerned. A classic example from Sweden's history is Erik Dahlbergh's account of the Swedish army's crossing of 'Det Store Bælt' (The Big Strait) in Denmark, in the year 1658. (Erik Dahlbergh (1625–1703) was a

Swedish soldier, finally field marshal). That year the winter was very cold in Scandinavia, so cold that the Big Straight froze and made it possible for the Swedish army to march towards Copenhagen and conquer Denmark. In his so-called ‘diary’,⁵ Dahlbergh describes the event as if he is one of the central actors of this brave military action. This ‘diary’ was long considered a reliable document. In modern times, however, this account has been proven false. Dahlbergh’s ‘diary’ is a retrospective construction, written long after the events transpired. It is now seen as an important part of Dahlbergh’s campaign to promote his own views and boost his reputation. Critical analyses of this kind are based on an author’s personal bias and aim to pinpoint the intentions for writing his or her work. Thus any content that expresses, or is in line with, the author’s personal bias is automatically discarded from a historical study, unless that bias is the subject of particular investigation.

Dependence A third form of source criticism regards the possibility that the author was instructed, threatened, or otherwise forced into producing the source. Note that this includes situations in which the author is unconsciously coerced, as in the case of societal conventions. For instance, modern historians do not believe accounts of witchcraft, coitus with the devil, or religious miracles, as these phenomena conflict with general scientific knowledge and can be satisfactorily explained in psychological terms. Another example is found in the Marxist repertoire. Contained in the Marxist perspective is the view that a person’s reasons for his actions are secondary effects of objective, material and societal processes. Such a critique is based on the notion that actions of agents are determined by objective historical processes. Thus according to the Marxist perspective, the reasons and arguments an agent gives for her actions or points of view are not relevant, no matter whether she is honest or not. Motives are not historical causes. (Opponents of Marxism were therefore, by definition, deemed representatives of the bourgeoisie, and therefore unable to see the truth. This is a typical *argumentum ad hominem!*) This source criticism thus builds upon the author’s dependence on various norms, values, assumptions, historical mechanisms, etc. Furthermore, it is this dependence that makes us doubt the truth of the author’s statements.

It is apparent that the utilization of these principles of source criticism in particular cases leaves plenty of room for the historian’s personal judgment. Can one then truly say that what is left after a critical source evaluation of some historical material is indeed historical fact? It depends on what one means by ‘fact’. If we equate fact with ‘considered to be a fact’, then it is so by definition. However, if by ‘fact’ we mean the content of true statements, then it is not the case.

⁵ An analysis of the paper in the diary shows that it could not have been written earlier than the 1670s, though it was probably written much later. The book draws on actual diaries and almanacs that Dahlbergh authored, but the contents of the so-called diary do not agree with other descriptions of the crossing. See Englund, P. (2000). *Den öövertinnerlige: om den svenska stormaktstiden och en man i dess mitt*. Atlantis, p. 565.

This is in no way unique to the historical sciences. In all empirical sciences one is painfully aware that what is presently considered secure fact can later show itself to be false. Truth exists, but we do not have any guarantees of finding it.

Whether history treats its material as a source or as relics is not a trivial matter. An informative text can be entirely unreliable as regards the events described therein, but it may still be useful as an expression of the author's attitudes, values and beliefs. An example from Thucydides' *The Peloponnesian War* should illustrate this point. Pericles, eulogizing over the fallen soldiers, describes the norms and habits of Athenian citizens:

Nor are these the only points in which our city is worthy of admiration. We cultivate refinement without extravagance and knowledge without effeminacy; wealth we employ more for use than for show, and place the real disgrace of poverty not in owning to the fact but in declining the struggle against it. Our public men have, besides politics, their private affairs to attend to, and our ordinary citizens, though occupied with the pursuits of industry, are still fair judges of public matters; for, unlike any other nation, regarding him who takes no part in these duties not as unambitious but as useless, we Athenians are able to judge at all events if we cannot originate, and, instead of looking on discussion as a stumbling-block in the way of action, we think it an indispensable preliminary to any wise action at all. Again, in our enterprises we present the singular spectacle of daring and deliberation, each carried to its highest point, and both united in the same persons; although usually decision is the fruit of ignorance, hesitation of reflection. But the palm of courage will surely be adjudged most justly to those, who best know the difference between hardship and pleasure and yet are never tempted to shrink from danger.⁶

If we ask what the text says about Thucydides' view on Athenian life, instead of questioning the accuracy of what Pericles said, or even if it was Pericles who said it, we are not treating the text as an informative source, but rather as an ancient relic that tells us something about how the Athenians saw themselves. Such a treatment can be just as interesting as asking what Pericles actually said.

4.6 Summary

All empirical sciences are tested against data in the form of observation reports. For such testing to be meaningful, observation reports must be independent of the theory to be tested. This can be achieved, although observation reports, in the usual sense of the word, often are theory-laden to some extent. It is crucial for the objectivity and trustworthiness of any science that data are acceptable as such both by friends and foes of the particular theory, which is tested by those data.

⁶ Thucydides: The History of the Peloponnesian War, book 2, chapter 6, (trans: Crawley, R.). <http://classics.mit.edu/Thucydides/pelopwar.2.second.html>

Further Reading

- Ellis, B. (1968). *Basic concepts of measurement*. Cambridge: Cambridge University Press.
- Quine, W. V. O. (1990). *Pursuit of truth*. Cambridge, MA: Harvard University Press.
- Quine, W. V. O. (1995). *From stimulus to science*. Cambridge, MA: Harvard University Press.
- Suppes, P., & Zinnes, J. L. (1963). Basic measurement theory. In R. D. Luce, R. R. Bush, & E. H. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 3–76). New York: Wiley.
- Tosh, J., & Lang, S. (2006). *The pursuit of history: Aims, methods and new directions in the study of modern history* (4th ed.). Harlow: Pearson Education.