The Art of Doing Science

2007-08-14

Department of Philosophy and the History of Technology, KTH (www.infra.kth.se/phil)
soh@kth.se

Contents

Preface

The philosophy of science is largely devoted to great discoveries that have changed the ways in which we see our world. It is an important task to investigate the intellectual and social processes that have changed our ways of thinking about the world we live in and about ourselves. Therefore, scientific revolutions, as well as the special issues concerning quantum mechanics, relativity theory, and evolution, take up a large part of many texts in the philosophy of science.

But the scientific revolutions and the great theories just mentioned are a relatively small part of science. Very few of us have a chance to contribute to their development. Many more have a profession where we need to make some sort of scientific judgement. This applies not only scientific research. Scientific method is useful and often necessary in also many other types of professional activities. As citizens we all have reasons to take a stand in social issues where scientific arguments often have an important role. This text is devoted to clarifying what it means to have a scientific attitude and to use a scientific method, both when performing research or investigations and when drawing conclusions from the research and investigations performed by others.

A discussion on scientific method can be either descriptive or normative. The descriptive approach is typical for the sociology of science, in which studies are made of how science in conducted in practice. The normative approach is more typical of the philosophy of science, the discipline that discusses methods and approaches that are characteristic of good science. In this text, the approach is normative. This means that we will be concerned with understanding what it means to conduct science in a good way. However, the approach taken here differs in several ways from how the philosophy of science is usually presented.

Research can be conducted in order to satisfy our curiosity about the world around us. It can also be performed in order to provide us with guidance for action. Most of the research conducted for instance in medicine and in the technological sciences belongs to the latter category. The philosophy of science has mostly had a strong focus on so-called pure science. In this text the applied sciences will be treated on an equal footing. The text is primarily written for students in technology, and therefore gives more attention to research in technological science than what is common in the philosophy of science.

In university education there is very little contact between the philosophy of science (theory of science) and practical scientific work. Between them is a large area of knowledge, that can be called general methodology in science. It includes for instance general principles for the planning of experiments and measurements, general principles for computer simulations, the use of historical sources, and methods to deal with various types of sources of error that are common to many different areas of science. This text has a strong emphasis on these general methodological issues that have seldom been treated systematically.

On occasions, the philosophy of science has been described as some sort of a superior judge over the so-called special sciences. This is an over-estimate of the abilities and possibilities of the philosophy of science. It would be better to describe the philosophy of science as an auxiliary science, studying the fundamental methodological problems that are common to the different sciences.

This text is also available in a Swedish version. It will be subject to further changes and amendments. Comments and proposals are very welcome.

Stockholm, August 2007-08-14
Sven Ove Hansson

1 What knowledge do we need?

Science is a means to acquire knowledge. But there are many types and forms of knowledge. In this chapter we will attempt to specify what types of knowledge science can provide us with. We will also begin our investigation of the sources of scientific knowledge.

1.1 Episteme and techne

The circus artist who balances a plate on a stick does not have to be able to explain what forces act on the plate. It is sufficient if he can keep the plate in place. For the physicist the opposite applies. The physicist should be able to explain what forces combine to keep the plate in place, but she need not be able to master them in practice.

We can therefore distinguish between two types of knowledge. One of these can be called knowledge of facts. It consists in being able to give an account of how things are. You prove your knowledge of facts by making the right type of statements. The other type of knowledge can be called action knowledge. You prove your action knowledge by performing actions that lead to a desired result. In some cases, like that of the circus artist, we consider it sufficient to have an intuitive, non-verbal form of action knowledge. In other cases, like those of the engineer and the physician, we usually demand that action knowledge should be expressible in rules or principles that others can learn to apply.

The distinction between knowledge of fact and action knowledge is very old. The ancient Greeks, including Aristotle (384 – 322 B.C.) distinguished between knowledge of facts that they called *episteme*, and action knowledge that they called *techne*. The latter was described by Aristotle as "a productive state that is truly reasoned".[1]

1.2 The notion of science

Far from all knowledge can be described as scientific. The distinction between scientific and non-scientific knowledge is in part conventional, as we can see from the difference in meaning between the English word "science" the corresponding word "Wissenschaft" in German with its close analogues in some other Germanic languages.

Originally, the English word "science" had a very wide meaning. It was used to denote almost anything that you had to learn in order to master it: everything from scholarly learning to sewing and horse riding. But in the 17th and 18th century the meaning of "science" was restricted to systematic knowledge. The word was for instance used about the knowledge you need to make a living in a particular practical trade. In the 19th century the meaning of "science" was further restricted, and essentially meant what we would today call natural science.[2] Today, the word "science" is still primarily used about the natural sciences and other fields of research that are considered to be similar to them. Hence, political economy and sociology are counted as sciences, whereas literature and history are usually not. In several academic disciplines there has been a movement to make the discipline accepted as a science.

[1] Nicomachean Ethics 6:4 , p. 208 in JAK Thomson's translation, Penguin books1976.
[2] Edwin Layton, "American Ideologies of Science and Engineering", *Technology and Culture* 17:688-701, 1976.

This applies for instance to social anthropology, that is often counted as a science although it is in many respects closer to the humanities.[3]

The closest synonym to "science" in German, "Wissenschaft", also originally meant knowledge. However, "Wissenschaft" (and the corresponding words in other Germanic languages) has a much broader meaning than "science". It is used to include all the academic specialties, including the humanities. The term "Wissenschaft" has the advantage of giving a more adequate delimitation of the type of systematic knowledge that the present text refers to. It does not exclude academic or otherwise systematized knowledge that is excluded from the "sciences" due to linguistic conventions. In other words, our subject matter is science in a broad sense ("Wissenschaft") that includes disciplines such as musicology, medieval literature and ancient history. What is important is of course not the name, but the fact that there is a community of knowledge disciplines that all strive to obtain reliable general knowledge and all respect the other disciplines in their respective areas of speciality.

This community of knowledge disciplines covers, in combination, a very large area of knowledge. Science in a broad sense seeks knowledge about nature (natural science), about ourselves (psychology and medicine) about our societies (social science and history), about our own physical constructions (technological science), and about our own thought constructions (language, literature, mathematics and philosophy).

In spite of this universality there are some conventional limits. Areas of knowledge that have not found their place at universities are usually not regarded as sciences or "Wissenschaften". Hence numismatics (the study of coins) is counted as an academic discipline, but not philately (the study of stamps). In a more principled discussion (like the one to be conducted here) we should pay attention to the substance and not to these conventional limitations. It is not the area of study but the methodology and the type of knowledge that should determine whether or not a field of knowledge should be regarded as scientific.

1.3 "Pure" and "applied" science

The philosophy of science has usually focused on episteme, i.e. "pure" science that aims at knowledge of facts. A large part of the science that is conducted in practice is however devoted to action knowledge, and therefore aims at techne rather than episteme. In other words its purpose is to make us able to achieve practical aims. Two of the largest areas of research, medicine and technological science, are basically devoted to action knowledge. This also applies to large parts of the social sciences.

In practice it is very difficult to find a sharp limit between the science that aims at knowledge of facts and that which aims at action knowledge. Chemistry is usually regarded as a "curiosity-driven" science, but there are good reasons to see chemistry as close to the technological sciences. About two thirds of all chemists devote their working hours to synthesizing new chemical substances.[4] In chemistry it is difficult to draw a clear line between pure and applied science. This also applies more and more to biochemistry.

[3] Merrilee H Salmon, "The rise of social anthropology", pp 679-684 in Thomas Baldwin (ed.) *The Cambridge History of Philosophy 1870-1945*, Cambridge University Press 2003.
[4] Joachim Schummer, "Challenging Standard Distinctions Between Science and Technology: The Case of Preparative Chemistry", *Hyle* 3: 81-94, 1997.

Instead of dividing scientific activities into two categories, pure and applied science, it would be more accurate to distinguish between two types of values that a scientific study can have. It can have greater or smaller value for our strivings to explain the world, i.e. contribute more or less to the "pure" or internal strivings of science. It can also have more or less value for our decisions on how to act in order to achieve different practical goals. These two types of values are not in contradiction to each other. Sometimes one and the same investigation can give valuable contributions of both types. Explanatory (intra-scientific) fruitfulness and practical (extra-scientific) usefulness are two independent properties that one and the same scientific investigation can have to different degrees.[5]

1.4 Universal versus special knowledge

When we look for facts in our everyday lives, we sometimes ask questions like "What is it like?". Sometimes we also ask questions like "Why is it like this?". Then we are looking for knowledge that connects and explains the individual facts. Both these types of questions are also asked in science. Sometimes we are interested in special events or special phenomena, sometimes in more general connections. In the latter case, special phenomena and events will be of interest only as examples of the more general phenomena that we are looking for.

Different sciences divide their interest quite differently between the special and the general. Physics is probably the clearest example of a science that focuses on universal issues. It is characteristic of physics to search for very general principles and laws, for instance concerning the properties that are common to all matter.

Biology has an interest both in the special and in the general. Describing a specific species carefully, or perhaps even discovering a new species, always has a value in biology. At the same time biology has a strong interest in general issues, such as the ecological and genetic mechanisms that apply to very different species and environments.

Most sciences share their attention between universal and special knowledge in about the same way as biology. Chemists look for knowledge both about individual chemical substances and about principles that apply to chemical substances and reactions in general. Historians are interested both in individual events and personalities and in larger trends in social development. Anthropologists are interested both in individual cultures and in more general principles that are common to different human societies.

Many sciences have developed from a previous focus on special knowledge to a growing interest in more universal issues. Perhaps this can be most clearly seen in biology. The biologist Peter Medawar expressed this by saying that "Biology before Darwin was almost all facts".[6] Evolutionary biology, molecular biology, and ecology have all provided biology with new tools that bind together previously isolated facts to a more unified body of biological knowledge.

There have been similar developments in many other sciences. This trend towards more universal knowledge has often been neglected in discussions of the disciplinary development of science. It is certainly true that science is divided between an increasing number of specialties and sub-specialties, but at the same time the connections and dependences between

[5] Sven Ove Hansson, "Praxis Relevance in Science", *Foundations of Science*, 13:139-154, 2007.
[6] Peter Medawar, *Pluto's Republic*, p. 29.

all these scientific areas increase continuously. Science becomes at the same time more specialized and more unified through overarching knowledge.

The increased specialization in science depends on the much larger size of the mass of scientific knowledge. In biology, the number of described species has increased, and so has the detailed knowledge about the species that were known since long. Therefore it is no longer possible to be an expert in as large parts of biology as you could before. The situation is similar in other sciences.

In summary, science is based on a combination of knowledge on different levels of generality. These levels support each other, but cannot replace each other. General knowledge has often grown out of a large mass of special knowledge, such as when knowledge of evolutionary processes has grown out of knowledge about individual species. At the same time evolution provides us with a conceptual apparatus that simplifies and amplifies our understanding of many individual phenomena.

1.5 Action knowledge

Research aiming at action knowledge can use science in essentially two ways. We can study these methods in both their medical and their technological applications.

The physician's art was for a long time dominated by the conclusions that physicians drew from rather speculative theories and from their own experiences from treating different patients. Medicine was taught at universities since the end of the 13th century, but it was not until the 19th century that university teachers in medical subjects tried to make medicine a science. There were two different conceptions of how this could be achieved, and on occasions there were deep conflicts between proponents of these two conceptions.

Joseph Dietl

Still in the middle of the 19th century the general view among physicians was that pneumonia depends on an imbalance between the bodily fluids. The most commonly recommended treatment was blood-letting. Some physicians favoured instead a somewhat less drastic means of bringing the bodily fluids into balance, namely by giving patients an emetic.

In 1849 the Austrian physician Joseph Dietl (1804 – 1878) reported an investigation in which he compared three groups of pneumonia patients. One group had received blood-letting, the second had received an emetic, and the third had received general care but no specific treatment. Mortality among those who had received blood letting was 20.4%, among those who had received an emetic 20.7% and among those who had not received any specific treatment only 7.4%.[7]

How could the clinical impressions have given rise to such misleading conclusions? One important explanation is wishful thinking. Most sick people feel sometimes somewhat better, sometimes somewhat worse. If the patient happened to feel somewhat better after blood-letting, this was noticed, whereas less weight was put on occasions when the treatment was not followed by an improvement.

Dietl's message was at first very negatively received, and Dietl himself lost his work. But in a longer perspective he was successful. In the 1870´s the major medical textbooks advised against blood-letting of pneumonia patients.

Dietl's study was one of the very few treatment studies that were performed in the 19th century. It was only about 100 years later that clinical studies of treatment effects began to be performed on a large scale and to determine the development of medicine in general.

[7] Dietl, J. (1849). *Der Aderlass in der Lungenentzündung*. Wien, Kaulfuss Witwe, Prandel & Comp.

One of these approaches can be called “mechanism research”. It consists in laboratory studies in which the causes and mechanisms of different diseases are investigated. On the basis of such investigations conclusions are drawn about how different diseases should best be treated. The other approach can be called “treatment research” or clinical research. It consists primarily in clinical experiments, in which patients with the same disease are treated differently in order to determine which treatment method yields the best result. (See the fact box.)

In modern medicine both mechanistic research and treatment research have central roles, but their roles are different. The innovative force is mostly found in mechanistic research. It is by finding out the mechanisms behind different diseases that new concepts for treatment can be developed. (Chemical synthesis and technological inventions have of course also important roles in this process.) But when it comes to judging which methods should be used, the final decision cannot come from mechanistic research. The human body is such a complicated system that we can never determine from mechanistic knowledge whether a treatment will have the intended effect. Here clinical trials have the last word. It does not matter how convincing the biochemical or physiological arguments are for a treatment method, if its expected effects cannot be shown in clinical trials then it has to be rejected.

A similar development has taken place in the technological sciences. When modern technological science emerged in the 19th century, two distinct strategies were used. First, results from the natural sciences were applied in order to determine beforehand how different constructions would work. With formulas from mechanics the movements of moving parts of machines could be predicted, and formulas from the theory of electricity provided guidance in the construction of electrical machinery. This corresponds to the use of mechanistic research in medicine.

Secondly, scientific studies were performed on individual technological constructions. This was often done by constructing a machine prototype on which careful measurements were performed in order to determine how its different parts should be arranged in order to obtain the desired result.[8] Such investigations are still performed to a large extent in innovation work; among the most well-known examples are crash tests on cars and wind tunnel tests of new airplanes. This type of investigation corresponds to the treatment studies performed in medical research.

Summing this up, in practice-oriented research such as medicine and technology, science is used in two ways. First, scientific theories are used to construct solutions to practical problems (mechanistic research in medicine, the use of physical theories in technology). Secondly, scientific methods are used to study how different solutions of the practical problem work in practice (treatment research, tests of technological constructions). These two approaches answer different types of questions, and neither of them can replace the other. It is by combining them that the most powerful results can be obtained.

[8] Walter Kaiser, “Die Entwicklung der Elektrotechnik in ihrer Wechselwirkung mit der Physik”, pp. 71-120 in Lothar Schäfer and Elisabeth Ströker, *Naturauffassungen in Philosophie, Wissenschaft, Technik, Band III; Aufklärung und späte Neuzeit*, Verlag Karl Alber Freiburg, München 1995. W Faulkner “Conceptualizing Knowledge Used in Innovation: A Second Look at the Science-Technology Distinction and Industrial Innovation”, *Science, Technology and Human Values* 19(4):425-458, 1994.

1.6 Intersubjectivity and objectivity

Scientific activities aim at obtaining knowledge that is valid for all humans. When discussing religions or world views you may sometimes be satisfied by saying “That may be true for you, but this is true for me”. This is not a way in which we can reason in science. In science we seek common (although provisional) truths. Science provides us with a common world view, or perhaps more precisely: common parts of our world views. In order to achieve this, scientific knowledge should be intersubjective, i.e. common to all humans.

But the purpose of science is not just to achieve any common opinion. Science does not for instance strive to achieve commonly accepted prejudices. Instead, its aim is to obtain knowledge about how things really are. This means that we are striving for objective knowledge.

This requirement of objectivity applies both to episteme and techne, both to knowledge of facts and action knowledge. When a historian makes an account of how World War II started, or a chemist explains a particular chemical reaction, we expect their statements to refer to and describe reality. They should not just report personal suggestions or ideas. Our requirements are similar when a medical researcher tells us which treatments are best against some disease, or when a civil engineer describes how a bridge should be built in order not to fall down. We are not satisfied with just any personal opinion that a doctor may have about treatments or with the engineer’s intuitive conviction that the bridge will hold. We expect, if they claim to base their statements on science, that the recommended treatment should actually be efficient and that the bridge should not fall down.

Objectivity can be difficult to achieve, and sometimes it can be impossible to achieve it more than partially. However, it does not follow from this that we should refrain from attempting to achieve as much objectivity as possible. If something worth achieving cannot be fully achieved, we should still try to achieve it as far as possible. It is for instance not possible to write a fully objective book about the history of the Vietnam war. But it does not follow from this that a historian would be justified in giving up the ideal of objectivity, and stop worrying about whether what she writes about the Vietnam war gives a good or bad picture of what actually happened.

The lack of this simple insight, that objectivity can be strived for even if it cannot be fully achieved, has sometimes led to the idea that we do not have to strive for objectivity in science. Nothing can be more erroneous. Science is devoted to finding out how things really are.

We can also express the demands of objectivity and intersubjectivity in terms of three presuppositions on which science is based. Together they imply that objective and intersubjective knowledge is possible. The three conditions are:

1. There is a real world that is independent of our senses.
2. This real world is common to all of us.
3. With combined forces we can achieve, or at least approach, knowledge about this real world that is common to us all.

The first two of these conditions concerns what there is (ontology). The third concerns what we can know (epistemology). All these three preconditions are taken for self-evident in everyday life. It is almost only in philosophy that they are seriously reflected upon.

Objectivity and intersubjectivity are related, but they do not coincide. Intersubjectivity is possible without objectivity. We can have ideas that are intersubjective without being objective, for instance we can all have the same prejudices or be subject to the same misunderstandings. On the other hand, objectivity includes intersubjectivity. The reason for this is that the demand of objectivity is the same for all persons, since it refers to a single reality that we try to describe as accurately as possible.

However, it does not follow from this that intersubjectivity is uninteresting. To the contrary, intersubjectivity is important since it is easier to deal with than objectivity. We have no direct access to physical reality (only to the impressions that it leaves in our senses). Therefore we have no simple and direct means of checking if a statement is objective. It is much easier to determine if it is intersubjective, i.e. if anyone could arrive at it and it holds against critical testing by anyone. In science we require that the conclusions one draws should resist the critical assessment of others. This is a way to express the requirement that scientific knowledge should be intersubjective.

1.7 The danger of belief in authorities

Since science strives for intersubjectivity it is incompatible with the belief that certain persons have a special ability to obtain knowledge, such that others only have to accept what they tell them. In other words, science is non-authoritarian, often anti-authoritarian. The value of an argument cannot depend on who puts it forward. Science and democracy share a strong commitment to the fundamental equality between human subjects. In order to be workable, both science and democracy presuppose a rational public discussion where arguments are tried against each other. Democracy and modern science also have common philosophical roots in the enlightenment tradition and its protests against old authorities.

Nevertheless we often find authoritarian patterns in science. A long experience shows that such patterns are in the long run dangerous for the development of science. Freudian psychoanalysis and dogmatic Marxism are examples of tendencies that have often led scientists into a dead end. The natural sciences also have examples of how individual researchers have achieved such a strong position that they have prevented new ideas from developing.

In practice it can also be difficult to see that science is intersubjective. Due to specialization nobody can walk into any scientific institution and check what the experts are doing. Neither can we build our own radio telescopes or particle accelerators. But this does not mean that there is no intersubjectivity. The requirement is not that anybody should in practice be able to check the statements made by others. Instead it is that we should be able to do so if we spend the time needed to learn enough about the issues.

In this respect our attitudes to science can be compared to our attitudes to practical trades. If I want to have my clock repaired, I do not do this myself but go to a watchmaker. I do not believe that this person has any type of special abilities that I would not be able to acquire. In fact I believe that if I took the time needed I would be able to learn this trade.

This is the common attitude that we have to practical trades. But many of us have a quite different attitude to science. They take it for granted that science is "above their level", so that only a special type of persons are able to understand it. This is not so. We should ask the members of the science trades for advice for the same reasons that we ask the members of practical trades for advice: not because we are unable to learn these trades but because it would be impractical to take all the time needed to learn them.

There is a risk that a dangerously uncritical attitude to scientific authorities can develop. There is also a risk of the opposite mistake, namely a lack of respect for the insights that long experience and extended studies give rise to. I can of course choose not to follow the advice of the watchmaker on how to mend my clock, but it would be unwise to take such a standpoint without first carefully considering the arguments and experiences that the advice is based on. Similarly it would be unwise to reject the opinion of scientific specialists in their fields of expertise without first very carefully studying what they base their opinions on.

1.8 Starting out from the best available knowledge

Some worldviews offer absolute and certain knowledge. Science is not one of these worldviews. This is important to realize, since many unproductive discussions have had their origin in the mistaken view that science claims to provide absolute truths about the physical world, truths that are as certain as simple arithmetical facts like 1 + 1 = 2. The claims of science are smaller, but still important enough. Science tries to offer the best available knowledge today, and in addition methods with which we can successively improve our knowledge.

Uncertainty and fallibility are not specific for science. All human knowledge is fallible and subject to reappraisal. In everyday life we take many things for granted that we cannot be absolutely certain of. If it turns out that we were wrong, we change our opinions. The same procedure applies in science.

It is self-evident but still in need of being said: That knowledge is fallible does not mean that anything goes. Hence, many elements of modern evolutionary biology may be in need of revision in the light of future research, but this is no reason to accept for instance creationism or its variant "intelligent design". (These are teachings that have their origin in religious traditions, not in attempts to understand nature with the help of modern research.) We must always try to find the best available knowledge. It is in most cases less certain than 1 + 1 = 2, but that is no reason to replace it by statements for which we have even less secure foundations.

Epistemologist Isaac Levi has pointed out that we must distinguish between "certain" and "incorrigible" opinions.[9] The chemist's opinion that gold is an element is certain knowledge in the sense that there is not now any doubt about it. It belongs to what a chemist, given what we now know, has no reason to doubt. On the other hand this opinion is at least in principle not incorrigible, since we can (hypothetically) think of empirical evidence that would lead us to reject it.

[9] Isaac Levi, I. *The Fixation of Belief and Its Undoing*, Cambridge University Press, Cambridge, Mass. 1991.

It is possible and even probable that in the future we will be forced to revise many opinions that we have no reason to put into doubt today. There is no reason to believe that we will have to give up the opinion that gold is a chemical element, but probably revisions will be required of some of the millions of other opinions that we take for given today to almost the same degree. One of the best historical examples of this is the reappraisal of previous opinions on space and time that relativity theory gave rise to. This applies for instance to Euclid's parallel axiom. It says that if we have a straight line and a point beside it on a flat surface, then there is exactly one line on the surface that goes through the point and never crosses the first mentioned line. Philosopher Immanuel Kant was convinced that the parallel axiom is true with necessity. Today we know that it does not even apply to the geometry of the physical world.

That the best available knowledge is not fully certain is of course no reason not to use it. We will never be able to achieve completely certain knowledge. The closest that we can come is to use on each occasion the best available knowledge, and at the same time try to improve it. Or, as philosopher John Locke expressed this: "If we will disbelieve everything, because we cannot certainly know all things, we shall do much what as wisely as he who would not use his legs, but sit still and perish because he had no wings to fly." [10]

1.9 Science is a human activity

The only way to guard oneself against being wrong is to never say anything determinate. There are in fact philosophers of science who have maintained that a scientist should behave in that way. Among others, Richard Jeffrey proposed that researchers should never claim anything to be true that concerns the real world. In his view, a scientist should never regard any factual statement as true or false, but instead be satisfied with assigning probabilities to different theories and hypotheses. These probabilities can be quite close to 0 or 1, but they can never be equal to 0 or 1, since nothing is fully certain.[11]

A chemist who lived according to these principles would not take for granted that gold is an element or, for that matter, that metals consist of atoms. She would regard these as highly probable hypotheses, but she would never say "We know that this is so" but only "We hold it highly probable that this is so". This may seem to be an attractive picture of science. But unfortunately, it will never work in practice. The reason for this is that it would make science into a highly complex net of uncertain hypotheses that were connected to each other through various probabilities. We humans are not able to keep very much open at the same time. Such a complex and uncertain mass of knowledge is beyond human comprehension. It is difficult enough to gain overview over a scientific area in the way that science is in practice organized, with statements held to be true or false.

Science is a human activity, and it is subject to all the limitations of human cognition and human intelligence. Any model of science that abstracts from these limitations misses much of its essential nature. In order for us humans to be able to carry through arguments and draw conclusions in science we have to make many simplifications, among them the following:

---

[10] John Locke, *An Essay Concerning Human Understanding*, Abridged and edited with an introduction by John W Yolton, Everyman's Library1976, p. 3. (Book I, Chapter I, Introduction).

[11] Jeffrey, RC (1956) "Valuation and Acceptance of Scientific Hypotheses", *Philosophy of Science* 23:237–249.

Most of what seems to be almost certain is not treated by us as almost but as fully certain. This means that we accept and reject statements and hypotheses, which is quite distinct from just assigning to them very high or very low probabilities.

The accepted scientific statements comprise together the scientific corpus, the mass of scientific knowledge. This corpus consists of everything that is taken for true in science until we obtain a good reason to doubt it. Alternatively, we can describe the corpus as containing all the statements that could be made, without reservations, in a sufficiently detailed scientific textbook.

1.10 The big and the small doubt

Science proceeds largely by questioning and doubting that which was previously taken for given. Since our knowledge is not infallible we sometimes have to question what we already believe in. How can we organize this doubt so that it contributes efficiently to the development of new and better knowledge?

There are two major ways to doubt that in which we believe. One way, the big doubt, consists in "tearing everything down". We rid ourselves of as much as possible of what we believed in previously, and try to build more reliable knowledge from the start. According to the other method, the small doubt, we should continue to believe provisionally in most of what we believed in before, but doubt and question such parts of the corpus of knowledge that do not seem to tally with the general picture. Through a series of many small adjustments, the whole corpus can gradually be transformed.

The big doubt can also be called "philosophical scepticism". Probably the best known proponent of this method was the French philosopher René Descartes (1596 – 1650). He started out from a statement that many had made before him, namely that there does not seem to be anything that one can be fully certain of. How can you know if that which you see is an optical illusion or not? And how can you know if that which you experience is reality or just a dream? In the worst of all cases perhaps the world was created by a God who wanted to deceive us?

But there is one thing according to Descartes, about which the sceptic can be quite certain. The very fact that I am thinking sceptical thoughts proves that I exist. "I think, therefore I am". From this very simple idea Descartes tried to build certain knowledge and "regain" his (hypothetically) lost belief in everyday knowledge. Putting it briefly, the gist of his argument was that since he existed someone must have created him, and the entity that created him must have done this out of good will and must therefore have provided him with reliable senses that give him an accurate picture of the world in which he lives.

But even if Descartes' argumentation was ingenious, other philosophers pointed out that none of the crucial steps in his derivation of everyday knowledge was certain in the sense that 1 + 1 = 2 is certain knowledge. The very first step could be questioned by a sceptic. The experience of thinking does not guarantee the existence of a delimited entity ("I") that performs the thinking. The next step is on even weaker ground. That something exists does not prove that a thinking being has created it. Descartes´ road back from the big doubt to everyday knowledge did not provide the certain knowledge about the world that sceptics ask for.

Many other philosophers, both before and after Descartes, have used the big doubt as a method, as a tool for philosophical inquiry. There is no reason to believe that Descartes or any other philosopher took it to be more than a tool. They did not seriously doubt the existence of the outer world, or that of other living and thinking humans who would be able to read their books. With the big doubt as a tool they tried to provide an account of the basic preconditions of our knowledge. No new knowledge for instance about nature or about human societies was achieved in this way, and this was not either intended.

Instead, science - both the natural sciences and the humanities - are based on the method of the small doubt. Science started out from everyday knowledge, and gradually modified it, sometimes beyond recognition. Modern science would not have been possible unless scientists managed to make an intelligent selection of the small parts of previous knowledge that have been questioned on each particular occasion. One example of this is Galileo's investigations of gravitation. He took much for given, for instance that a measuring rod does not change in length when it is moved. But he questioned other opinions that were taken by many to be almost equally self-evident, for instance that heavy objects fall faster than light objects.

To question everything at the same time is not a recipe for success in science. Instead, what is typically required is to find on each occasion that which it is most constructive to subject to a critical analysis. An analogy can be made with the restoration of a building: You cannot remove all the bearing parts at once.

Hence, the small doubt is an efficient tool for improving our knowledge about the outer world. The big doubt is an efficient tool for improving our understanding of how this knowledge is constructed.

It often happens that people when pressed in a discussion say something like: "Yes, but everything is relative" or "You cannot be certain of anything." Those who say so are not usually prepared, however, to use the big doubt to consistently doubt everything. They only use it against arguments or statements that they select because they have other reasons for not believing in them.

Those who want to apply the big doubt to science usually do not apply it to everyday knowledge. But this is not a tenable position. There are ways to define the word "certain" so that no scientific knowledge is certain: it is not "certain" that evolution has taken place, that DNA carries our inherited properties, etc. But in this sense of "certain", you cannot either be certain of how many fingers you have on your left hand or on whether your own life is real or just a dream.

1.11 Senses and reason

If we want to achieve common, intersubjective knowledge, we have two roads to knowledge to our disposal, namely our senses and our reasoning capabilities. Almost all scientific work is based on the combination of these two roads to knowledge. This means that we combine observations made with our senses with rational arguments that are presented in such a way that others can check whether the conclusions follow. It is only the mathematician and (sometimes) the philosopher who believe themselves to manage with only one of these roads to knowledge.

The two roads to knowledge, senses and reason, are not unique to science. To the contrary, they are the same methods that we use in everyday life in order to find out what the world around us looks like and in what ways we can influence it.

In spite of this, science repeatedly comes up with conclusions that are incompatible with common sense and with our everyday experiences. This is what happened with the discoveries that the earth rotates around the sun, that a light object falls as fast as a heavy one of the same form, etc.

Some of the conclusions from science have become so incorporated into our everyday conceptions that they are nowadays counted as common sense. This applies for instance to our knowledge that the stars are "suns" in space at different and very large distances from us, and likewise that many diseases are spread by small invisible organisms (bacteria and viruses). But there are also many conclusions from science that have not (yet) been incorporated into our everyday conceptions. The insights of relativity theory about the nature of time and space are examples of this.

Is it then really possible that we use the same roads of knowledge in everyday life and in science, and yet science must all the time correct our everyday conceptions? Yes, that is possible, since in science the two roads of knowledge, senses and reason, are employed in a much more systematic way than in everyday life. Science is a systematic search for knowledge.

1.12 Empiricism and rationalism

Although it is unavoidable in science to use both roads to knowledge, there are differences in how much emphasis is put on each of them. In the history of philosophy the term empiricism is used to denote an opinion that puts much weight on empirical observation and draws strict limits for how much can be achieved with reason alone. By rationalism is meant an opinion that puts more emphasis on the exercise of rational reasoning and argumentation. (It is important to distinguish between "rationalist" and "rational". Being rational means to be capable of reasoning in a good way, being rationalist means to give a high status and an important role to arguments based on reason.)

The rationalists of previous periods often claimed that there was no great need of empirical observations. Modern rationalists do not go that far, but they emphasize that empirical observations are not much worth if they do not have their starting-point in a theory telling us which observations are relevant.

Both these traditions have their roots in ancient philosophy. Plato (428 – 348 B.C.) was a proponent of the rationalist view. In his famous story of the cave he claimed that our senses betray us. What we perceive are only shadows from the world of ideas, which is the only real world. The knowledge that we can obtain about it is obtainable with reason, not with our senses. Plato's pupil Aristotle (384 – 322 B.C) is usually counted as an empiricist since he put great weight on our senses as sources of knowledge. He also wrote several important treatises in the empirical sciences.

In the medieval period both rationalism and empiricism had proponents, but rationalism had by far the strongest position. The common view among university teachers was - although this may seem incredible today - that issues in the natural sciences should be settled in the same way as theological and philosophical issues, namely through arguments based on reason. The major starting-point for these arguments was a thorough study of the works of ancient thinkers. University teachers did not perform experiments or other empirical work. Such manual activities were considered improper for members of the higher strata of society to which the academics belonged. Similarly, academically educated physicians did not perform any surgery. They left this task to barber-surgeons who had a much lower status than the physicians.

There was mathematical physics in the Middle Ages. Mathematical models of the movements of bodies were developed, but these models were only used to provide a more precise account of theories developed with purely reason-based methods. Mathematical models were not used to predict the outcome of measurements, and therefore the crucial method was missing that would later be used to test theories against reality. The essential step, which was combining a mathematical model with practical measurements in physical experiments, was first taken by Galileo (1564 – 1642).[12]

1.13 The contributions from craftsmen

The technology of the Middle Ages was not as static as has often been believed, but developed throughout the period. The cathedral builders and the constructors of astronomical clocks were among the important contributors to this development. Even more important progress was made in agricultural technology.[13] Such progress would not have been possible if people had been satisfied by always using available technology. There was a tradition of making experiments and trying out new ideas.

Robert Norman

In 1581 the English instrument maker Robert Norman published a book on magnets, *The Newe Attractive*. Here he laid the foundations of a new understanding of magnetism. It was Robert Norman who came up with the idea of putting magnets on pieces of cork in a bowl of water in order to see how they influenced each other. He found out that the south pole and the north pole are attracted to each other, whereas two south poles or two north poles repel each other. He collected magnetic stones from different parts of the world and disproved the then common opinion that magnets from different mines point in different directions. He also weighed objects of iron before and after they were magnetized, and concluded that magnetism was weightless.

Norman also discovered the magnetic inclination, the angle between the direction of the magnetic force and the horizontal plane. He hung an unmagnetized iron thread in a small holder on its middle so that it balanced horizontally. When he magnetized it, its northern end tilted downwards towards the ground.

In the year 1600 the royal physician William Gilbert published a book in Latin about magnets. There he repeated much of what Norman had written, but without revealing his source. In today's encyclopaedias, and in many books on the history of science, Gilbert is mentioned as one of the major scientists of his period, but Norman is not mentioned at all.[14]

---

[12] Livesey, Steven J. "The Oxford Calculatores, Quantification of Qualities, and Aristotle's Prohibition of Metabasis", *Vivarium* 24:50–69, 1986. AC Crombie, "Quantification in Medieval Physics", *Isis* 52:143-160, 1961.
[13] Jean Gimpel, *The Medieval Machine*, New York 1976, especially pp. 29-58.
[14] Edgar Zilsel, "The Origin of William Gilbert's Scientific Method", *Journal of the History of Ideas* 2:1-32, 1941. Duane H D Roller, *The De Magnete of William Gilbert*, Amsterdam 1959.

Some craftsmen also performed experiments that had no particular practical purpose, but satisfied their curiosity about how nature works. Important discoveries about magnetism were for instance made by an English instrument maker named Robert Norman. (See the fact box.) Other craftsmen laid the foundations of important advances in areas like optics, mechanics, anatomy, and biology.[15]

Many of the qualified craftsmen who performed these experiments had what we would today call the job of an engineer.[16] Others were what we would today have called artists. In the Middle Ages artists were counted as members of a practical trade, and they were organized in guilds in the same way as members of other practical trades.[17] Artists like Albrecht Dürer (1471 – 1528) investigated the linear perspective and thereby also optics. Many artists made careful observations about anatomy and about the proportions of the human body.

For a short period academics were surpassed by craftsmen and artists in the investigation of nature. However, academics soon caught up and started to perform experiments themselves. Among the pioneers we find the anatomist Andreas Vesalius (1514 – 1564), who started to dissect himself, and Galileo Galilei (1564 – 1642), the physicist and astronomer. Galilei was in close contact with craftsmen and learned from their experience. In 1599 he founded the world's first university laboratory (in Padova), and hired the instrument maker Marc´ Antonio Mazzoleni as the world's first laboratory assistant.[18]

In the 17th century when academics started to make experiments, optics and mechanics belonged to the areas that they most often investigated. In both these fields the road had been paved for them by artists and other craftsmen. It was largely from these professional groups that the academics learned to weigh and measure, to try things out in a laboratory, and to make exact observations in nature.

However, modern science was not only a product of the practical trades. Galileo developed a mathematical theory for the movement of bodies, and showed how this theory could be tested against reality through experiments and measurements. His method was based on the experimental tradition from practical trades, but also on the mathematical tradition in physics that had been developed by academics. A particularly important contribution from the latter source was the idea to treat the position of an object as a mathematical function of the point in time. Modern science grew out of this combination of the experimental tradition of craftsmen and the academic rationalist tradition.

In the development of science this interaction between empiricism and rationalism has been further strengthened. Successful science in characterized by the combination of precise and well selected observations with well developed rational arguments.

---

[15] Edgar Zilsel, "The Sociological Roots of Science", *American Journal of Sociology* 47:544-562, 1942.

[16] Hans-Liudger Dienel, "Herrschaft über die Natur? Naturvorstellungen deutscher Ingenieure im 19. und frühen 20. Jahrhundert", pp. 121-148 in Lothar Schäfer and Elisabeth Ströker, *Naturauffassungen in Philosophie, Wissenschaft, Technik, Band III; Aufklärung und späte Neuzeit*, Verlag Karl Alber Freiburg, München 1995.

[17] Arnold Hauser, *The Social History of Art*, Volume One, *From Prehistoric Times to the Middle Ages*, London, Routledge & Kegan Paul 1968, pp. 222-230.

[18] Stillman Drake, *Galileo at Work*, Chicago 1978, pp. 46 and 83. See also: Edgar Zilsel, "The Sociological Roots of Science", *American Journal of Sociology* 47:544-562, 1942, p. 555.

1.14 Episteme and techne approach each other again

With the exception of some individual scientists who were also active as inventors, the techne of practical life and the episteme of the universities developed in isolation from each other.[19] The experimental tradition in practical trades also continued independently of academic activities. Hence, mill builders in the 18th and 19th centuries made advanced experiments and measurements but they had very little contact with academic science.[20] Antonio Stradivari (ca. 1644 – 1737) developed the violin to such a degree of perfection that the best violins made today are essentially faithful copies of his instruments.

Beginning in the 19th century episteme has increasingly "paid back" to techne by producing scientific knowledge that can be used for technical purposes. The chemical and electro-technical industries were pioneers in the systematic use of natural science in technical work. Important new products like the telegraph were the result of discoveries in university laboratories.[21]

It has become more and more difficult to distinguish between episteme and techne. Scientific work has become increasingly dependent on advanced technological equipment. This applies not only to the natural sciences, but for instance also to linguistics, in which computerized analysis of texts has become an indispensable tool. At the same time scientific method is used to develop action knowledge, for instance in medical treatment research, agricultural research, and research on teaching methods. In addition, technology depends more and more on scientific discoveries. Episteme and techne are more closely connected to each other than ever before.

---

[19] Gernot Böhme, Wolfgang van den Daele and Wolfgang Krohn, "The 'scientification' of technology", pp. 219-250 in Krohn, Layton och Weingart (ed.) *The Dynamics of Science and Technology, Sociology of the Sciences vol II*, 1978.

[20] Edwin T Layton, "Millwrights and engineers, science, social roles, and the evolution of the turbine in America", pp. 61-87 in Krohn, Layton and Weingart (ed.) *The Dynamics of Science and Technology. Sociology of the Sciences, vol II*, 1978, Reidel, Dordrecht.

[21] Gernot Böhme, Wolfgang van den Daele and Wolfgang Krohn, "The 'scientification' of technology", pp. 219-250 in Krohn, Layton och Weingart (ed.) *The Dynamics of Science and Technology, Sociology of the Sciences vol II*, 1978, Reidel, Dordrecht.

2 Reasoning rationally

In the previous chapter we noted that rational reasoning is one of the two major roads to scientific knowledge. Scientific communication takes the form of arguments so presented that they can be followed step by step. But this way of communicating is not unique for science.

2.1 Rational conversations

Even in simple everyday conversations we expect a certain level of rationality. It is disappointing to talk to someone who is unwilling to explain what she means or dismisses others' arguments without confronting them. Our demands on ourselves and others, concerning communication, include expressing oneself as clearly and precisely as possible, taking the arguments of others seriously, and accepting a correct argument independently of who put it forward.

In some contexts the risk of misunderstandings is particularly severe. This applies for instance to scientific discussions and many public debates. In order to communicate well in such contexts we sharpen our requirements, and try to conduct a well-ordered rational conversation.

The requirement of rationality in scientific discussions does not import a new element into our everyday conservations. Instead it consists in strongly emphasizing an aspect that is already there: making ourselves as well understood as possible. Sometimes this can only be achieved at the cost of other qualities of everyday conversations, such as spontaneity. Nevertheless, the forms of rational thought are basically same in science as in everyday thinking. Just as in science, many everyday situations require a strong focus on precision and intelligibility, such as when two car mechanics try to figure out why a motor does not start.

Such demands of rationality are of course not unique to Western culture. !Kung hunters discuss animal behaviour in a way that has much in common with scientific discussions. They distinguish clearly between data (for instance tracks that they have seen) and theories, and also between observations and hearsay. They are willing to admit gaps in their own knowledge, and they are often sceptical against each other's statements. (Not surprisingly they often know much more about animal behaviour than what Western zoologists do.)[22]

A rational conversation demands logical and analytical ability of its participants. But this is not all that it demands. Most of us have observed how analytically highly trained and capable persons fail to communicate with each other because neither of them is willing to engage sufficiently in the other's endeavours. A rational conversation also requires human empathy of its participants. Unempathetic rationality does not function in the social contexts where rationality is intended to fulfil its tasks.

Therefore rationality is in part a matter of emotions, in particular one's emotional attitudes to one's own arguments and convictions and to those of others. Rationality requires presence

[22] Blurton Jones, N and Konner M "!Kung knowledge of animal behavior", pp. 326-348 in RB Lee and I DeVore (eds.) *Kalahari hunter-gatherers*, Cambridge, Harvard UP, 1976.

The history of university seminars

The modern university seminar has its origin in German classical linguistics. A linguistic seminar that started in Göttingen in 1737 is usually counted as the world's first university seminar.[23] Other subjects followed suite, first history and biblical exegetics.[24] In the first half of the 19th century seminars became increasingly common in German universities.

The university seminar merged at least three different traditions in higher education. One of these was the medieval disputation exercises, the public defence of an assigned thesis against the attacks of an opponent. Another was the private tuition that many university teachers gave against payment. It was often less authoritarian than the official lectures, and contained various kinds of exercises. A third tradition was the informal discussions in associations for various disciplines that had been formed by the students themselves.[25]

The seminars broke with the strict academic hierarchy, since it was knowledge rather than status that counted in the seminar room.[26] However, the seminars were not open to everyone, but only to particularly merited students. A visitor in the German university town of Halle in the 1790's wrote: "Most of the seminarians affect peculiar and atypical mannerisms, by which they very noticeably distinguish themselves; you can spot them at a great distance on account of their attire and other small details".[27]

As late as in the late 19th century seminars usually took place in the professor's own living room, probably primarily because of the unsuitability of university lecture halls for the purpose. Usually, the seminars were followed by a joint meal.

of reason, but it does not require absence of feelings. The emotional component in a rational attitude was emphasized already in the text that is usually called Plato's seventh letter.

> "After much effort, as names, definitions, sights, and other data of sense, are brought into contact and friction one with another, in the course of scrutiny and kindly testing by men who proceed by question and answer without ill will, with a sudden flash there shines forth understanding about every problem, and an intelligence whose efforts reach the furthest limits of human powers."[28]

2.2 Forums for scientific discussions

There are many forums for scientific discussions. Scientific journals often contain articles that criticize or otherwise comment on the results and conclusions put forward by other researchers. At least as importantly, these journals normally only publish articles that have been favourably reviewed by specialists in the relevant area. (See the fact box.) Funding applications for research are usually subject to a similar review procedure. Scientific

---

[23] Wilhelm Erben, "Die Entstehung der Universitäts-Seminare", *Internationale Monatsschrift für Wissenschaft, Kunst und Technik* (Berlin) 7:1247–1263, 1335–1347, 1913. William Clark, "On the Dialectical Origin of the Research Seminar", *History of Science* 27:111–154, 1989, p. 119.

[24] Bonnie G Smith, "Gender and the Practices of Scientific History: The Seminar and Archival Research in the Nineteenth Century", *American Historical Review* 100:1150–1176, 1995. Walter Prescott Webb, "The Historical Seminar: Its Outer Shell and Its Inner Spirit", *Mississippi Valley Historical Review* 42:3–23, 1955–1956, p. 5.

[25] William Clark, "On the Dialectical Origin of the Research Seminar", *History of Science* 27:111–154, 1989.

[26] Bonnie G Smith, "Gender and the Practices of Scientific History: The Seminar and Archival Research in the Nineteenth Century", *American Historical Review* 100:1150–1176, 1995.

[27] William Clark, "On the Dialectical Origin of the Research Seminar", *History of Science* 27:111–154, 1989, p. 127.

[28] *Epist*. 344b. Pl. *Epist.* 344b (translation J. Harward), http://classics.mit.edu/Plato/seventh_letter.html.

Ten commandments for seminar participants.

As a seminarist you should…
1. … engage in others' intellectual endeavours: read their texts, listen to them, and try to understand how they think.
2. … never appeal to your own authority or that of others. Competence has to be proved anew on each occasion.
3. … let every issue be thoroughly discussed, and never heap up more arguments than what can be handled in the discussion.
4. … never try to avoid criticism by changing topics or asking a diverting counter-question.
5. … focus on the strongest, and most certainly not the weakest, arguments in favour of a standpoint you are trying to refute.
6.… take note of the developable parts of ideas that you do not fully endorse.
7.… strive for precision, and rather say something precise that needs to be corrected than something imprecise that evades criticism.
8. … never hesitate to reveal your own lack of knowledge.
9. … try to find weaknesses in your own standpoints and always be prepared to give them up.
10. … be merciless towards all arguments and ideas, and merciful to all seminar participants.[29]

conferences provide occasions for public discussions. Usually, each conference talk is followed by a short question-and-answer period. In addition there are many informal forums for scientific discussions: personal meetings, correspondence, and not least the scientific seminar. The seminar is the forum for scientific discussions that students and young researchers first come into contact with. (See the fact box.)

A seminar is, at least ideally, a form of well-ordered rational discussion. Therefore the seminar discussion should be impersonal in one important respect: Each argument should be judged according to its own value, independently of who put it forward. The purpose of the discussion is not to find out *who* is right or wrong, but *what* is right or wrong. The discussion should therefore focus on standpoints and ideas as abstracted from their bearers. Intellectual feuds and other entrenched conflicts should be avoided, and every participant should search for mistakes in her own standpoints and arguments, not only those of others. (See the fact box.)

The impersonal ideal of the seminar stands in contrast to common forms of political debate. In the latter case attempts are seldom made to treat standpoints as independent of the persons who put them forward. To the contrary, each participant's goal is often to show that the opponent is wrong. Seminar participants should be on their guard, so that the seminar does not take the common form of a political debate. The suitable form of political debates is a quite different issue that will not be treated here.

[29] Sven Ove Hansson, *Verktygslära för filosofer*, Thales 1998, p. 102. Sven Ove Hansson, "Editorial: Philosophical Seminars", *Theoria* 71(2):89-91, 2005.

Scientific journals

The French *Journal des Sçavants* and the English *Philosophical Transactions*, both first published in 1665, are usually counted as the first scientific journals. Already in the first decades the editors of these journals acquired the help of experts in various areas to review manuscripts. This system has since then been developed and systematized, and is now called peer review. Almost all international scientific journals apply it.[30]

When an article has arrived in the editorial office the editor sends it out for review to one, two or more experts. (Their number differs between journals). These experts, who are called referees, are colleagues of the author, but they need not be more knowledgeable or more experienced than the author herself. The task of refereeing for journals is unpaid and time consuming but it is usually considered a matter of honour to accept such assignments.

The referees should be anonymous for the author. Many journals also try to keep the authors' identities hidden from the referees.

When the editor has received the reports, she makes a decision. This decision can be to accept the article without changes (uncommon), to accept it conditionally on certain changes proposed by the referee(s), to recommend the author to submit a revised version for renewed review, or to reject the article. Articles that receive strongly positive reviews are normally accepted, and those that receive strongly negative reviews are rejected. In intermediate cases the views and tastes of the editor can determine the issue. When informing the author of her decision, the editor usually appends the referee reports.

It is unavoidable that some articles will be unfairly rejected. This would have been unbearable if it were not for the pluralism of the system. Articles that are rejected in one journal can have a second chance in another journal (and a third chance in a third journal…). Almost all good research (and some bad research) is published either in peer-reviewed international journals or in books published by international publishers that apply a similar referee system to ensure scientific quality. This is true of all areas of science.

2.3 Stepwise argumentation

In a rational discussion the arguments should be put forward sufficiently clearly and in sufficiently small steps to make it possible to follow and check them. This is needed in order to avoid logical mistakes. It is also necessary to make it clear on what assumptions the conclusions are based.

The need for such a careful stepwise presentation is not unique for science. It is also used for instance in legal argumentation and when we try to find an error in a computer program.

However, a conversation or a written text would be unbearable if all arguments were presented in the fullest detail possible. In practice it is both allowed and advisable to use more abbreviated expressions. What rationality demands is that all such abbreviations should be expansible. In other words there should be a more extended argumentation that can be put forward when needed. Uncontroversial or self-evident assumptions and steps of thought can then be eliminated. It is of course essential that no unclear or controversial elements of the argumentation are hidden behind such abbreviations.

[30] Harriet Zuckerman och Robert K Merton, "Patterns of Evaluation in Science: Institutionalisation, Structure and Functions of the Referee System", *Minerva* 9:66-100, 1971. John C Burnham, "The Evolution of Editorial Peer Review", JAMA 263:1323-1329, 1990. David A Kronick, "Peer Review in 18th-Century Scientific Journalism", JAMA 263:1321-1322, 1990.

2.4 Ambiguity and vagueness

One of the demands on a rational conversation is that the words that are important for the argument should be as well-defined as possible. Lack of clarity in the meaning of a word or a phrase can be of two different types.

First, a word can have two or several quite different meanings. This is called ambiguity. As an example, "approach the bridge slowly" has different meanings in a driving lesson and a violin lesson.

The word "normal" is ambiguous in a practically important way. What does it mean for instance that a blood pressure is normal? There are at least two clearly distinguishable interpretations:

Within the limits of what is common.
Within the limits of what is harmless.

Failure to distinguish between the two concepts of normality has caused problems in medical practice. Clearly, whether a condition should be treated or not does not depend on how uncommon it is. Nevertheless, conditions that are harmless but unusual have sometimes been treated, whereas conditions that are harmful and usual have not been treated.[31]

The other form of unclarity is vagueness, the absence of a sharp limit between cases covered and not covered by the word. The word "bald" can be used about a person who has no hairs on his head, but also about a person who has a few hairs on his head. There is no sharp limit, as can be seen from the following list of pronouncements:

A person with 0 hairs is bald
A person with 1 hair is bald
A person with 2 hairs is bald
A person with 3 hairs is bald…

Arguably, a single hair cannot make the difference between baldness and its opposite. Therefore, it seems as if this list could be extended to any number of hairs. If that were the case, then we would all be bald. In practice there is of course a limit for the uses of this word but the limit is not a very sharp one. The word is vague.

In a similar way it is unclear where to draw the limit between a hill and a mountain, between sand and gravel, between heavy and light traffic, between copying and printing, between publicity and information, etc. In practice most words are vague at least to some extent. It is almost only specifically defined technical terms that have a fully precise and well-delimited meaning.

2.5 The uses of precision and imprecision

Our attitudes to ambiguity and vagueness differ between linguistic contexts. Many jokes and most puns are based on ambiguity. Poetry is another linguistic context where ambiguity is

[31] ML Johnson Abercrombie, The Anatomy of Judgement, London, Hutchinson 1960 p. 93-109.

used to achieve the desired effects. In diplomacy and negotiations vagueness and ambiguity are used to find formulations that parties with different standpoints can all accept.

Jokes, poetry, and diplomacy are a diverse set of linguistic situations, but they have the positive use of ambiguity and vagueness in common. The usefulness of ambiguity in politics may be more debatable. On one hand it can be used just as in diplomacy, for instance in coalition-building. On the other hand it can also be used to give promises that are conceived differently by different groups of voters.[32]

In science there are strong reasons to avoid ambiguity and vagueness. Only by doing this can we achieve a reasonable degree of intersubjectivity when communicating about results and theories. Similarly, the need for precision is usually large in engineering. It is not sufficient to say that a pressure vessel can hold "high pressure". This pressure must be specified in precise and measurable terms. In the same way, medical records employ a technical language that makes it possible to describe symptoms more in detail than with everyday language.

It should be no surprise that the elimination of vagueness and ambiguity in scientific and technical language is usually bought at the price of losing the poetic qualities of language. Such qualities are important enough, but their full expression will have to be achieved in other contexts.

2.6 Definitions

Each branch of science has a long list of technical terms with special meanings. There are different ways to reach agreement on the definition of these terms. In many cases a definition becomes accepted through the same type of open competition as between scientific hypotheses. In other cases a committee for terminology or standardisation issues terminological recommendations. This applies for instance to the established names of chemical substances, that are the outcome of extensive committee work. However, most other chemical terms, such as the names of chemical reactions, are not decided by committees. Due to the large amount of technological standardization, the technological sciences have many terms that have been defined by committees.

There are two major types of definitions: lexical and stipulative definitions. In principle, the difference between them is simple: A lexical definition reports common usage. A stipulative definition states how the author herself intends to use the word (and at least by implication recommends others to use it).

In general it cannot be seen from a definition if it is lexical or stipulative. Therefore, this should be explicitly stated (which it is usually not). Lexical and stipulative definitions have the same structure. In both cases, the definition always contains a *definiendum* (that which is to be defined) and a *definiens* (that by which it is defined). They are connected with a *defining connective* for instance in the following way.

[32] Raymond Dacey, "The Role of Ambiguity in Manipulating Voter Behavior", *Theory and Decision* 10:265-279, 1979.

An ion…	*definiendum*
is…	*defining connective*
an electrically charged atom or molecule.	*definiens*

Definitions of properties and other qualifiers contain an additional component, a delimiter.

A chemical reaction is…	*delimiter*
reversible…	*definiendum*
if and only if…	*defining connective*
it can be made to change direction by an arbitrarily small change of one of the factors that has an influence on the equilibrium (for instance pressure, temperature, or concentration).	*definiens*

The function of the delimiter is to determine the definition‘s area of application. The word “reversible” that is defined in the above example is also used in many other contexts for instance about disease processes and environmental change. The delimiter “a chemical reaction” clarifies that the above definition of “reversible” is limited to this particular usage of the term.

The terminology is complicated by the fact that the word “definition” has two meanings. First, it can designate the whole complex of a definiendum, a defining connective, a definiens and sometimes a delimiter. This is how the term was used above. Secondly, the word “definition” is also often used about the definiens, for instance when “electrically charged atom or molecule” is said to be a definition of “ion”.

Large scholarly dictionaries such as the *New Oxford English Dictionary* are excellent sources of lexical definitions. Such dictionaries often contain fairly well identified characterizations of the various usages that a word or an expression has in common usage.

There are two major ways in which a lexical definition can be deficient: It can be too broad or too narrow. A definition is too broad if the definiens is applicable to more than what the definiendum covers according to common usage. Consider a definition of “traffic accident” as “accident on a road where vehicles are involved”. This is a too wide definition since it would include an accident in which a drunk person is hurt as he stumbles into a car that is parked at the roadside. This is hardly something that we would call a traffic accident.

A lexical definition is too narrow if the definiens does not cover everything that the definiendum covers according to common usage. Consider a definition of “traffic accident” as “accident on a road in which a motor vehicle is involved”. This is a too narrow definition since it would exclude a road accident where two cyclists collide at high speed. A definition can have both deficiencies, i.e. it can be too narrow and too broad at the same time.

If a word has several different meanings they should be separated and listed in a lexical definition. However, a lexical definition cannot add more precision than what was available from the beginning in common usage. Since it should reflect common usage, it will also reflect the vagueness and ambiguity that a word has in common usage. If more precision is needed, a stipulative definition should be constructed instead.

2.7 Three roads to more precise concepts

When developing a stipulative definition in order to increase precision, there are essentially three options in the choice of the term to be defined. First, we can retain but redefine the original word whose imprecision caused the problem. The English language contains many words that have been redefined in order to have a more precise meaning in a scientific context: "force", "weight", "density", "risk", "intelligence", "dialect", etc. A particular type of redefinition is that which consists in eliminating vagueness by the introduction of exact although arbitrary limits in a key variable. This applies for instance to the scientific redefinitions of "sand" (particle size), "storm" (air velocity), and "obesity" (weight in relation to square height).

Redefinitions of everyday words will lead to the same word having both a colloquial usage and a scientific, more precise usage. This situation can also arise when scientific words are taken over by everyday language where they receive a less precise meaning (for instance "energy").

Many scientific redefinitions have become accepted in everyday language after a longer or shorter period. This applies for instance to the word "fish". We do not any longer count whales as a fishes, and consequently the word "whalefish" is obsolete. The word "star" has gone through a similar development. Nowadays we all count the sun as a star, but not Venus. A few generations ago the opposite standpoint was taken for granted. More recently, the word "planet" obtained a new, more precise meaning. Pluto was counted as a planet from its discovery in 1930 to 2006, when the International Astronomical Union (IAU) adopted a new definition of "planet", with criteria not satisfied by Pluto. It will probably take a few years before this new definition has changed common usage.

The redefinition of everyday words has the advantage that simple and well-known words can be used in science, but on the other hand this is often a somewhat deceptive form of simplicity. A person who reads a scientific text may well believe that she understands the key terms better than what she actually does. Redefinitions can also give rise to unnecessary conflicts about usage. Words such as "weight" and "force" have more restricted meanings in physics than in common usage. Physicists may be tempted to require that everyday usage of these words be adjusted to the scientific definitions. However, it is far from evident that everyday language is always best served by taking over redefinitions that have been made for quite different linguistic contexts.

The second method to achieve precision is to assign a qualifier to the original term before redefining it. Examples of this are "electrical resistance" and "physical pendulum". This method has the advantage of giving rise to fewer misunderstandings than simple redefinitions. On the other hand the specialized terminology that is achieved in this way will often be clumsy, which is one of the reasons why this method is not often used.

The third method consists in introducing a word that is new for the context, to be used as a precise, specialized term instead of the less precise term in everyday language. Examples of this are leukaemia (blood cancer), sodium chloride (salt), monochromatic (one-coloured), and putrefy (rot). This method is very common. It has the disadvantage of resulting in a specialized language that may seem more difficult. On the other hand it has the essential advantage that the reader will not erroneously believe herself to understand the specialized terms because of confusion with their everybody meanings.

It has often been said that stipulative definition is free, so that you can choose to redefine words arbitrarily. If the purpose with the definitions is to use them in communication with others, then this is certainly not true. In practice, stipulative definitions have to be in fairly close contact with common usage. It may be useful to redefine the word "sand" by giving it exact delimitations for instance it terms of particle sizes. However, it would be meaningless to redefine "sand" so that it also includes large blocks.

2.8 Value-laden words

The prevalence of value-laden scientific terms differs between the sciences. In the natural sciences they are relatively unusual. In the social sciences they are very common, and some terms such as "justice" and "welfare" have a strong value component. In medicine several central terms are value-laden, including "disease", "handicap", and "normal". Technological science also uses value-laden words to a much larger extent than the than natural sciences. Examples are "risk", "disaster", "user friendly", "environmentally friendly", and "natural".

A word's value-ladenness can be either positive ("environmentally friendly") or negative ("unnatural"). Some words have different value-ladenness for different persons, for instance "socialist", "market economy", and "religious". Independently of its type, value-ladenness is usually strongly associated with the word and almost impossible to remove.

A lexical definition of a value-laden term should reflect its value-ladenness. Value-neutral lexical definitions of words like "disaster" or "betrayal" would be misguiding. In contrast, stipulative definitions can at least in principle be constructed through the elimination of values from value-laden concepts. For this to have any chance of success, a new word should be chosen to replace the value-laden one. Hence, it is a vain endeavour to introduce a value-neutral definition of the word "bureaucracy". Its derogative connotations cannot be washed off. Chances are much better if another word (perhaps "administration") is chosen for the value-neutral concept.

2.9 Creativity and criticism

The creation of new ideas is an essential part of the research process. The creative process is seldom described in scientific articles or other research reports. There is a good reason for this, namely that the nature of this process should not have any influence on how others judge the validity of the research output. Hence, a mathematician only gives a proof of her new results. She does not tell her readers which special cases she tried out first, or what mental pictures she used. This is sensible, since in the end it is only the actual proof that counts. For similar reasons the historian does not say much about the rather vague reasons that made her start reading exactly the sources that she did. The experimentalist does not tell us about the series of failed experimental set-ups that in the end led her to the one that she could use. This restricted way of reporting often gives the impression that the road to scientific results was much straighter than it actually was.

It is an unavoidable part of the research process that researchers come up with many ideas that they later reject. Ideas that are put forward to be tried out are often called hypotheses. Sometimes it is sufficient to evaluate an hypothesis theoretically or against previous observations, but often new experiments or other observations are needed. An empirical test of scientific hypotheses is called hypothesis testing. (More about this in Chapter 5.)

Hence, research proceeds through a combination of creative and critical processes. To some extent there can be a division between persons of these processes. Some researchers are much more creative and develop "wild ideas" which they need the help from others to sort out. Others are less innovative, but more skilled in criticising or testing the ideas of others.

However, such a division of labour cannot be driven too far. A fast feedback from creativity to criticism is needed, so that creativity is led into constructive tracks of thought. Therefore a researcher must be capable of criticising and sorting among her own ideas. There must be in her mind an "ongoing dialogue between two voices, one imaginative and the other critical".[33] But this inner dialogue is not enough, It has to be linked to an outer dialogue with other researchers. In other words the combined creative-critical process should operate both on an individual and a collective level.

2.10 Intuition

Creativity cannot build on logic alone. It requires an innovative process in which intuition has an important role. Intuition is an elusive concept. For our purposes it is sufficient to note that having an intuition means to have a conviction that cannot be intersubjectively justified. If I have a general feeling that a wooden beam of a certain dimension is sufficient to carry a certain load, but I am unable to justify this conviction, then it is an intuition. If I can support this claim with systematically collected observations, or with calculations based on such observations, then this is not (only) an intuition but also a justified standpoint. If a mathematician has a general feeling that a certain theorem is true, but cannot prove it, then her belief in the theorem is (merely) intuitive. Contrary to a proof, such intuitions do not constitute a reason for other mathematicians to believe in the theorem. In other words, intuitions are not intersubjectively valid.

Many scientific insights have begun as vague intuitions. Therefore it would not be wrong to describe intuition as another scientific source of knowledge, in addition to the senses and reason. However, it differs from them in providing only hypothesis to test, not facts or data to be relied on. Therefore intuition is not a source of knowledge on par with senses and reason.

Romanticists of the 19th century cherished the myth of the intuitive genius who relied completely on immediate inspiration. Major artworks such as novels and symphonies were claimed to have been written in just a few days. In practice these are rare exceptions. Art, just like science, requires hard work in addition to inspiration. Sometimes scientific intuition is idealized in a similar way. Great researchers are claimed to have come up with their ideas in dreams or in other effortless ways. These descriptions are often much exaggerated. (See the fact box.) Creative ideas do not come suddenly to the unprepared mind, and it takes hard work

[33] Peter Medawar, *Pluto's Republic*, Oxford University Press 1984, p. 46.

to prepare the mind. Furthermore, not only good ideas but also bad ideas can come up in dreams. Scientific creativity is often a highly collective process. Studies of successful research strongly indicate that scientific discoveries usually have their origin in the dynamics of a creatively interacting group of researchers, rather than in the mind of a lonely researcher.[34]

Doing research while asleep?

The best known example of a research idea said to have originated in a dream is Friedrich August Kekulé von Stradonitz's (1829 – 1896) proposal that aromatic compounds like benzene have six carbon atoms in a ring. This is his own description of his moment of inspiration:

"I turned my chair to the fire and dozed. Again the atoms were gambolling before my eyes. This time the smaller groups kept modestly in the background. My mental eye, rendered more acute by repeated visions of the kind, could now distinguish larger structures, of manifold conformation: long rows, sometimes more closely fitted together; all twining and twisting in snake-like motion. But look! What was that? One of the snakes had seized hold of its own tail, and the form whirled mockingly before my eyes. As if by a flash of lightning I awoke; and this time I also spent the rest of the night in working out the consequences of the hypothesis."[35]

There was however a hitch: The Austrian engineer and chemistry teacher Josef Loschmidt hade proposed the ring-formed structure several years before Kekulé. A letter by Kekulé reveals that he had in fact read Loschmidt's text on the subject.[36] In all probability, Kekulé's insight did not originate in a dream.

[34] Kevin Dunbar, "How Scientists Build Models. In Vivo Science as a Window on the Scientific Method", pp. 85-99 in Lorenzo Magnani, Nancy J. Nersessian, and Paul Thagard (eds.) *Model-Based Reasoning in Scientific Discovery*. New York, N.Y.: Kluwer Academic/Plenum Publishers, 1999.

[35] Quoted from JT Davies, The Scientific Approach, 1965, pp. 15-16.

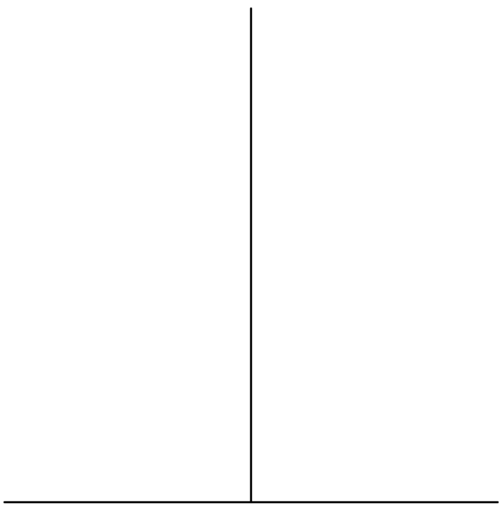
[36] Christian R Noe and Alfred Bader, "Facts are better than dreams", *Chemistry in Britain* February 1993, pp. 126-128.

3 Making observations

We call science empirical because it requires that we base our knowledge about the world we live in on observations that we make with our senses. Sometimes we can observe directly that which we want to investigate. This happens for instance when an ethologist studies the behaviour of animals or when a political scientist follows a decision process as it goes on. On other occasions, our observations will have to be more indirect. They can be indirect in different ways. The study of historical sources only gives indirect knowledge about life in previous times. The traces of particles in a bubble chamber only give us indirect knowledge about the micro-world.

3.1 The imperfection of the senses

Although we use our senses as a source of knowledge we do not use them uncritically. Look at these two lines:



The two lines seem to differ in length, but with a ruler we can convince ourselves that they are of equal length. Why are we more convinced by a measurement with a ruler than by the direct visual impression? Obviously we assume that our vision is imperfect, but also that there is something objective behind our subjective perception of length. If we had relied strictly on our senses as infallible, then we would have believed that the two lines are of unequal length, and consequently that the ruler changed its length when we turned it around.

A similar phenomenon occurs when we visit the house of mirrors. There we would rely more on a plummet or a water level than on the eye's impression of what is perpendicular or horizontal. As these examples show, we assume already in everyday life that there is even more regularity in nature than what we observe directly with our senses. By experience we know that this mode of interpretation is efficient and provides us with better predictions than what we could obtain by relying on our senses as infallible.

The concept of temperature is another, somewhat more complex example of the same phenomenon. We have temperature senses that give rise to subjective experiences of hot and cold. A classical physiological experiment shows that these senses are also subject to illusions. Place one hand in a bowl with ice cold water and the other in a bowl with hot water. Lift them up after a while and place them both in a bowl with water of medium temperature. The hand that came from hot water will feel cold, whereas the other hand will feel hot. If we believed fully in our senses we would be forced to believe that the two hands were in surroundings with different temperatures. It would be difficult to obtain a coherent picture of the world if we reasoned like that.

The solution is in this case to assume that there is an objective temperature, that can only be approximatively perceived by our senses, but can be fairly accurately measured with a thermometer. There is a good but far from perfect correlation between such temperature measurements and our subjective experience of temperature. Through additional investigations we can find regular patterns in the deviations between the two and explain them in a credible way.

Hence, the concept of temperature has been derived from our sensory experiences through a process that is more abstract than the one used to derive the concept of length. Other concepts, like power and energy, have been extracted through similar, even more abstract processes.

3.2 Observations are theory-dependent

Temperature measurement exemplifies that behind most scientific observations there are theories about how various features of the world are connected to each other. There would not have been much reason to read off scales beside mercury pillars, if we did not believe that there is an objective temperature to measure. More generally speaking, our observations are largely *theory-dependent* in the sense that they start out from our conceptions about what types of empirical observations are most suitable for finding out the regularities in nature.[37] These conceptions, in their turn, are not “purely theoretical”, but are based on previous empirical observations.

It would therefore be wrong to say that scientific observations always precede scientific theories. However, it would be equally wrong to say that theories should always come before observations, since without observations there is no basis for theorizing. The relationship between theories and observations is comparable to that between the hen and the egg; there is not much point in debating which of them came first.

Sometimes observations have been questioned because there was no reasonable theoretical basis for interpreting them. When Galileo observed four moons around Jupiter in his telescope, there was no optical theory to explain the function of the telescope. In the same observations strange coloured patterns could be seen, which Galileo rightly interpreted as instrument errors. How then could one know that the four moons were not similarly instrument errors? In addition to the observation, theory was needed to justify its

[37] NR Hanson, *Patterns of Discovery*, CUP 1958.

interpretation.[38] The sceptical attitude that some of his contemporaries took towards the telescope was not at all as strange as it can seem today when we have much better telescopes and in addition a well-established optical theory that explains the anomalies in early telescopes.

The microscope was subject to much more severe epistemological problems than the telescope. It is easy to confirm that a telescopic picture corresponds to what you can see with your bare eyes. Just direct a telescope at a distant object, and compare what you see in the telescope to what you see if you walk close to the object and look at it from a short distance. There is no corresponding method to confirm directly the function of a microscope. Our reliance on microscopes depends much more on optical theory. This gave rise to a widespread philosophical scepticism against the microscope as a source of scientific knowledge. It took much longer time for the microscope than for the telescope to become a generally accepted scientific tool. The microscope was invented in the late 16th century, but it was only in the 1830s that it was generally accepted and used in biological and medical research.[39]

Today we have no problems with interpreting what we see in an optical microscope. We know that the visual impressions that we receive through microscopes have the same type of origin as those that we obtain with the unaided eye. Optical technology offers a continuation of the process that begins when we approach an object from far away in order to see it more in detail. It is much more difficult to account for the pictures obtained in electron microscopes. Spontaneously we tend to regard them as the same type of pictures as those obtained in optical microscopes, only with additional magnification. This is an erroneous interpretation from a physical point of view, since it would not be possible with ordinary light to distinguish the small structures that are revealed in an electron microscope. The interpretation of pictures from electron microscopes is problematic for us in much the same way as the interpretation of pictures from optical microscopes were problematic in the 17th century.[40]

Gradually, more and more complex and theory-dependent operations have come to be counted as observations. We cannot observe the interior of the sun in the everyday meaning of the word “observe”. Nevertheless astrophysicists have been able since the 1960s to detect neutrinos originating inside the sun. The registration of such neutrinos is often considered as an observation of the parts of the sun that they came from.[41] This exemplifies how the concept of observation evolves as theories and technologies develop.

3.3 Technology helps the senses and the memory

Natural science would be helpless without measurement instruments. A large part of all discoveries have been made thanks to an increasingly advanced technology for scientific observations. In this sense, episteme is strongly dependent on techne. Two early technological advances should be mentioned that were particularly important for the development of

[38] Yaakov Zik, “Science and Instruments: The telescope as a scientific instrument at the beginning of the seventeenth century”, *Perspectives on Science* 9:259-284, 2001.

[39] Catherine Wilson, *The Invisible World: Early Modern Philosophy and the Invention of the Microscope*, Princeton University Press 1995. David Wootton, *Bad Medicine. Doctors Doing Harm Since Hippocrates.* Oxford University Press 2006, pp. 110-138.

[40] Olaf Breidbach, “Schattenbilder: Zur elektronmikroskopischen Photographie in den Biowissenschaften”, *Berichte zur Wissenschaftsgeschichte* 28:160-171, 2005.

[41] Dudley Shapere, “The concept of observation in science and philosophy”, *Philosophy of Science* 49:485-525, 1982.

modern science. One of them is the mechanical measurement of time, the other the glass technology that made both the telescope and the microscope possible.

With a measurement instrument we can restrict the use of our senses to tasks that we can perform with high precision, for instance reading off a thermometer instead of using our own temperature senses to determine how hot or cold it is. Measurement instruments can also be used to observe phenomena that could otherwise not be observed at all, such as very small and very distant objects. We can observe electromagnetic radiation with wavelengths other than the narrow range that our eyes are sensitive to. In the same way we can register sounds with wavelengths that we cannot hear. With chemical analysis we make a wide variety of observations that would not otherwise be possible. In these and many other ways technological progress has multiplied our opportunities to observe the world that we live in.

Technology extends not only our senses but also our memory. Human memory is not at all as reliable as we would like it to be. We often compare memories with books or recordings, but these are highly misleading analogies. It is more accurate to regard human memory as a mechanism for reconstruction from fragmentary parts.[42] Mechanical registration, most elementarily in the form of writing and drawing, is much more resistant against decay. Therefore it is important for researchers to make meticulous notes of their observations before memory fails them. Laboratory scientists keep laboratory diaries, and field workers for instance in biology and social anthropology keep field diaries. In archaeology it is a fundamental principle to document the location of an object before you remove it.

Before the invention of photography, drawings were indispensable for scientific documentation in most fields of knowledge. They still have an important role in areas such as botany and anatomy. Drawings have important pedagogical advantages, since it is much easier for a draughtsman than a photographer to give prominence to important structures. But as an aid for research drawings also have clear disadvantages. A person who draws for instance a microscopic preparation tends to "see" and draw structures that she expects to see. In this way drawings can reinforce and perpetuate fallacious interpretations of natural phenomena – a risk that many scientists have been acutely aware of. In order to avoid it the German histologist C. F. Link went as far as to base his research on the microscopic structures of plants on drawings made by a person without botanical knowledge. In a book published in 1806 he wrote: "I have mostly left the observations to be performed exclusively by my drawer, Mr Schmidt. The open-mindedness of an observer who is ignorant of all botanical theories guarantees the accuracy of the drawings." [43]

In the middle of the 19th century photography provided new opportunities for the documentation of scientific observations. One of the most interesting developments was the photography of microscopic preparations. It was relatively simple to connect a camera to a microscope. Already in 1845 a book was published with pictures that had been taken in this way.[44] Proponents of the new technology claimed that photographs offered a "mechanical

---

[42] Leonard Zusne och Warren H Jones, *Anomalistic Psychology*, Lawrence Erlbaum, Hillsdale 1982, pp. 149-150.

[43] CF Link, Vom inwendigen Bau der Gewächse, und von den Saftbewegungen in denselben, Göttingen 1806. Quoted in MJ Schleiden, Grundzüge der wissenschaftlichen Botanik, Theil I, Leipzig 1845, p 105. In its turn quoted in Olaf Breidbach, "Representation of the Microcosm – The Claim for Objectivity in 19th Century Scientific Microphotography", *Journal of the History of Biology* 35:221-250, 2002, p. 223.

[44] Olaf Breidbach, "Representation of the Microcosm – The Claim for Objectivity in 19th Century Scientific Microphotography", *Journal of the History of Biology* 35:221-250, 2002, p.224.

objectivity" not obtainable with drawings.[45] However, for several decades most microscopists preferred drawings to photographs. One of their arguments was the extensive manipulations of pictures that microphotographers devoted themselves to, both in the photographic processes and through retouching of negatives.[46] It was not until the 1880s that microphotography was generally accepted. New printing technology probably contributed to this break-through; at this time it became possible to reproduce photographs directly in the press.[47]

The famous bacteriologist Robert Koch also contributed to the success of microphotography. In an article published in 1881 he pointed out that published photographs made it possible for researchers to critically assess the interpretations that colleagues made of their observations. Hence, a bacteriologist could study photographs published by another scientist and discover if a bacteria was incorrectly identified.[48] Koch proposed that published photographs should be completely unretouched, which was an important way to increase their credibility. It is interesting to note that Koch differed from previous proponents of photographic registration by focusing on the intersubjectivity, rather than the objectivity, that it made possible.

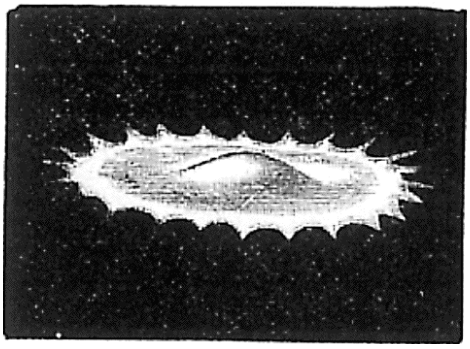
The new printing technology also increased the use of photographs in other areas of science.[49] A spectroscopist who could now publish photographs instead of drawings of the spectrum of the sun wrote enthusiastically: "The spectrum is absolutely untouched. It represents therefore the work of the sun itself, and is not a drawing either made or corrected by hand." [50]

Photographic technology also made it possible to study fast and short-lived events that could not be seen with the un-aided eye. In 1878 and 1879 the American photographer Eadweard Muybridge (1830-1904) managed to document in detail the movements of galloping horses. He placed 24 high speed cameras in a row after each other on a racing track. With pictures taken closely after each other he could solve the old controversy whether a galloping horse has at least one foot on the ground all the time. (It does not).[51] Today, fast events can easily be followed with film or video technology.

In 1907, the English physicist Arthur Mason Worthington published the book A Study of Splashes with photographs showing how a drop of liquid hits a liquid surface. He achieved extremely short periods of exposure by performing the experiment in a dark room that was

---

[45] Klaus Hentschel, "Wissenschaftliche Photographie als visuelle Kultur" *Berichte zur Wissenschaftsgeschichte* 28:193-214, 2005, esp. p.195.
[46] Olaf Breidbach, "Representation of the Microcosm – The Claim for Objectivity in 19th Century Scientific Microphotography", *Journal of the History of Biology* 35:221-250, 2002. Frank Stahnisch, "Die Photographie als Hülfsmittel mikroskopischer Forschung", *Berichte zur Wissenschaftsgeschichte* 28:135-150, 2005. Elke Schulze "Zeichnung und Fotografie – Statusfragen", *Berichte zur Wissenschaftsgeschichte* 28:151-159, 2005.
[47] Colin Clair, *A History of European Printing*, Academic Press, London 1976, pp. 404-405.
[48] Robert Koch, "Zur Untersuchung von pathogenen Organismen", *Mittheilungen aus dem königlichen Gesundheitsamte* 1:10, 1881. Quoted in: Olaf Breidbach, "Representation of the Microcosm – The Claim for Objectivity in 19th Century Scientific Microphotography", *Journal of the History of Biology* 35:221-250, 2002.
[49] Colin Clair, *A History of European Printing*, Academic Press, London 1976, pp. 404-405.
[50] Klaus Hentschel, "Wissenschaftliche Photographie als visuelle Kultur" *Berichte zur Wissenschaftsgeschichte* 28:193-214, 2005, quotation from p.197.
[51] Paul C Vitz and Arnold B Glimcher, *Modern Art and Modern Science. The parallel analysis of vision*, Praeger New York 1984, pp. 113-115. Helmut Gernsheim and Alison Gernsheim, *The history of photography from the camera obscura to the beginning of the modern era*, revised and enlarged edition, London 1969, pp. 435-438. Eadweard Muybridge, *Descriptive Zoopraxography, or the science of animal motion made popular*, University of Pennsylvania 1893.



To the left a drawing and to the right a photography from Worthington's publications on what happens when a drop of water falls on a water surface.

momentarily lit by an electric discharge. Previously he had performed the same experiment without photographic registration. Instead a person sitting in the dark room made drawings based on what he saw at the moment when the room was lit. A comparison between the drawings and the photographs reveals an interesting difference: The patterns on the water surface appear to be much more regular on the drawings than on the photographs.[52] The obvious explanation is that the draughtsman emphasized what he believed to be fundamental structures and evened out what he believed to be inessential irregularities.

Towards the end of the 19th century photography was supplemented by film and sound recordings, that are equally important for the documentation of scientific observations. Sound recordings make it possible for musicologists to study details in sounding music that are difficult if not impossible to notate with sufficient precision. Sound recordings are equally indispensable for linguists studying spoken language. In the natural sciences films and sound registrations are used routinely for a wide variety of purposes. We do not any longer have to rely on memories of the behaviour of animals or the appearance of sun spots. Just as Robert Koch pointed out, registration through visual and oral media contribute to making science more intersubjective. Instead of relying on the individual researcher's interpretation of her observations, we can recapitulate what she saw and heard, and reassess her interpretations. Increased Internet publication of research articles will accentuate this development. On the Internet, space restrictions are less severe than in printed journals. Therefore a much larger material of photographs, films and sound recordings can be made accessible to the readers.

Another type of documentation has substantially increased the accuracy of observation reports in the natural sciences, namely direct mechanical registrations from measurement instruments. Originally such registration was achieved by connecting the instrument directly to a pen writing for instance on a moving paper. Now it is usually achieved by sending measurement data directly to a computer for storage. This form of mechanical registration contributes to increasing the intersubjectivity of scientific observations.

[52] Peter Geimer "Fotografie as Wissenschaft", *Berichte zur Wissenschaftsgeschichte* 28:114-122, 2005.

3.4 Selected observations

The observations that are made with a scientific purpose are only a very small part of the many observations that humans make of nature or social life. However, even if these observations are comparatively few, they are in general better prepared, better performed, and better documented than most other observations. In addition, they have been selected to cover the needs of science. Therefore, science should primarily be based on them rather than on other types of observations.

It has often been proposed that science should pay more attention to various types of spontaneous observations. People who have seen UFOs or ghosts often require that their reports be taken as scientific proof that what they saw was real. They are supported in this by promoters of occult ideas. One famous example is Sir Arthur Conan Doyle (1859-1930), best known as the author of the Sherlock Holmes novels. Doyle collected stories from people who believed themselves to have seen fairies. His purpose was not, as one might have expected, folkloristic. Instead he tried to prove that these creatures really exist. The many reports that he received were evidence enough for him. He wrote:

> "[T]hese numerous testimonies come from people who are very solid and practical and successful in the affairs of life. One is a distinguished writer, another an ophthalmic authority, a third a successful professional man, a fourth a lady engaged in public service, and so on. To waive aside the evidence of such people on the ground that it does not correspond with our own experience is an act of mental arrogance which no wise man will commit."[53]

It may seem sympathetic to take such stories about fairies, or ghosts, UFOs, or other strange phenomena, as reliable sources. However, experience shows that such an attitude will very often lead to untenable conclusions. It is for this reason, and not due to the arrogance that Doyle accused researchers of, that this type of observations are treated as unreliable in a scientific context. These observations have the weakness of being spontaneous and unplanned. They have been made in an unsystematic way, are usually badly documented and lack support in mechanical registration. (The photos that Conan Doyle published were forgeries that had been sent to him.[54]) Long experience and many psychological experiments show that we are bad at observing something we are not prepared for, such as an accident or an unknown object on the sky.

Science should be based on observations that are as reliable as possible. Such observations should be well planned and made systematically and with as good technical support as possible.

3.5 Four types of observations

The observational ideal differs between the sciences, largely according to how they distribute their interest between general knowledge and knowledge of particular facts. In the search for

[53] Arthur Conan Doyle, *The Coming of the Fairies*, Samuel Weiser, New York [1921] 1972, pp. 140-141.
[54] David Hewson, "Cottingley fairies a fake", *The Times* 18/3 1983, p.3.

knowledge about a particular phenomenon, the ideal is to observe it as directly and undisturbedly as possible. Hence, a social scientist investigating how the French president makes his decisions would ideally follow all his meetings and consultations.

In the search for general knowledge we need to make observations that support conclusions about more than the particular cases that are observed. For that purpose it is mostly advisable to manipulate the study object in various ways in order to find out how it reacts under different conditions. In other words: in order to obtain general knowledge we need to perform experiments.

An experiment is an observation in which the observer (experimenter) controls the major conditions that determine that which is observed. If I happen to see two chemical substances being mixed, I can observe what happens, but this is not an experiment. If I carefully control the identity of the substances, weigh them, mix them in a planned way and then observe what happens, I have made an experiment. In the latter case I have controlled the conditions of my observation.

Observations can be divided into four types according to the degree of control.

1. *Experiment*: A planned observation where the observer (experimenter) manipulates the variables that are believed to influence the outcome.
   Example: In order to investigate the effect of artificial irrigation of the yield of rye, a number of rye fields are used. Half of them are selected for irrigation, and the yields of irrigated and un-irrigated fields are compared.
2. *Controlled observation*: A planned observation where the variables believed to influence the outcome are registered (measured), but cannot be manipulated in order to see what happens if they are changed.
   Example: We sit in an observation tower close to an eagle's nest, making notes about the behaviour of the eagles according to pre-determined rules for what to note and how.
3. *Uncontrolled observation*: An observation where the variables believed to influence the outcome are not known.
   Example: During a thunderstorm someone observes that some flashes of lightning seem bluer than the others and that all of these seem to go down into a lake.
4. *Rumour*: Observations that are reported to have been made but cannot be traced down to their origin.
   Example: It is well known in this area that in previous years wolves attacked humans, but nobody can tell us when it happened or who was attacked.

There are no sharp limits between the four types of observations. In practice they are contiguous fields on a continuous scale representing degrees of control. Sometimes it is difficult to determine if an observation is an experiment or a controlled observation. The term "natural experiment" is often used for controlled observations where the relevant variables are adjusted through a natural process in a way that is similar to how an experimenter would herself have adjusted them in an ordinary experiment. Suppose for instance that we want to know how foxes are influenced by two different diseases. We do not want to spread the disease ourselves, but we find two isolated islands. One of the two diseases has started to spread on one of the islands, and the other disease on the other island. Then nature itself has started an "experiment", and it remains for us to make the appropriate observations on the two populations. However, natural experiments are seldom if ever as well-arranged as ordinary (artificial) experiments can be. In this case we would have liked the foxes to be distributed randomly between the two groups. This, of course, nature will not do for us.

In the search for general knowledge, experiments are the desirable form of observation. Controlled observations are second-best whereas uncontrolled observations come third and rumours of course come last.

In the search for knowledge about particular facts, experiments are not relevant. Instead controlled observations are the first hand choice, followed by uncontrolled observations and lastly rumours. Some examples when controlled (non-experimental) observations are the observational ideal are anatomical dissections, botanical inventories, geological surveys and demographic statistics. What is common to these examples is that their primary aim is knowledge about particular facts. It is only for general knowledge that we need manipulating methods.

3.6 When the observational ideal cannot be met

In science we should always use as well-controlled observations as possible. However, many scientific issues cannot be studied in a very well controlled way. This does not make these issues less important or their study less scientific. What is unscientific is to rely on less well-controlled observations than necessary, or to draw more far-reaching conclusions from uncontrolled observations than they allow for.

There are two major reasons why we may have to be content with observations that do not satisfy the scientific ideal: Observations satisfying the ideal can be practically impossible, or they can be unethical. Both these reasons can obtain in the humanities, in the social sciences, and in the natural sciences. It is impossible to study the origin of stars in experiments or listen to spoken language from the Middle Ages. It would be unethical to expose humans to dangerous toxic substances or to massive totalitarian propaganda. It is also unethical to make close observations of people who do not know that they are observed.

What is possible or impossible changes with time and, not least, with technological development. In the 1930s elementary particles with high energy could not be studied experimentally. Instead, physicists sent up balloons in order to catch high-energy cosmic radiation on photographic negatives. Now many of these particles are studied in particle accelerators. This transition from non-experimental to experimental studies has made it possible to obtain more general knowledge about elementary particles.[55] Our views on research ethics have also evolved. Ethical restrictions on scientific experiments on humans are much stricter today than they were half a century ago, and the same applies to experiments on animals.

All types of sciences - the humanities, social sciences, natural, and technological sciences – are affected by the practical and ethical limitations that thwart the attainment of observational ideals. Historical sciences often have the largest problems with this, since it is

[55] Harré, Great Scientific Experiments, p.112.

never possible to observe directly what happened long ago. This applies both to historical humanities such as general history and history of languages, and historical natural sciences such as palaeontology and cosmology.

When it is impossible or unethical to make observations that conform with the scientific ideal there are two major ways to go. They can be called *methodological adaptation* and *object adaptation*.

Methodological adaptation means that the object of study is investigated with a less well-controlled method than what would have been scientifically ideal. The palaeontologist investigating how dinosaurs moved cannot use the methods employed on living animals, namely direct observation and filming of their movements. Instead she uses more indirect methods such as studies of skeletons and in lucky cases footprints. The physician studying the effects on the human body of extreme cold cannot expose humans to dangerously cold conditions. (Such experiments were performed by Nazi war criminals.) Instead she collects documentation from accidents in which humans were exposed to such conditions.

Object adaptation means that instead of investigating the primary object of study, investigations are performed on another object, a *substitute*. The substitute has to satisfy two criteria. First it must be sufficiently similar to the primary object of study so that its properties and behaviour can be used to predict those of the primary object of study. Secondly the substitute should be accessible to more controlled studies than the primary object of study. A simple example of a substitute is the reduced model of an airplane that is used instead of the full scale airplane in wind tunnel experiments. Theory and experiments have taught us that results from such experiments give a fairly good picture of the aerodynamic properties of the full scale airplane. For practical reasons a full size passenger plane cannot be studied in a wind tunnel, but with a model of the plane this is possible. Another example is the test dummies used in crash tests of cars as substitutes for human victims. A test dummy equipped with accelerometers and other instrumentation records information that can be used to draw conclusions about how a human would have been injured in similar situations.

Obviously the method for studying a substitute has to be adjusted to the substitute, even if this means that a method is chosen that would not have been optimal for the primary study object. Experimental archaeology provides an interesting example of this. The best way to study how a tool is used in practice is a controlled, non-experimental observation of an experienced, traditional user of the tool. When it comes to tools that no living person has used, such as an axe from the European Stone Age, this is of course not possible. Instead experiments are performed where different ways to use the tool are tested systematically. The test person then serves as a substitute for the original users of the tool. An interesting aspect of this procedure is that an experiment serves as a replacement for a controlled observation.

When the ideal observation cannot be performed it is often advisable to combine information from various types of sub-ideal observations. Therefore methodological adaptation and object adaptation should not be seen as alternatives but as supplementary strategies. The best result is in most cases obtained by combining them. For ethical reasons we cannot study the effects on humans of poisonous substances through experiments on humans. Instead we perform both non-experimental observations of humans who have been accidentally exposed to the substances (methodological adaptation) and experimental observations on animals or cell cultures (object adaptation). Since it is not possible to study the origin of stars experimentally, we perform both experimental studies in particle accelerators of related phenomena (object adaptation) and non-experimental observations through telescopes (methodological adaptation).

3.7 The observer herself

One factor is present in all observations: the observer herself. The presence of the observer is problematic in at least two different ways: through her influence on the study object, and through her interpretations of the observations.

The *influence problem* is usually the most difficult of the two to deal with. In many cases it is unavoidable that the observer will influence that which she observes in an uncontrolled way. The social anthropologist who visits a distant village will not see village life under normal conditions, but rather how it develops in the presence of an exotic and inquisitive visitor. A person who fills in a questionnaire will be influenced by the way in which the questions are asked. When a biologist studies the behaviour of animals her own presence can make them behave differently than otherwise.

In all scientific and technological measurements that which is measured is influenced by the instrument. The thermometer will heat or cool the object whose temperature is being measured, the speedometer reduces the speed of the vehicle, the ammeter "steals current", etc. Through quantum mechanics we know that all such influences on a measurement cannot be eliminated at the same time. However, in most cases other types of errors are much larger than the small effects of quantum mechanical uncertainty.

The observer's influence on the study object is normally an undesired effect. There is no general recipe for eliminating it. Instead this is a methodological problem that every branch of science has to attack with its own specific methods. A large part of the methodological discussions in science concerns how such influences can be reduced or avoided.

The *interpretation problem* arises whenever an observer's interpretations of her observations are affected by her expectations. If a physician investigates the effects of a treatment on a group of patients, it is in practice impossible for her not to be influenced in her assessment of what she believed beforehand about the treatment.

In a classic experiment, psychology students were assigned the task of performing learning experiments on two groups of rats. They were told that one of the groups belonged to a strain with a high capacity for learning, whereas the other group had normal learning abilities. This was not correct, since the rats belonged to the same strain and had been randomly distributed between the two groups. Nevertheless, the students reported better results with the rats that they believed to be faster learners.[56]

Contrary to the influence problem, the interpretation problem is often simple to eliminate. If we are searching for correlations between property A and property B in different individuals, then the investigator who determines whether an individual has property A should not know whether that individual has a property B, and vice versa. An assessment is called blind if information is withheld from the investigators so that their expectations of the

[56] Rosenthal, R., & Lawson, R. "A longitudinal study of the effects of experimenter bias on the operant learning of laboratory rats, *Journal of Psychiatric Research*, 2: 61-72, 1964.

outcome cannot influence their judgments. In the example with the rats the students should have divided tasks among themselves so that those who assessed the learning capabilities of the rats did not know which rat belonged to which group.

Clinical drug tests are usually performed as double-blind tests. Different patient groups receive different drugs (and often one of the groups receives placebo, a drug with no physiological effect). Double-blindness means that neither the physician who treats the patient nor the patient herself knows which treatment she receives. The purpose of this arrangement is to suppress the effects of both the physician's and the patient's expectations, so that the physiological effects of the drugs come out more clearly.

In many animal experiments histopathologists assess tissue samples from the animals in order to determine if there are any differences between animals that received different treatments. Traditionally, such assessments have been made without blinding, which means that the person who assesses the tissues knows which animal group each slide comes from. Histopathologists who oppose blinding have pointed out that small and subtle changes may be difficult to identify unless one knows which type of changes to look for. However, this is not a tenable argument for unblinded assessments. It is much better to let the actual assessment be preceded by an unblinded pre-assessment that aims at identifying and describing the type of tissue changes that can be observed in the experimental group. With this description as a starting point, the actual assessment is then performed by another person who does not know the origin of the tissue samples she is assessing.[57]

Unfortunately there is still a rather wide-spread resistance among histopathologists against blinded assessment procedures.[58] Therefore histopathology is in many cases still unblinded, which of course reduces its reliability. The situation is similar in many other areas of scientific and technological research where blinding should be used in order to prevent the investigator's expectations from influencing the outcome. Blinding is still an exception in many research areas where it should be the rule, in spite of the fact that blinding is seldom expensive or very cumbersome to introduce.

3.8 Being prepared for the unexpected

Even well-planned scientific investigations often give rise to surprising results. Many important discoveries have been made unexpectedly in investigations aimed at finding out something else. The term "serendipity" is often used for the ability to make use of unexpected discoveries and occurrences.[59] Two of the best-known examples are Roentgen's discovery of X-rays when he was looking for quite other phenomena and Fleming's discovery of penicillin when his bacteria cultures were attacked by moulds. In 1979 the Australian clinical pathologist Robin Warren discovered that gastric ulcer is caused by bacteria. He did this when performing routine work – not research – in a hospital laboratory.

These are rather drastic examples. Many more unexpected results are obtained in a less dramatic way, for instance when an investigation yields the opposite result of what was

[57] Keith Prasse, Paul Hildebrandt, and David Dodd "Letter to the Editor", *Veterinary Pathology* 23:540-541, 1986.
[58] Tom Holland, "A Survey of Discriminant Methods Used in Toxicological Histopathology", *Toxicological Pathology* 29:269-273, 2001, especially p. 272.
[59] Duncan C Blanchard, "Serendipity, Scientific Discovery, and Project Cirrus", *Bulletin of the American Meteorological Society* 77:1279-1286, 1996.

expected. In a large study in the sociology of research it was concluded that about half of the results obtained by the researchers were unexpected.[60]

The role of unexpected results in science is often misunderstood. Such results are described as "chance discoveries". This expression is misleading, since these discoveries put high demands on the researcher's ability to change tracks and draw the right conclusions from unexpected information. At least one researcher before Röntgen had seen traces of X-rays, but only thought of them as an unidentified source of error that disturbed his experiment. Several researchers before Fleming had seen the effects of moulds on bacteria cultures, but they did not choose to investigate the issue further.[61]

Everyday research gives rise to a large number of surprises and opens a large number of possible side tracks. We cannot follow them all. To be successful the researcher has to strike a delicate balance between persistency in pursuance of her research goals and openness to promising side tracks.

3.9 Critical analysis of sources

Experimental scientists have developed methods to make as well-controlled observations as possible. It has fallen to the historian's lot to develop methods for dealing with the opposite situation namely when the observational record is only know to us through sources with low or uncertain reliability.

Historical method is useful not only in the humanities but also in the natural sciences. Astronomers are often interested in ancient observations that may refer to comets or supernovas. Biologists may be interested in old texts revealing the distribution of plants and animals. A chemist who wants to know the taste of a toxic substance does not have to risk her health by putting it in her mouth. Instead she can find this information in literature from the 18th century, when chemistry was more a matter of tasting and smelling than weighing and measuring.[62] All this requires a critical appraisal of the old sources. Furthermore, in all scientific fields we need to critically assess more recent research reports written by others. For this purpose as well, the methods developed by historians can be highly useful.

Historical sources are usually assessed according to four criteria, namely authenticity, dependence, temporal distance, and tendency. In scientific contexts we often have reasons to add a fifth criterion, namely competence.

Authenticity. The first question to ask in a critical appraisal of a text is: Is it authentic? Was it written by the alleged author and at the stated point in time? Suppose for instance that we read an eye-witness report of the great eruption of Etna in the year 1169. We must then ask ourselves if the text is really as old as it is said to be. In order to determine this linguistic and historical expertise is needed.

When it comes to modern scientific texts another type of authenticity assessment may be needed: Does the report describe correctly the observations made, or is it fraudulent? In order

[60] Kevin Dunbar, "How Scientists Build Models. In Vivo Science as a Window on the Scientific Method", pp. 85-99 in Lorenzo Magnani, Nancy J. Nersessian, and Paul Thagard (eds.) *Model-Based Reasoning in Scientific Discovery*. New York, N.Y.: Kluwer Academic/Plenum Publishers, 1999, especially p. 90.
[61] Samuel Rappaport och Helen Wright, eds, Science: Method and Meaning, New York University Press 1963, pp. 141-142.
[62] Lissa Roberts, "The Death of the Sensuous Chemist: The 'New' Chemistry and the Transformation of Sensuous Technology" *Studies In History and Philosophy of Science Part A* 26:503-529, 1995.

to determine whether fraud has taken place, both the published texts and filed materials must be carefully studied. In many cases inspections of instruments and other equipment may be needed. In addition, attempts to repeat the suspected experiments or observations may be called for.

Dependence. If a source is in dependent of other sources then this increases its value. The requirement of independence has two major aspects. First, a good source is *direct*, i.e. it has no intermediates. This is why preference is given to eye witness reports. It is common to distinguish between primary and secondary sources, where the primary sources of those that satisfy the criterion of directness. A person who saw the eruption of Etna in 1169 with her own eyes is a primary source, whereas someone who heard eyewitnesses tell their stories in a tavern in Palermo is a secondary source.

Secondly, a good source should be *unadjusted*, which means that it has not been adjusted to other accounts of the same events. When the police interrogates witnesses of an accident they do not want them to talk to each other before the interrogation. If they have not done this, then any concordance between their testimonies will have much higher evidential value. The same principle applies to the assessment of historical sources.

Distance in time. Our memories are subject to decay. Therefore memories written down many years after an event are much less reliable than memories that were written down immediately. This is the reason why letters and diaries are much better sources for the historian than autobiographies. If an old person tells us about a lemming migration she saw in her youth, then this story has much less scientific value than an old letter in which she described what she saw when her memory was still fresh.

Tendency. Every source has to be judged also according to the informant's purpose. Has the person who claims to have seen the Loch Ness monster possibly something to gain from attention given to the monster? (Local hotel and restaurant owners have been very active in some cases like this.) Did the farmers who told stories about wolves killing their livestock have a particular purpose, such as obtaining compensation for the losses or a permit to hunt the wolves? What connections does the expert who wrote the report on the nuclear accident have to the power industry, alternatively to political groups with strong standpoints on nuclear power?

Competence. Some types of observations require training to be made reliably. If a person who claims to have seen a very rare bird has little ornithological competence then this report will have much less value than if the observation was made by a person with good ornithological competence.

On some occasions the competence needed to make an observation is non-academic. Perhaps the best example of this is the need for knowledge about conjuring in observations of alleged psychic phenomena. Persons who claim that they have supernatural powers sometimes create seemingly inexplicable phenomena with techniques that are well-known to conjurers. The best observations of so-called psychics have been made by professional conjurers, not by academic scientists.

3.10 Measurements

Many observations are made in such a way that it is useful to summarize the outcome of each particular observation with a number. With a *measurement* is meant an observation whose result is expressed in numbers. Measurements can be divided into three major groups according to the type of measurement scale that is used: ordinal, interval and ratio scales.

An *ordinal scale* is a scale in which *a higher number corresponds to a higher degree of a given property*. We can for instance measure people's subjective perceptions of outdoor heat and cold by letting them describe the weather on different occasions on the following scale:

1. unbearably cold
2. inconveniently cold
3. cold
4. cool
5. hot
6. too hot
7. unbearably hot.

An ordinal scale like this provides no other information than the order between its items. Numerical differences between the positions on the scale have no particular significance. In our example it would be misleading to say for instance that the difference between cold and hot is two units, and thus larger than the difference between too hot and unbearably hot that is only one unit. In fact, there are no units in this sense in an ordinal scale. This is measurement without measurement units. Ordinal scales are often used to summarize subjective assessments for which no units are available. Other examples, that are more practically relevant, are measurements of pain and satisfaction in work life.

It is a common mistake to use ordinal scales as if they were measurements with units. One of the most common errors is to calculate the average (arithmetic mean) of ordinal measurements. The average is not even well defined for ordinal measurements. Instead, the median (the middle value) can be used to indicate the central tendency in a distribution of ordinal measurements.

An *interval scale* is a measurement scale with a *constant measurement unit*. One example is temperature measurement on the centigrade scale. It is quite reasonable to say that the difference between 19°C and 21°C is twice as large as that between 24°C and 25°C. (Each difference of one degree corresponds, according to the original definition, to an equal increase in the volume of the liquid used in the thermometer.) For similar reasons the average (arithmetic mean) of a set of interval measures is well defined. Therefore it is quite reasonable, and often useful, to calculate the average of a collection of temperature measurements.

Hence, it is meaningful to subtract values on an interval scale. However, it is in general not meaningful to calculate the ratios between such values. 20°C is not twice as hot as 10°C, and neither is 2°C twice as hot as 1°C. The reason for this is that the zero point on the centigrade scale does not correspond to total absence of the measured property. As everybody knows, there are temperatures below 0°C.

A *ratio scale* is an interval scale with a *zero point* that corresponds to total absence of the measured property. There is a ratio scale for temperature, namely the Kelvin scale, whose zero point is equal to the absolute zero, the lowest possible temperature. It is quite correct to say that 400° Kelvin (127°C) is twice as hot as 200° Kelvin (- 73°C).

Most scientific measurements are made with ratio scales. This applies to length, energy, mass, momentum, etc. An object with a length of 0 meters has no extension. If the luminous intensity is 0 cd, them it is dark, etc.

In all measurements, errors are possible. If an interval scale (or ratio scale) is used, then the size of an error can be expressed in the measurement unit. An error in a temperature measurement can be specified by saying that the measured temperature was for instance 0.15°C too high.

If a measuring instrument tends to yield either too high values in most measurements or too low values in most measurements, then there is a systematic error. Such an error can be corrected by *calibrating* the instrument, i.e. adjusting it so that it yields on average the right value.

Even with a well calibrated instrument errors remain, but they are randomly distributed between too high and too low values. The instrument error of a well calibrated instrument is usually expressed as an interval (confidence interval). That an analytic balance has the error interval ± 0.05 mg means that the true value is expected to be at most 0.05 mg higher or lower than the measured value.

Modern scientific measurement instruments often have very small error intervals. However, it does not follow that the investigations performed with these instruments have the same high degree of accuracy. Often the major sources of error can be found in the preparation of a measurement rather than in the actual use of the instrument. Errors should be searched for, assessed and minimized in the whole process that leads to an outcome in the form of a measurement.

4 Making experiments

In the previous chapter we concluded that for many, but far from all, scientific purposes we should attempt to make observations in the form of controlled experiments. In this chapter we are going to have a closer look at how experiments should be constructed in order to provide the desired information.

The word "experiment" is used in everyday language in a much wider sense than in science. If we say that a person "experiments with drugs" we do not usually refer to a scientific or other intellectual activity. Here, when talking about experiments, we use the term in its more limited, scientific sense.

4.1 Many types of experiments

When you say "experiment" most people will probably think of work performed in laboratories by people in white coats. Probably most experiments are performed in various types of laboratories, but it is not location that determines if an observation is performed under experimental conditions. The essential criterion is instead that the conditions under which the observation takes place are sufficiently well controlled.

In some cases it is an advantage to perform an experiment outside of the laboratory. Suppose for instance that we want to investigate how eagles react to different types of food. Then we could enclose eagles in a cage, give them different types of food and study their behaviour. It is much better, however, to build an observation platform close to an eagle's nest, put different types of food in the vicinity and study the behaviour of the eagles. Then we will find out how they react in their natural habitat. The study of animal behaviour was for a long time hampered by too many experiments being performed in laboratories rather than in the natural environments of the animals.[63]

Some of the sciences are based on experiments to such a high degree that they are called "experimental sciences". This applies in particular to the natural sciences and to psychology. But experiments are also performed in other sciences. Hence, experimental archaeology is a discipline of its own. The best way to determine how stone age man can have produced and used his tools is to try this out in an experiment. In experimental history of technology, replicas of ancient machines are built and their technical properties are investigated. In experimental economics, investigations are made of how humans react on markets with different conditions. Experiments are also performed for instance in linguistics and musicology. The basic principles and problems of experimental work are common to all sciences where experiments are made.

Experiments are also made in other traditions than Western academic science. As was mentioned in Chapter 1, the experimental tradition among European craftspeople is older than modern natural science. Furthermore, experimental traditions prevail among many so-called primitive peoples. Hence, the Mende people in Sierra Leone have a special word, "hungoo", for experiment. A "hungoo" can consist in planting two seeds in adjacent rows, and then

[63] Harré, *Great Scientific Experiments*, p. 61.

measuring the output in order to determine which seed was best. This was probably an original habit, not one brought to the Mende by visiting Europeans. Similar experiment also occur in other parts of the world.[64]

Both farmers and craftspeople have experimented primarily with the purpose of solving practical problems. The same focus on techne still prevails in modern agricultural and technological science. We try out new technical constructions and new seeds in order to find out if they yield better results than their alternatives. Similarly: new treatment methods are tried out in clinical experiments in order to determine if they are better or worse than the old methods.

In "pure" science, that is directed at episteme, experiments are performed with another purpose, namely to find out how the world is constructed. There are large similarities, but also differences, between the two types of experiments. In experiments aimed at understanding conditions are often simplified in order to exclude factors from the experiment that are considered to be less important. Hence, physical experiments can be performed in vacuum in order to eliminate the effects of air. In a technical experiment that aims at testing a new construction this simplification is not allowed, (unless you are either developing technology to be used in vacuum or have good reasons to believe that the air has no influence in the case in question).[65]

4.2 Constructing an experiment

Before performing an experiment we have to build an experimental set-up and make it work in practice. This construction work is often more time consuming that the actual observations. It is largely a technological activity. Many of the phenomena that we study experimentally are produced with technical means, and technology is also used for measurement and observation.[66] Many scientific advances have been achieved thanks to a technological invention that made a new type of experiment possible.[67] It would therefore not be wrong to count this type of technological work as one of the sources of scientific knowledge. However, since this activity is "only" used as a preparation for observations that we make with our senses, it will not be treated here as a basic source of knowledge on par with the senses and reason.

The role of technology in experimental science is so great that some have wished to regard natural science as applied technology, rather than technology as applied science.[68] However, it is more constructive to regard the relationship between science and technology as one of interaction. A clear example of this interaction can be gained from the joint development of

[64] Paul Richards, "Farmers also experiment: A neglected intellectual resource in African science", *Discovery and Innovation* 1:19-25, 1989.
[65] Ronald Laymon, "Applying Idealized Scientific Theories to Engineering", *Synthese* 81:353-371, 1989.
[66] Peter Kroes, "Physics, Experiments, and the Concept of Nature", pp. 68-86 in Hans Radder (ed.), *The Philosophy of Scientific Experimentation*, University of Pittsburgh Press 2003, esp. pp. 70-71.
[67] Srdjan Lelas, "Science as Technology", *British Journal for the Philosophy of Science* 44: 423-442, 1993.
[68] Peter Janich, "Physics – Natural Science or Technology", pp. 3-27 in Krohn, Layton and Weingart (eds.) *The Dynamics of Science and Technology. Sociology of the Sciences, vol II*, 1978, Reidel, Dordrecht, p 13.

electro-technology and the physics of electricity. New machines provided the conditions for making new discoveries, that in their turn gave the conditions for new machines, etc.[69]

It is not possible to provide a general recipe for the construction of an experimental set-up. This is to a very high degree a creative process, and the problems that have to be dealt with differ between different areas of research. Generally speaking, however, the tasks of an experimental apparatus can be summarized as follows:

realize: to make the phenomenon appear that you intend to study.
separate: remove disturbing factors, so that the phenomenon under study appears as clearly as possible.
control: bring those factors under control that can have an influence on the phenomenon.
observe: obtain as exact observations as possible.

For example we can consider an experimental investigation of the propelling force of ship propellers with different constructions. Realizing the phenomenon, i.e. the propelling force, is fairly simple in this case: you install the propeller on an axis that is driven around by a motor, and place it under water. What the phenomenon needs to be separated from in this case is primarily water movements that do not depend on the propeller. This problem is solved by performing the experiment in an experimental basin rather than in open water. The factor that we have to control is primarily the effect of the motor. This can be done either by using a motor with a constant effect or by applying an apparatus that can set the effect of the motor on different precise levels. Finally we need to find a way to observe the propelling force that the propeller gives rise to.

Another experiment can be mentioned, namely one that aims at finding out if a particular chemical substance gives rise to liver damage in fish. In order to realize the phenomenon we expose fish to the substance, for instance by adding it to the food given to fishes in an aquarium. Separating the phenomenon means here primarily to distinguish between the effects of the substance and other changes in the liver, for instance normal age changes. This is done in this case by comparison with a control group that is not exposed to the chemical but is otherwise subject to the same conditions as the experimental group. The major factor to control in this experiment is the dose (which is difficult to control in the feeding of fish[70]). Finally we need a method to observe the possible damages. This is done with microscopy and chemical analysis of the liver, both in exposed fish and in the control group.

We are now going to consider more closely two of these four components, namely separation and control.

---

[69] Walter Kaiser, "Die Entwicklung der Elektrotechnik in ihrer Wechselwirkung mit der Physik", pp. 71-120 in Lothar Schäfer and Elisabeth Ströker, *Naturauffassungen in Philosophie, Wissenschaft, Technik, Band III; Aufklärung und späte Neuzeit*, Verlag Karl Alber Freiburg, München 1995.

[70] Magnus Breitholtz, Christina Rudén, Sven Ove Hansson and Bengt-Erik Bengtsson "Ten Challenges for Improved Ecotoxicological Testing in Environmental Risk Assessment", *Ecotoxicology and Environmental Safety* 63: 324–335, 2006.

4.3 Separation

It is often difficult to separate out the phenomenon under study from other phenomena that can give the same or similar effects. There are two major ways to do this, *elimination* and *effect separation*. By elimination is meant that some experimental apparatus is used that eliminates one or several of the phenomena whose effects can be confused with the effects of the phenomenon under study. We are going to look closely at an instructive historical example of elimination, namely an early investigation of the use of pendulums as "measuring instruments".

Pendulums have been used for dowsing since far back in time, both in Europe and China. The Romans made predictions by holding a pendulum over an alphabet. In later years it has become common to hold a pendulum over a map in order to obtain "information" about where to find oil, archaeological sites, or lost objects.[71]

The chemist Michel Eugène Chevreul (1786 – 1889) made observations in the 1830´s that initially made him believe that a pendulum reacted to some previously unknown natural force. As a pendulum he used an iron ring hanging in a string. When he held it over mercury, it gave a clear indication. When he put a pane of glass between the iron ring and the mercury, the indication disappeared. It seemed as if the pendulum was influenced by a force that was shielded by the pane of glass.

But Chevreul was not fully convinced. Could this possibly be what we would today have called a psychological phenomenon? Perhaps his own expectations made the pendulum swing? In order to settle the issue he had to perform an experiment in which he could separate a possible unknown physical force from a possible expectation effect. He did this by eliminating the expectation effect. He repeated the experiment with blind-folded eyes. An assistant now and then introduced the glass pane between the pendulum and the mercury, but without letting Chevreul know when he did this. If it was only the effects of expectations that ruled the pendulum, the effect would then disappear, otherwise it would remain.

It turned out that when Chevreul, who was himself holding the pendulum, did not know that the iron was shielded from the mercury, then shielding had no effect on the movement of the pendulum. Chevreul drew the correct conclusion from this, namely that his own expectations, conveyed through small subconscious muscular movements, ruled the pendulum. He wrote about his experiment in a letter to Ampère in 1833. Since then, a large number of experiments, both with the pendulum and with the dowsing rod, have shown that dowsing is fully explained by the expectation effect.[72]

The elimination of confusable phenomena is a very general principle in experimental work. The technology that Chevreul used is blinding, that was discussed in the previous chapter. Blinding is commonly used to eliminate expectation effects. There are many other technologies that can be used to eliminate different types of effects. Hence, physicists perform experiments in a vacuum in order to exclude the effects of air pressure and air movement, and in Faraday cages in order to exclude the effects of external electromagnetic fields.

There are also cases in which confusable phenomena cannot be eliminated. Often, effect separation can then be used instead. This means that an experimental set-up is constructed

[71] James Randi, *Flim-Flam!*, Buffalo 1982, pp. 273-279.
[72] Leonard Zusne and Warren H Jones, *Anomalistic Psychology*, Lawrence Erlbaum, Hillsdale 1982, pp. 249-255.

such that the confusable effects and those of the studied phenomenon can be separately registered.

We can illustrate effect separation with another elegant study of a supposed supernatural phenomenon. Table dancing (also called table levitation), was a popular activity in the middle of the 19th century. In table dancing, a group of persons sit down around a light, round table. Everybody hold their hands around the rim of the table, waiting for it to start moving. After some time (anything between a couple of minutes and several hours) the table is lifted up from the floor although nobody believes her- or himself to contribute to lifting it. Sometimes the table starts to rotate with such velocity that the participants have difficulties in following it.

There are two potential explanations of this. One of them is that the table is moved by some type of mystical power that common physics could not explain. The other is that the participants in the session, perhaps even without knowing it, together exert sufficient physical force on the table to make it lift.

Michael Faraday (1791 – 1867), one of the greatest experimental physicists of his time, decided to find out which of these explanations was the correct one. For this purpose he covered the table with a sheet of paper, that was fastened with a soft, sticky substance. If the table rotated away from the participants by its own force, then the paper would lag behind the table. On the other hand, if the participants (perhaps unconsciously) moved the table with their own muscular power, then the paper would move somewhat ahead of the table.

These experiments showed clearly that the latter was the case. In 1853 Faraday published an article in which he reported his experiments and drew the conclusion that table dancing depended on the unconscious muscular activity of the participants.[73]

An important modern example of effect separation is the use of magnetic fields to separate particles with different charges. If an experiment aims at measuring neutron radiation, it must be distinguished from radiation with charged particles. This can be achieved by letting the radiation pass through a magnetic field, where the charged particles will deviate.

4.4 Controlling the variables

There are often many variables that can influence the phenomenon under study. One example of this is the large number of variables that can influence the outcome of a chemical reaction. In order to optimise the outcome of a reaction we need to know how the reaction rate is influenced by various variables such as the proportions of input chemicals, temperature, pressure, presence of catalysers etc.

The first step towards controlling variables is always to identify those variables that have to be taken into account. In an established experimental area there is usually background knowledge available that indicates what factors need to be considered. Chemical reactions is one of these areas. In new fields, however, such background knowledge is often lacking. Many scientific pioneers have missed important factors that their successors have discovered to be in need of control.

One example of this is a famous plant physiological experiment that was performed in the 17th century by the Belgian Jan Baptista van Helmont (1580 – 1644). In order to find out

[73] Michael Faraday, "Experimental investigation of table turning", *Atheneum*, 801-808, 1853.

where growing plants take their new material from he planted a weighed willow plant in a pot with dried and weighed earth, and then had it watered with rain water. After five years the earth and the willow plant were dried and weighed again. The plant had increased considerably in weight, whereas the earth had only lost a small amount of weight. Since only water had been added, van Helmont concluded that the plant mass had been formed solely by water. Only later was it realized that air could contribute matter to the plant.[74]

When it is not known what factors are important, it is often useful to perform preliminary experiments with extreme values of the factors suspected to have an influence. Suppose that we want to know if variations in lighting, room temperature, and humidity have any influence on an instrument. We can then perform the experiment in darkness and in very strong light, at much lower and much higher temperatures than what is expected in the laboratory, and with and without a humidifier. The factors that have an influence in these high "doses" are selected for being controlled in continued experiments.

Once a factor has been selected to be controlled there are two ways to do this, *constancy* and *controlled variation*. If we want to measure the pulse of a group of experimental persons, we need to control their degrees of physical exertion. This can be done by keeping the level of exertion constant, which is in practice achieved by measuring the rest pulse. Alternatively we can obtain control by varying the degree of exertion, using for instance a test bicycle, and monitoring how the pulse varies with physical exertion. The choice between these two methods depends on the purpose of the experiment. Since the degree of physical exertion has a high potential to influence the result it has to be controlled in one way or the other.

Some experiments aim at investigating the effects of one single variable, for instance the effects of a drug. This is often best done by keeping all other factors constant, and only vary the variable under investigation, in this case the dose of the drug. In animal experiments this is achieved by using animals that are genetically as similar as possible, and treating them as equally as possible with respect to food, care, temperature, etc. The only difference should be that the experimental groups receive the drug, whereas the control group does not.

In a clinical experiment, where the drug is tested on patients, we can of course not keep genetics or food constant. Instead, patients are distributed randomly between groups that receive different treatments. If the number of patients is sufficiently large, the background factors will then be reasonably evenly distributed between the groups. The effects of this are essentially the same as those of constancy, namely that these factors will not have any major influence on the outcome. With respect to the patient groups as a whole, randomisation can be seen as means to achieve constancy.

Controlled variation means that the experiment is performed with several different input values. You can for instance perform a drug test with different doses of the drug, in order to find the dose that yields the best balance between therapeutic effect and side effects. Similarly, a chemical reaction can be investigated at different combinations of pressure and temperature, a crash test can be performed at different speeds and directions of collision, an economic experiment with different prices and other initial conditions, etc.

Some experiments takes a long time to perform. This applies for instance to studies of long term biological effects. In such cases, controlled variation should normally be arranged

[74] D Hershey "Misconceptions about Helmont´s Willow Experiment", *Plant Science Bulletin* 49(3):78-84, 2003.

Documenting experiments

The established way to document experiments is, as already mentioned, to keep a laboratory diary with extensive details about each experiment and its outcomes. Just like patient records, the laboratory diary should be filled in according to well determined rules. In particular, it should not be allowed to make any part of such a diary unreadable by rubbing out or crossing over previous text. It is therefore recommended that the laboratory diary be written in a bound book with numbered pages. All entries should be made with a pen, and changes should be made by striking over the incorrect text and writing the correction beside it. All entries (including corrections) should be dated and signed.

Most experimentalists record carefully the variables that they have changed in the experiment and the outcomes they obtained. However, their notes about the experimental set-up and the background variables are often less complete or may even be missing. Only if these notes are sufficiently detailed can the experiment be reconstructed and reassessed, if it is later put into doubt.

It is advisable to write down the name of the supplier and the batch number of chemicals, and the same applies to serial numbers of measuring instrument. Electrical circuits should be recorded. It is often advisable to make notes of room temperature, air humidity and other external circumstances that might possibly influence the experiment. Generally speaking, the experiment should be documented so well that major sources of error can be identified and assessed and that a repetition of the experiment is possible.

in parallel, i.e. several variants of the experiment are run at the same time. If the experiment only takes a short time, it is often better from practical viewpoint to run the different variants of the experiment one after the other. It is then important to check that the experimental set-up remains unchanged throughout the series of experiments.

When several variables have an impact on the outcome it would in principle be best to try out all possible combinations of these variables. We should for instance register how a chemical reaction proceeds at all possible combinations of pressure, temperature and other factors that can have an influence on the reaction. However, if the number of factors is large it is in practice impossible to investigate all their combinations. Instead, statistical methods should be used to select a number of combinations of the different factors, so chosen that they provide good information about the effects of the various variables and about their interactions.

4.5 Experiments should be repeatable

An experiment aims at determining general patterns, patterns that are present under many different circumstances. Therefore the experiment has to be performed under well-controlled conditions that can be recreated with essentially the same result. A unique event that cannot be repeated cannot fill the function of an experiment in science. Therefore a report on a scientific experiment must give the reader the information that is required to repeat the experiment.

The requirement of repeatability is not unique for scientific experiments. We impose the same requirement on cookbooks and technical manuals. Suppose that this book contained a recipe of a new type of cakes. You and other readers followed these instructions carefully, but only managed to retrieve a heap of scorched crumbs from the oven. Furthermore suppose that you asked me to correct the recipe. You would then probably not be impressed if I said that "this recipe worked when I used it on April 17, 2007, and therefore there is nothing wrong with it". By a recipe we mean an instruction that others can follow and obtain essentially the same outcome. The same argument applies to the instructions in a technical instruction book. If we do not manage to install the new carburettor on the car when following the instruction, it does not convince us if somebody says: “yes, but the manual gives a correct picture of how it was once done”. An instruction is not a description of a particular event that has happened once.

Suppose that we instead try to perform a chemical experiment that is intended to result in a very beautiful, bright blue liquid. When we follow the instructions, we instead obtain a brown liquid. It would not be a good defence of the instructions that they describe what once happened in a laboratory. Just like a recipe and a technical instruction, a description of a scientific experiment should provide sufficient information for others to obtain the same result if they follow the instructions carefully.

Cookbooks and instructions books for professional usage can be difficult to follow for someone who does not have the technical expertise. The same applies to many descriptions of scientific experiments. In all these cases, abbreviated descriptions may be used, but it must always be possible to deduce from them what one has to do in practice in order to obtain the desired outcome. Most descriptions of scientific experiments differ from food recipes and instructions books in their linguistic form, and this in a rather unfortunate way. The cookbook and the instruction book use the imperative tense: “Let it boil in twenty minutes”. In scientific articles, the past tense is used instead: “The object was immersed in water at 100°C, during 1 200 seconds.” But in spite of the past tense experimental reports, just like recipes, are intended to describe a general method that others can use to obtain essentially the same result.

4.6 Repetition in practice

It has sometimes been claimed that the demand of repeatability is not applied in practice, since only few scientific experiments are actually repeated. True, scientists are not very fond of replicating investigations that have already been performed by others. It is much more rewarding to search for new results than to confirm old ones. Therefore, exact replications of experiments are exceptions. They usually take place only when it is believed that something may have been wrong in the original experiment, for instance some uncontrolled variable or even fraud or other misdemeanour. Most experiments are performed only once.

It could then seem as if the requirement of repeatability is not applied in practice. But there are at least four reasons why it is applied to a much higher degree than what can be seen from the number of published repetitions of experiments.

First, the requirement of repeatability does not mean that scientific experiments actually have to be repeated, only that they should be possible to repeat with essentially the same result. Competent experimentalists usually know what conditions they have to specify in order to make an experimental report determinate enough for repeatability.

Secondly, many repetitions are performed but not published. If a repetition yields a different result than the original experiment, it is much more plausible that it will be published.

Thirdly, many new experiments are extensions or modifications of old ones. Hence, a chemist can adopt a method of synthesis that has already been developed by a colleague, but modify it to produce another, related molecule. If this fails, she may return to the original

The journal *Organic Synthesis*

All scientific journals of importance have referees that assess the articles before they are published. The journal *Organic Synthesis* goes one step further. The journal publishes methods of chemical synthesis. Before a method can be published it has to be reproduced by another researcher (a member of the board of editors). The methods are also described in more in detail than in most other chemical journals.[75]

The reason for this unusual practice is that is difficult to describe a complicated organic synthesis so precisely that the description can serve as a guide for repeating the synthesis. To obtain repeatability an extra control has therefore been introduced, that is not common in other areas.

experiment, and try to discover possible errors in the published description. Generally speaking, it is much more probable that an experiment will be repeated in a modified version than in the original version.[76]

Fourthly, there are some types of experiments that are often repeated. This applies to treatment experiments in medicine. A new treatment method is not introduced into routine clinical practice on the basis of a single clinical trial that speaks in its favour. It usually takes several well conducted clinical studies before the clinical expertise is fully convinced. In all areas of science, experiments that are regarded as unusually important tend to be repeated. One example of this is Aspect's experiment that provided decisive information about the validity of quantum mechanics. It has been repeated – and confirmed – several times. Another example is "cold fusion", fusion energy on a small scale at room temperature. In March 1989 two American researchers announced that they had been able to achieve cold fusion in their laboratory. This lead to intense activities in laboratories around the world, and hundreds of attempts were made to repeat the experiment. It soon turned out that cold fusion, as described in the original report, does not work. The seemingly positive outcome that was first reported resulted from deficiencies in the experimental technique, in particular in the measurement.[77]

In addition to all this, experimental outcomes often receive indirect support from other experiments that confirm their underlying principles. One example of this is Mendel's original experiment with peas. In this experiment he found out how properties are inherited through sexual reproduction, thereby laying the foundations of modern genetics. Few if any geneticists have been interested in repeating Mendel's experiment. However, thousands of them have performed other experiments showing how inherited properties are transferred to the next generation in various species. Even if these investigations are not repetitions of Mendel's experiment, they provide an important although indirect support of the experiment by confirming the same pattern of inheritance.[78]

Nevertheless it is important to keep the discussion about repeatability alive, and in particular to study new experiments from that point of view. When a reported investigation confirms what we already believe or want to believe, we all tend to take the demand of repeatability too lightly. But in all probability the number of incorrect (non-repeatable)

---

[75] www.orgsyn.org

[76] David T Lykken, "Statistical Significance in Psychological Research", pp. 269-279 in DE Morrison and RE Henkel, *The Significance Test Controversy*, 1970. Michael Mulkay and G Nigel Gilberg "Replication and Mere Replication", *Philosophy of the Social Sciences* 16:21-37, 1986.

[77] Gerd Graßhoff and Michael Schneegans, "Experimentation and Methodology, Applied to Cold Fusion Research" *Philosophia Naturalis* 32:47-70, 1995.

[78] Peter Urbach, "On the utility of repeating the 'same' experiment", *Australasian Journal of Philosophy* 59:151-162, 1981.

experimental reports is higher, even in the best scientific journals than the number of non-serviceable recipes in the cookbooks that we keep in our kitchens.

5 Proving connections

As was pointed out in the previous chapter, experiments and other scientific observations are not performed just in order to register single events. Instead, the purpose is to find more general connections between events and phenomena. One possibility could be to start out from simple and isolated theory-independent observations. Generalizations could then be drawn according to the principle "in all the cases that I have observed it is like this, so this is how it is generally". However, it is difficult to find any clear proponent of such hypotheses-free science. Successful researchers do not make their observations randomly, but plan carefully what to look for. Most of those who have theorized about science have also been well aware of this.

The most discussed form of empirical investigation is hypothesis testing. Here the starting point is an hypothesis. Experiments and other observations are planned so that they can confirm or falsify this hypothesis. Hypothesis testing has been accounted for more or less clearly by thinkers far back in history. The English philosopher William Whewell (1794 – 1866) developed it in detail in a book published in 1840.[79] This mode of working is usually called the "hypothetic-deductive method", since it consists in deducing consequences from the hypothesis, and then testing these consequences.

5.1 Testing hypotheses

It is easy to be wise after the event. Consider the following two examples:

> *Case 1*: Andres and Beata bring their cat to their summer house for the first time. The cat changes its behaviour, becoming wilder and hunting more. After a week Andres says to Beata: "This is no surprise. The explanation is simple. The hunting instinct of cats is weakened when they adapt to people, but this weakening only has effect at the place where the adaptation took place."
>
> *Case 2*: Already before Andres and Beata go to the summer house with the cat Andres says: "The cat is surely going to hunt more than here. The hunting instinct of cats is weakened when they adapt to people, but this weakening only has effect at the place where the adaptation took place." When they arrive to the summer house the cat becomes wilder and hunts much more than what it did before.

In case 1, when Andres made his proposal after the event, his general statement about the hunting instinct of cats had much less credibility than in case 2 in which he predicted what was going to happen. (Of course this does not mean that the explanation was correct in case 2. There are alternative explanations, such as that the surroundings gave better opportunities to hunt.)

[79] Medawar, *Pluto's Republic*, p. 101.

Einstein's attitude to empirical research

Albert Einstein put much effort into promoting empirical tests of general relativity. This applied not least to the prediction that starlight passing very close to the sun would be deflected somewhat due to the gravitation of the sun. Einstein was strongly involved in plans to make observations during a solar eclipse in the year 1914, but these plans could not be realized due to the outbreak of the first World War.

In Einstein's view, the general relativity theory would stand or fall with the observations that could be made of the deflection of light in the gravity field of the sun. In 1913 he wrote to the astronomer Erwin Finlay Freundlich: "Nothing can be done here through theoretical means. Next year, you astronomers can do a clearly invaluable service to theoretical physics in this regard. We will get reliable information on whether it is correct to continue generalize the relativity principle or whether we must stop at the first step." (With the first step he referred to special relativity theory.)

During the solar eclipse in 1919 Arthur Stanley Eddington made measurements that confirmed Einstein's prediction. This was a decisive confirmation of general relativity. When Einstein was informed of these results, he wrote several letters about the findings, including a postcard to his mother that began with the words "Today good news". It has often been claimed that Einstein was uninterested in these observations, and paid little attention to Eddington's measurements since he already knew that his theory was correct. This is a myth with no credibility that promotes a seriously wrong picture of how the best theoretical physics relates to empirical research.[80]

It is similar in science. Hypothesis testing means that an hypothesis is put to a test, the outcome of which was not known when the hypothesis was put forward. The testing (and confirmation) of Einstein's general relativity theory is a classic example of this. It follows from this theory that the observed position of a star would change somewhat when the light from that star passes very close to the sun. During a solar eclipse in 1919 measurements were made that confirmed this prediction. This was generally seen as a strong confirmation of the theory. Probably this would have been seen differently if Einstein had put forward his theory only after these measurements hade been made. The theory could then have been accused of being "ad hoc", i.e. adjusted in order to confirm observations that had already been made.

It is of course not wrong in itself to adjust or supplement a theory after it has been put forward. However, such a change or addition should do something more than just correct known deviations between the original theory and empirical observations. It should also provide new testable predictions. Otherwise the adjustment is regarded as ad hoc. Such adjustments weaken a theory.[81]

Einstein predicted not only that starlight would be deflected, but also the size of this deflection (namely 1.745 arc seconds). The measurements confirmed this prediction (within the expected margin of error).[82] The confirmation of general relativity that was obtained through the solar eclipse observations in 1919 was much strengthened by the fact that it was a rather precise prediction that was confirmed.

During new solar eclipses the same types of observations have been made as in 1919. These additional observations have again confirmed the predictions of general relativity

[80] Klaus Hentschel, "Einstein's attitude towards experiments: Testing relativity theory 1907-1927", *Studies in the History and Philosophy of Science* 23:593-624, 1992. Klaus Hentschel, "Das Märchen vom Zauberer im weißen Kittel", *Phys. Unserer Zeit* 34:225-231, 2003.

[81] Adolf Grunbaum, "Ad Hoc Auxiliary Hypotheses and Falsificationism", *British Journal for the Philosophy of Science* 27:329-362, 1976.

[82] JT Davies, *The Scientific Approach*, 1965, p 28-29.

theory, but these confirmations are not considered to be particularly important. Einstein's theory also gave rise to other predictions that concern quite different natural phenomena than the deflection of starlight. When such predictions have been confirmed, this is considered to be a much stronger confirmation than new data from solar eclipses. Confirmation is stronger if it is based on tests that are independent of each other in the sense of concerning different types of natural phenomena.

This example exhibits two generally accepted principles for assessing confirmations of a theory or hypothesis:

1 The confirmation is stronger, the more exact the confirmed prediction is.
2 If there are several confirmations, their joint force increases the more independent they are of each other.

The second of these principles can be contrasted with the requirement that experiments should be repeatable. The support of an hypothesis is stronger if we have confirmations from different types of experiments than if all confirmations come from repetitions of one and the same experiment.[83] This is often an important reason to test the conclusions of an experiment in an experiment of another type, rather than repeating the original experiment. (If the new test yields a result that does not confirm our conclusions from the first experiment, then a repetition in the proper sense will be much more interesting to carry through.)

A reservation is in place concerning the principle that tests are valid if they have been performed after the hypothesis was put forward. Sometimes a theory can explain other phenomena than those for which it was originally developed. This is usually taken to speak strongly in favour of the theory in question. Sometimes confirmations from another field are treated as equally important as confirmations that are obtained after the hypothesis was put forward. This too can be illustrated with general relativity theory. The planet Mercury has minor irregularities in its orbit (perihelion precession) that were known in the 19th century but could not be explained until Einstein's theory of relativity was available. This has often been seen as a major confirmation of general relativity.[84] A plausible argument for this view is that Einstein probably did not consider the orbit of Mercury when developing his theory, and at any rate he did not adjust the theory to suit previous observations of the planet.

Cases like this are unusual. The vast majority of scientific hypotheses and theories have a rather limited area of application, and those who proposed them were well aware of these limitations. In such cases a confirmation based on previously known observations could not be put on par with a test that was been performed after the hypothesis was put forward.

[83] Louis Boon, "Repeated Tests and Repeated Testing: How to corroborate low level hypotheses", *Zeitschrift für allgemeine Wissenschaftstheorie* 10:1-10, 1979.
[84] Elie Zahar, "Why Did Einstein's Programme Supersede Lorentz's?", *British Journal for the Philosophy of Science* 24: 95-123, 1973, p. 101.

Positivism

The concept of positivism is usually connected with the French philosopher and sociologist Auguste Comte (1798 – 1857). He claimed that knowledge and science should begin with "the positive", i.e. that which is given by experience.

Logical positivism (logical empiricism) denotes a philosophical school that arose in the so-called Vienna Circle in the 1920´s. A basic idea among these philosophers was that only sentences whose contents can be tested against sensory experiences can tell us anything about the world.

Positivism is today a highly ambiguous concept. Positivism can be said to include the following four standpoints that are not always in accordance with each other.

1. Science is the only acceptable form of human knowledge. (Scientism, or antimetaphysics.)
2. Science should only deal with observables. All theoretical concepts should refer directly or indirectly to observables.
3. Science should provide political directives, since political issues can and should be settled by science.
4. Science is value-free.[85]

5.2 Verification or falsification?

When an hypothesis is tested, it can be either verified (shown to be true) or falsified (shown to be false). In order to be useful in science an hypothesis has to be testable, i.e. possible to verify or falsify. But which is most important, verification or falsification? In the 20th century two schools of thought emerged, one of which gave absolute preference to verification and the other to falsification.

The first of these schools was, logical empiricism, also called logical positivism or the Vienna school. Its proponents claimed that a sentence cannot be at all meaningful unless it is possible to verify. They even identified the meaning of a sentence with its method of verification. ("The meaning of a proposition is the method of its verification.") Science should only deal with observables, i.e. things that can be observed. All theoretical terms refer, directly or indirectly, to observables, and theories should be logically derivable from empirical knowledge. (This is why the school was called "logical empiricism".)

The logical empiricists made important, lasting contributions through their criticism of other schools in the philosophy of science. They removed excessive metaphysics and increased the requirements of stringency. But they were not able to show how their own demands on scientific work could be realized in practice. In particular, they did not manage to show how theories could be verified in practice through logical conclusions based on empirical observations.

The other school was founded by Karl Popper (1902 – 1994). He proposed that we should stop trying to verify theories or hypotheses. His major argument for this was that theories and hypotheses consist of universal statements, and such statements, he maintained, are not possible to verify. Consider for instance the hypothesis "all swans are white". It is not possible to confirm this statement in a logical sense. It does not matter how many swans we have seen that are all white, there may still be swans of other colours that we have not seen. On the other hand, this hypotheses can be falsified. It is sufficient to document one single swan that is not white in order to reject the hypothesis that all swans are white. Therefore, he

[85] Russell Keat, *The politics of social theory*, pp. 16-18.

Falsification in practice?

In the year 2000 70 scientific articles were published as "articles" in the journal Nature. This is one of the most prestigious forms of publication in science. (Most articles in Nature are published as "letters", which is a less prestigious form of publication.)

In only 17 of these 70 articles did the authors state a scientific hypothesis that was tested in the research reported in the article. Most of the other articles reported exploratory empirical research, i.e. investigations that are open-ended rather than based on hypotheses that have been put forward beforehand. Many of these articles reported discoveries of gene sequences, molecular structures, or biochemical reaction mechanisms.

Of the 17 hypothesis-testing articles, 13 resulted in verification of the hypothesis and 3 in its falsification, whereas the result was inconclusive in the remaining case.[86] This investigation puts in question the conception of good science as science that falsifies hypotheses.

(Later, one of the 70 articles was revealed to be a case of scientific misconduct. This was one of the 13 articles in which the hypothesis had allegedly been verified.[87])

said, scientific activities should be so organized that we try to falsify hypotheses, not to verify them.

It should be mentioned for completeness that Popper had a predecessor whom he was probably not aware of. In 1868 the zoologist August Weisman (1834 - 1914) wrote that "although a scientific hypothesis can never be proved, if it is false it can be falsified".[88] However, Popper was the first to develop this thought into a more complete theory of science.

Popper's work highlights the important insight that a theory can only be useful in science if it is in accessible to disproof in the case that it is wrong. If a theory is compatible with whatever observations you make, then it is empty from the viewpoint of empirical research. (One example is astrology, that is based on vague statements that can be reinterpreted in multiple ways, and contains no mechanism for correction or disproof.)

5.3 The problems of falsification

Examples of scientific hypotheses are easily found that contradict Popper's theory by being easier to verify than to falsify. One example of this is the toxicological hypothesis "Nitro methane is carcinogenic". (Or more precisely: "There is a dose and a route of entry at which nitro methane increases the frequency of cancer in humans.") According to Popper, if this is a proper scientific hypothesis, then it should be falsifiable but not verifiable. In other words, it should not be possible to show that nitro methane is carcinogenic (if this is true), but it should be possible to disprove the same statement (if it is false).

In actual fact it is the other way around. It can never be shown with empirical investigations that a substance does not cause cancer. The reason for this is that no empirical investigation can exclude the possibility that an exposure gives rise to an increase in cancer frequency that is too small to be distinguished from random variations. On the other hand, it can in many cases be verified that a substance causes cancer. This will be the case if a

[86] Sven Ove Hansson, "Falsificationism Falsified", *Foundations of Science* 11:275-286, 2006.
[87] Schön, J. H., Kloc, C. & Batlogg, B. Superconductivity at 52K in hole-doped C_{60}. *Nature* 408:549-552, 2000. Barbara Goss Levi, "Investigations Find that One Lucent Physicist Engaged in Scientific Misconduct", *Physics Today* 55(Nov):15-17, 2002.
[88] Quoted in Franz Stuhlhofer, "August Weismann – ein 'Vorläufer Poppers", Conceptus 50:99-100, 1986, p. 99.

sufficiently large number of people have been exposed to dose levels that give rise to a large increase in cancer frequency.[89]

A Popperian can of course respond to this example by claiming that all scientific hypotheses should have the form "All *x* are *y*". The hypothesis "Nitro methane is carcinogenic" does not have this form. This, however, is an attempt to mould science in a form that it is not suited for. Science is devoted to general statements, but general statements can have many other logical forms than "All *x* are *y*".

More generally speaking, a scientific issue is often concerned with the choice between two statements that are each other's negations. (In this case: "Nitro methane is carcinogenic" and "Nitro methane is not carcinogenic".) To verify one of these statements is synonymous with falsifying the other. Often but not always there is *verification asymmetry*, which means that one of the two statements is for practical reasons easier to verify than the other. There is also often *individuation asymmetry*, which means that one of the statements is more suitable than the other as a scientific hypothesis. It is for instance, according to prevailing views in the field, more natural to regard the statement "nitro methane is carcinogenic" as a scientific hypotheses than to give that role to its negation, "nitro methane is not carcinogenic". This is not a mere matter of terminology, but a substantial matter of where to put the burden of evidence. As was pointed out in chapter 1, the burden of proof rests with those who make a new specific statement, i.e. those who put forward an hypothesis. The distinction between an hypothesis and its negation is well entrenched in scientific praxis, and it cannot be removed or reversed without giving up the conventional meaning of the term "hypothesis".

With this terminology, Popper's view can be summarized in the following three conditions:[90]

1 that the verification asymmetry is always very large,
2 that there is an individuation asymmetry, and
3 that the two asymmetries are always related to each other in such way that what we count as the hypothesis is more difficult to verify than its negation.

It is easy to show that none of these three conditions applies to scientific practice in general. Suppose for instance that a new fish has been found in an African lake. For some reason we want to find out whether or not it is a livebearer. There is no verification asymmetry, since it is about equally easy to verify that a fish species is live-bearing as that it is egg-laying.

The same example can also be used as a counterexample to individuation asymmetry. We may treat either "the fish is live-bearing" or "the fish is egg-laying" as an hypothesis to be tested. This will depend on our background knowledge and the theoretical context. Perhaps we have found an anatomical trait in the fish that has previously only been found in live-bearing species. This would give us reason to set up the hypothesis that the new species is a livebearer. Or we can have some reason to set up the opposite hypothesis. It is important to note that we cannot determine from the linguistic form of the two opposite statements which of them is the hypothesis. This depends on what is already known and what we try to find out.

---

[89] Sven Ove Hansson, "Can we reverse the burden of proof?", *Toxicology Letters* 90:223-228, 1997.
[90] Sven Ove Hansson, "Falsificationism Falsified", *Foundations of Science*, 11:275-286, 2006.

We have already seen an example in which the third condition was not satisfied. In the case “Nitro methane is carcinogenic” there is both verification asymmetry and individuation asymmetry, but the two asymmetries are coupled to each other in the opposite way to what Popper stipulated. Generally speaking there is no simple connection between the two asymmetries. Some scientific hypothesis are about as accessible to verification as to falsification, others are more accessible to verification and yet others more accessible to falsification.

5.4 The necessary weighing of evidence

As we have now seen, the empirical testing of hypotheses cannot be reduced to a matter of falsification. It would be equally misleading to reduce it to a matter of verification. In actual science we have to search for evidence both for and against the hypotheses under study. When assessing the hypothesis, the evidence speaking in its favour should be weighed against that which speaks against it. Doing otherwise would be just as unreasonable in science as it would be in a court of law to consider only the arguments speaking in favour of the defendant, alternatively only those that speak against her.

There is a further complication: When an empirical observation yields results that contradict a theory, it is not necessarily the theory that is wrong. Another possibility is that the observation was incorrectly made, or incorrectly interpreted.

In the history of science literature many examples are given of experiments that have disproved established theories. Many of the most important experiments have been of this kind. But it would be wrong to believe that this is what always happens when theory and observations do not tally. At least as often it turns out that something was wrong with the experiment or with its interpretation. Two examples can illustrate this.

In 1906 the famous physicist Walter Kaufmann (1871 – 1947) reported experimental evidence indicating that special relativity theory was wrong. Only ten years later was it shown that his experiment had failed.[91] A leak in the vacuum system had led to a misleading outcome. Hence, in this case it was the experiment and not the theory that was in need of correction.

The other example is much older. Tycho Brahe predicted that if Copernicus was right, and the earth rotates around the sun, then the observed angle between a star that is close to us and a distant star would change as the earth moves around the sun. This change in angles is called the parallax. Brahe made measurements to test the hypothesis, but he found no such effect. This in his view spoke against Copernicus’ theory. We now know that the stars are too far away from us for any parallax to be discovered with the instruments that Brahe had access to.[92] It was not until 1838 that another astronomer, the German Friedrich Bessel, managed to measure the parallax and use it to determine the distance to a close star. There was nothing wrong with Tycho Brahe’s measurements, to the contrary he made observations with

[91] Gerald Holton, The Scientific Imagination, Case Studies. Cambridge, Harvard UP, 1978, pp. 98 and 326-327. Mario Bunge, Treatise on Basic Philosophy, vol 8, p. 139. Gerald Holton, Thematic Origins of Scientific Thought, Harvard UP 1973, essay 9.

[92] Pierre Oléron, Le Raisonnement, Que sais-je, p. 120-121.

admirable precision. In this case it was the interpretation of the observations that had to be corrected.

Many theorists of science have paid much attention to so-called crucial experiments. An experiment is crucial if it alone tells us whether a theory should be accepted or rejected. But in practice a single experiment is seldom sufficient to decide an important theoretical issue.[93] The normal procedure is to search for evidence in both directions and weigh together all the available evidence. Often it is only much later that it can be concluded in hindsight that a single experiment was decisive for the choice of a scientific theory.

5.5 Simplicity

There are often several hypotheses or theories that concord with the available empirical evidence. There is no simple rule for choosing between theories that have equal empirical support, but the following maxim is often of great help: choose the simplest of the theories or hypotheses that are most compatible with the empirical evidence.

An hypothetical example can illustrate this: It would in principle be possible to replace the theory of gravitation by a theory in which there are two forces, one attracting and one repelling force. It would not be very difficult to provide such a theory with a mathematical model that predicts exactly the same observations as those predicted by common gravitation theory. But such a two-force theory would be pointless unless it was possible to isolate the two forces from each other in some kind of empirical observation. The two-force theory should be rejected because it contains more theoretical concepts than we need.

The requirement of simplicity is historically connected to William Occam (ca 1285 ca 1349), who contributed significantly to its formulation. It is often called “Occam’s razor”, and its classical form is the sentence “You should not unnecessarily multiply the entities”. (In Latin: *Entia praeter necessitatem non esse multiplicanda*.) As can be seen from this formulation, the simplicity strived for does not have much to do with comprehensibility or ease of understanding. It is not required that hypotheses be easily understood but that they should contain no dispensable theoretical components.

The requirement of simplicity can also be expressed in another way: Use as little theory as possible to account for as much empirical evidence as possible. No theoretical elements should be introduced unless empirical evidence shows that they are needed. Furthermore, theoretical concepts should be discarded whenever it turns out that we do not need them any more. Many important discoveries have led to a decrease in the number of basic theoretical concepts. Hence Newton’s theory of gravitation reduced the number of theoretical concepts needed to describe the movements of bodies. Before Newton there had been two sets of laws of motion, one for bodies on the earth’s surface (terrestrial mechanics) and another for heavenly bodies (celestial mechanics). Newton created a unified theory for these seemingly quite different forms of movement. More empirical evidence could be explained with fewer theoretical concepts.

The requirement of simplicity can be derived from the requirement that science should provide us with intersubjective knowledge, knowledge that is common to us all. Theoretical assumptions that are not forced upon us by the facts can be chosen differently by different

[93] Klaus Hentschel, “Das Märchen vom Zauberer im weißen Kittel”, *Phys. Unserer Zeit* 34:225-231, 2003.

persons. In order to achieve intersubjectivity we should refrain from making assumptions that have no empirical justification. It is also often easier to find the complexity of the real world if you start out from a too simple model than the other way around.

The simplicity requirement is a research strategy. It is not a postulate about reality, and it should not be interpreted as an assumption that reality has to be simple. We do not have a basis for such an assumption. It is only in the choice between theories that describe reality equally well that Occam's razor should be used.

There are many examples of how empirical evidence has led science to more complex theories, rather than simpler ones. Boyle's law for the pressure, temperature, and volume of gases ($PV = RT$) was replaced by another formula that was more complex ($(P + a/V^2)(V - b) = RT$).[94] Einstein's theory of gravitation is more complex than Newton's. These, however, are complications that have been forced upon us by observations. The introduction of such complications does not contradict the requirement of simplicity.

Clearly, the requirement of simplicity can never replace observations and experiments. In the choice between two hypotheses or theories there is not much to be gained by arguing about which is the simplest one. The crucial issue is always if and in that case how observations can be made that adjudicate between the rival hypotheses.

5.6 Randomness

When testing hypotheses we often have reason to ask the question: Are the patterns we observe "just" random ones, or do they reveal some real correlation? It is often difficult to determine this, and our intuitive judgement of randomness are quite unreliable.

The so-called birthday paradox a classic example of this. Suppose that 15 persons tell each other on what day of the year they were born, and it turns out that two of them have the same birthday. Most of us would consider this to be quite improbable, but in fact the probability of this happening is as high as 1 in 4. If the number of persons is 23, the probability that at least two of them have the same birthday is just over 1 in 2.

Another example is the so-called straight line phenomenon. In the early 20th century a group of British amateur archaeologists were surprised when they discovered how often three, four or even more archaeological sites turned out to lie on the same straight line on a map. One of them, Alfred Watkins, published a book about this in 1925.[95] In 1958 the Frenchman Aimé Michel published a book in which he claimed that places where UFOs had been observed are concentrated on straight lines on the map.[96]

On maps with marks indicating for instance archaeological sites or UFO sites it is often possible to draw an impressive number of straight lines connecting these points. However, no scientific conclusions follow from this unless we can exclude a random phenomenon (or a bias in the choice of points). It turns out that many more straight lines can be drawn between randomly placed points on a map than what might be expected. On a map containing 330 archaeological sites you can for instance expect to find about 1 100 straight lines that pass

---

[94] Harré, *Philosophies of Science*, p. 45.
[95] EC Krupp, "Observatories of the gods and other astronomical fantasies", pp. 219-256 in EC Krupp (ed.) *In Search of Ancient Astronomies*, Penguin 1984.
[96] Aimé Michel, Flying saucers and the straight-line mystery, S.G. Phillips, New York 1958.

through at least four of these sites.[97] (Of course, the ancients did not locate their buildings at random, but according to the terrain, access to water, etc. The probability of straight lines between archaeological sites would hardly diminish due to these selection mechanisms.) Since our statistical intuitions are so unreliable, it is important to be on guard against over-interpretation of patterns that may have arisen due to random effects. The best way to do this is to test hypotheses statistically.

5.7 Statistical hypothesis testing

Suppose that a car manufacturer decides to make a small change in one of their models. They decide to fasten the inner mirror on the windshield. It was previously attached to the ceiling. Some of the engineers fear that the new construction will make the windshield less impact-resistant. In order to find out if this is so, they carry out an experiment in which 30 windshields without a mirror (the control group) and 30 windshields with a mirror (the experimental group) are subjected to a standardized impact test. It turns out that 7 of the windshields with the mirror break, but only 5 of the windshields without a mirror.

This outcome does not allow the conclusion that the mirror weakens the windshield. The difference between the two groups, is so small that it could just as well be a chance effect. However, if 20 of the windshields with a mirror had broken but only 5 of the control windshields, then the experiment would obviously have supported the hypothesis that the windshield is weakened by the mirror.

But how should we reason if 10 of the windshields with a mirror had broken? Or 15? Obviously our intuitions are not sufficient here. We need some type of rule, some method to determine whether chance effects are a reasonable explanation of the differences between the control group and the experimental group. It is only if chance effects can be excluded as a reasonable explanation of the difference that the experiment can be considered to confirm the hypothesis that the mirror weakens the windshield.

For this purpose a zero hypothesis is used. This is the alternative hypothesis that the pre-test difference between the groups (here the mirror) had no impact on the outcome variable (here the frequency of broken windshields). The crucial issue is: Under the assumption that the zero hypothesis is true, how probable are the results obtained?

In our example the zero hypothesis says that the mirror has no effect. If this is true, then the probability is 37 percent to obtain a difference between the two groups at least as large as that between 5 and 7. The probability is 12 percent to obtain at least a difference such as that between 5 and 10. It is 1 percent to obtain one at least as big that between 5 and 15, and 0.01 percent to obtain one at least as big that between 5 and 20.

The next question is of course where to draw the limit. If the observed outcome is sufficiently improbable given the zero hypothesis, then we can reject the zero hypothesis. But how improbable does it have to be? There is a strong tradition, common to a large variety of scientific disciplines, to set this limit at 5%. Results that pass this limit are called *statistically significant*.

[97] EC Krupp, "Observatories of the gods and other astronomical fantasies", pp. 219-256 in EC Krupp (ed.) *In Search of Ancient Astronomies*, Penguin 1984, p. 236.

How can scientists accept to be wrong so often?

The conventional limit for statistical significance, 0.05, is often incorrectly interpreted as “the probability that the result is wrong”. This gives rise to the question: How can scientists accept to be wrong so often?

However, the significance level is not the probability of a false positive. Instead, it is the probability that the outcome obtained would have been obtained if the null hypothesis were true. The probability of a false positive is usually much lower.

When we test an hypothesis, we usually have good reasons to believe that this hypothesis is true and that the null hypothesis is wrong. Suppose that in a particular area of science, nine out of ten tested hypotheses are true. Then a false positive is impossible except in the ten percent of the cases when the null hypothesis is correct. If we apply the 0.05 limit for statistical significance, then in the long run one out of 20 of these will be a false positive. Therefore, the frequency of false positives will be 1 in 200, not 1 in 20.

More generally speaking, 0.05 serves as an upper limit of the frequency of false positives. Only if the null hypothesis is true in all tests that are performed, will the long-run frequency of false positives be 1 in 20. This is the reason why the significance limit 0.05 is compatible with the high standards of accuracy in scientific research. It is important to note, however, that this only holds as long as most of the tested hypotheses are true.[98]

Hence, a result is statistically significant if the probability is low (below 5%) that an effect at least as big as the one observed would come about by pure chance. Another way to express this is that the probability would be less than 5% to obtain the effect in a comparison between two control groups that received exactly the same treatment. If the observed difference between an experimental group and the control group would have appeared in at least 1 out of 20 experiments with two control groups, then the result is not considered to be a significant. (It is important to observe that the level of statistical significance does not say anything about how probable it is that the effect depends on chance. See the fact box.)

In our case this means that the outcomes with 7 or 10 windshields in the experimental group would not be significant, whereas 15 or 20 broken windshields would have been a significant result (assuming in all cases that 5 of the control windshields broke).

It is important not to confuse statistical significance with scientific credibility. A statistically significant result has little value if the investigation was badly performed. A result can be a highly significant and yet highly implausible. A statistically significant result in a single experiment that contradicts the laws of gravitation is not a sufficient reason to abandon these laws. The new evidence has to be weighed against all the previously collected evidence in favour of them.

It is also important to observe that tests of statistical significance are intended for testing *previously stated* hypotheses. It is a common but nevertheless serious mistake to use these tests in a search by all means for statistically significant results. If you perform a large enough number of statistical tests on a large material you can be almost sure to find correlations that are statistically significant but are nevertheless just chance effects. The risk of over-interpreta-

[98] Maxwell, C. “Clinical trials, reviews, and the journal of negative results”, *British Journal of Clinical Pharmacology* 1:15-18, 1981. Fisher, R. A. *Statistical Methods and Scientific Inference*, Oliver and Boyd, Edinburgh, 1956, pp. 41-42.

tion of such correlations is a serious problem in all scientific work that is based on complex data.

One example of this is the statistical evaluations that the sociologist of music Alan Lomax made of large tables in which he had collected information about the musical and social characteristics of many different cultures. In 1972 he and a co-worker published an article in Science where they reported a correlation between high production of milk and energetic singing. They wrote that "this extra source of protein accounts for many cases of energetic vocalizing"[99]. In actual fact they had looked for so many correlations that there is no need for any other explanation of their findings than random variations in the data.

This phenomenon is usually called *mass correlation*. In Lomax's case it was easy to discover what had happened, since he reported what correlations he had investigated. It can be much more difficult to discover mass correlation if a researcher only publishes those parts of her material that yield significant results. In order to avoid mass correlation and other inappropriate uses of statistical tests, good research planning is needed. Only those statistical tests that have been selected before the collection of data should be counted as hypothesis-testing. If other tests are performed, their outcomes should instead be reported as the basis for new hypothesis for further testing.

5.8 Research is not always hypothesis-testing

Hypothesis testing has an important role in science, and it has sometimes been claimed that all good science has to be hypothesis-testing. This, however, is not a realistic conception of science. As we saw above, a large part of the best research has no hypothesis. You do not need an hypothesis to determine a DNA sequence or the structure of a protein. You do not either need any hypotheses to determine whether a particular policy measure increases or decreases unemployment, or does not influence it at all.

All good research must have a clear research issue, something that the investigation aims at finding out. In some cases this issue is an hypothesis that can be falsified or verified. This means that the possible outcomes have been divided in advance into two groups, outcomes that verify respectively falsify the hypothesis. In other cases it is not meaningful to make such a classification in advance of the possible outcomes. Such research can be called explorative.[100] Explorative research can often lead to formulations of hypotheses that can be tested in subsequent investigations. Science proceeds by a combination of explorative and hypothesis-testing investigations. Which type of investigation one should perform depends on the state of knowledge in the particular issue under investigation.

---

[99] Alan Lomax och Norman Berkowitz "The Evolutionary Taxonomy of Culture" *Science* 177(4045):228-239, 1972, p. 232.
[100] Edmond A Murphy, "The analysis and interpretation of experiments: some philosophical issues", *Journal of Medicine and Philosophy* 7:307-325, 1982. I.J. Good, "The Philosophy of Exploratory Data Analysis", *Philosophy of Science* 50:283-295, 1983. Stanley A Mulaik, "Exploratory Statistics and Empiricism", *Philosophy of Science* 52:410-430, 1985. Friedrich Steinle, "Entering New Fields: Exploratory Uses of Experimentation", *Philosophy of Science* 64:S65-S74, 1997. Friedrich Steinle, "Experiments in History and Philosophy of Science", *Perspectives on Science* 10:408-432, 2002.

6 Using models

In the previous chapter we studied how scientific theories and hypotheses can be tested. We are now going to look more closely at the contents of such theories and hypotheses. This chapter is devoted to scientific models and the two following chapters to explanations and causes.

6.1 Three types of models

Theories and hypotheses often contain models, i.e. physical or mental systems that have been constructed to reflect important properties of the phenomena under study. There are many types of models, but they all have some sort of structural similarity with that which they represent.

Three types of models are commonly distinguished between.[101] The simplest form are *iconic models* (pictures), enlarged or diminished renderings or projections of the object. Some examples are maps, anatomical models, molecular models, and architectural models. The structural similarity between an iconic model and its object is primarily spatial, and includes geometrical form.

Iconic models are often used to gain overview over objects that are complex or otherwise difficult to visualize. This is why chemists use molecular models, either in physical and computerized forms. This usage of iconic models prevails not only in science but also in various practical activities. Both engineers and architects make ample use of drawings and often also of three-dimensional models.

A scale model is an iconic model that is proportionately reduced or magnified in size. Scale models are widely used in technical applications. In order to investigate the aerodynamic properties of an airplane, a reduced model of the plane is placed in a wind tunnel where it is exposed to winds with different velocities and directions. Similarly scale models of ships have been used since very long to test their seaworthiness.

Analogue models have some non-spatial structure in common with the object. Some examples: Atoms can be described as planet systems on the micro level. Light can be described as waves of the type that we know from water surfaces. The programmer can describe the function of a computer program with a flow diagram. Some types of models can be used to represent widely divergent phenomena. Hydraulic systems can be used as models of electrical systems. Towards the end of the 1950's several hydraulic models were constructed to represent the economy; the flow of fluid represented flows of money.[102]

In *symbolic models*, different parts or aspects of the object under study are represented by symbols. The symbols on the control panel of the car are examples of this. However the symbolic models used in science are almost exclusively of another type: they are mathematical models. The variables in the formulas of the mathematical model correspond to

[101] Russell Ackoff, *Vetenskaplig metod*, 1972, p. 91-92.
[102] Herbert A Simon, *The Sciences of the Artificial*, third edition, MIT Press, p. 14.

different quantities in that which the model represents. The relations between different phenomena are expressed in mathematical equations.

Mathematical models have been indispensable since the 19th century in physics, chemistry and technology. Economic science was mathematized in the 20th century, and in the same century the role of mathematics in biology increased rapidly, not least due to the impact of ecology and populations genetics. Currently the role of mathematical models increases in many other sciences.

Mathematical models are usually more flexible than analogue models, and they also have greater power of expression.[103] Therefore they make more exact predictions possible, which allows for more conclusive testing of the model. If the variables of a mathematical model correspond to measurable entities, precise empirical testing is possible. Furthermore, mathematical language allows us to investigate structures that we could not imagine on our own or with any available analogue model. This makes mathematics an indispensable tool in particular in those sciences that have moved furthest away from everyday experience.

Analogue models are often easier to understand than mathematical models, but they cannot usually replace mathematical models. The different types of models are used for different purposes. One important usage of analogue models is as starting-points in the construction of symbolic models.

6.2 Idealization

> "What do you consider the largest map that would be really useful?"
> "About six inches to the mile."
> "Only six inches!" exclaimed Mein Herr. "We very soon got to six yards to the mile. Then we tried a hundred yards to the mile. And then came the grandest idea of all! We actually made a map of the country, on the scale of a mile to the mile!"
> "Have you used it much?" I enquired.
> "It has never been spread out, yet," said Mein Herr: "the farmers objected: they said it would cover the whole country, and shut out the sunlight! So we now use the country itself, as its own map, and I assure you it does nearly as well." [104]

This is of course not how we do in the real world. The iconic models that we make in the form of maps and three-dimensional constructions are always simplified, i.e. they exclude parts or aspects that are considered to be less important. Similarly, analogue and symbolic models are always idealizations in the sense of a "deliberate simplifying of something complicated (a situation, a concept, etc.) with a view to achieving at least a partial understanding of that thing. It may involve a distortion of the original or it can simply mean a leaving aside of some components in a complex in order to focus the better on the remaining ones."[105]

One example of such an idealization is the treatment in mechanics of bodies as if they had all their mass concentrated into one point. Another example is the assumption often made in economic theory that all actors on the market are well-informed and strictly egoistic. As we

---

[103] Sven Ove Hansson, "Formalization in philosophy", *Bulletin of Symbolic Logic*, 6:162-175, 2000.
[104] Lewis Carroll, From Sylvie and Bruno Concluded, 1893, chapter 11.
[105] Ernan McMullin, "Galilean Idealization", *Studies in History and Philosophy of Science* 16:247–273, 1985, p. 248.

all know, all objects in the real world have their mass distributed, and no humans are perfectly informed or completely self-interested. Nevertheless, these idealizations have turned out to be very successful. They make it easier to perform calculations in mechanics respectively economics. It is of course important to keep track of these and other idealizations and to introduce further complications into the model when they are needed and can reasonably be added.

Idealizations are particularly successful when an experiment can be constructed that eliminates in practice that which the idealization excluded from the model. Hence, a model of a chemical reaction between two molecules usually excludes the impact of other molecules that happen to be in the vicinity. Comparisons between theory and experiment can be hampered by the effects in the experiments of other molecules than the two reacting ones. In many cases this problem can be solved by performing experiments in gas-phase. This means that the reaction takes place far away from other molecules. Idealized theoretical models can then be studied against experiments that are "idealized" in the same way. (Unfortunately, most chemical reactions cannot be studied experimentally in gas-phase.)

As was pointed out in a previous chapter, the scope for idealizations is smaller in the technological sciences than in the natural sciences. The physicist can be satisfied with a theory of electromagnetism that does not take gravitation into account. The engineer who constructs a machine based on electromagnetism cannot afford to exclude gravitation from her model unless she knows that its impact is negligible. In the same way it would be quite reasonable from the viewpoint of theoretical mechanics to neglect the effects of winds and storms when calculating the mechanical stability of an arch. A bridge builder who made the same idealization would make a serious mistake.[106] The complications of reality are therefore much more unavoidable in the models of technological sciences than in those of natural sciences. (On the other hand the engineer is usually much less concerned than the natural scientist with finding a mathematically exact solution to her problem. A sufficiently close approximation is just as good for technological purposes as a mathematically exact solution.)

6.3 Problems with models

Models and idealizations are necessary in science, but they can also be dangerous if we forget that the real world is more complex than the models. Two examples can illustrate this.

Ever since Faraday, physicists have drawn pictures where a force field is represented by lines. (Usually, they are so defined that in every point the vector of the vector field is a tangent of the line). The density and the exact position of these field lines are chosen arbitrarily in order to produce an illustrative picture. Clearly, the force fields are not concentrated to the lines. Believing this would be just as absurd as to believe that air pressure is concentrated to the isobar lines that meteorologists draw on their maps. Nevertheless, field lines have been misinterpreted in exactly this way. Dowsers claim that they have identified "fields" in the form of parallel lines (Curry lines), i.e. fields that look like the illustrations in physics books. (The dowsing rod or pendulum only reacts to the dowser's expectations. There is no reason to believe that these lines exist.)

[106] Ronald Layman, "Applying idealizing scientific theories to engineering", *Synthese* 81:353-371, 1989, p. 355.

The other example is quantum mechanics. This is a physical theory with a comparatively complex mathematical model that yields very exact predictions for a wide spectrum of experimental situations. Unfortunately there is no good analogue model that provides us with a unified account of quantum mechanics. Instead, two analogue models are used to account for different experiments. According to one of these models electrons and light behave like particles; they can be visualized small balls. According to the other analogue model they behave like wave movements; they can be visualized as waves on a water surface. Neither of these two models can account for everything that mathematical quantum mechanics can account for. It is only in its mathematical formulation that quantum mechanics is unified and consistent. In some popular texts it has been claimed for instance that light is in some mystical way “both a particle and a wave movement”. It is in actual fact neither, but something third, that is very difficult to express in non-mathematical language.

6.4 Simulation

It is one of the major advantages of mathematical models that they can be used for computer simulations. The most well-known examples of this are the mathematical models of movements used by meteorologists to make weather forecasts. Simulations are also used in many other contexts. Hence, a mathematical model of deformations in a car can be used to simulate different types of collisions.

The computer programs used in simulations are usually based on many iterations of one and the same type of calculation. Two major types of iterative processes are used in simulations. One of these is processes in which each step corresponds to a period of time. This method is used in weather forecasts. In each step the machine calculates the weather for some short period of time. The outcome of this calculation is then used as an input for the next calculation, etc. The same method is used in other simulations of temporal developments.

The other method employs iterative processes to gradually improve an approximation. The calculation begins with a very rough approximation. In each step the approximation is improved, and the improved version is used as an input into the next computational step. The calculation goes on until the approximation has become stable, i.e. it is not changed much with additional computational steps. This method is used in particular in simulations that require the solution of very complex systems of equations.

Simulations make it possible to study complex cases that are inaccessible to exact mathematical methods. In this way they can reduce the degree of idealization. Quantum chemistry, that applies quantum mechanics to chemical problems, is an excellent example of this (even though the term “simulation” is seldom used about quantum chemical calculations). Quantum chemical equation systems cannot in practice be solved with pencil and paper unless they represent very simple systems such as a hydrogen molecule. With sufficient computer power even large biochemical molecules can be studied with quantum chemical methods.

Random processes can be simulated with a random generator (“Monte Carlo method”). Suppose for instance that you are simulating the flow of traffic in a crossing where a fifth of the cars go left and the others right. Then you can let a random generator distribute the vehicles between the two directions, so that on average every fifth car will drive to the left.

In almost all simulations it is important to investigate how robust the outcome is against small changes in the input variables. If very small changes of the input can lead to very big changes in the outcome, the system is chaotic. The weather is an example of a chaotic system. In a mathematical model of the weather of the whole earth, a small change in one place can have a large impact on the global weather several weeks later.

In some cases simulations can be used instead of experiments. Hence, simulations can be used to study chemical reactions that are too fast to be observed experimentally. Of course it is always required that the simulations accords with the experiments or observations that are available. Simulations do not eliminate the need for empirical testing.

7 Explaining

We develop scientific theories and models in order to understand the world that we live in. In other words, science attempts to explain that which it studies.

7.1 Science without explanations?

Some philosophers of science have regretted this focus on explanations in science. They claim that explanations should be eliminated from science. Instead, science should be restricted to the description and prediction of phenomena.

This attitude has a long history, in particular in astronomy. Babylonian astronomers seem to have contented themselves with predicting the movements of heavenly bodies. In contrast, Greek astronomers tried to explain what they observed, for instance with theories about rotating spheres supposed to surround the earth. An attitude similar to that of the Babylonians regained influence in the 16th century. Astronomers of that period concluded that no observations that they could make could settle the major contested issues of their discipline, such as whether the earth went around the sun or the sun around the earth. This led some of them to restrict the task of astronomy to that of describing and predicting the movements of heavenly bodies. It was beyond the reach of astronomers to explain what they observed.[107] This was all changed when telescopes made a new type of observations possible.

One of the most influential theoreticians who wanted to eliminate explanations from science was the physicist and philosopher Ernst Mach (1838 – 1916). In his view, the only conclusions that should be drawn from observations were those that could be expressed in the form of abbreviated descriptions of the observations. This means that "only the relation among facts is valuable – and this is exhausted by descriptions".[108] No explanations in the proper sense of the word would be needed. His British follower Karl Pearson (1857-1936) claimed that "nobody believes now that science explains anything; we all look upon it as shorthand description, as an economy of thought".[109] This conception of science had a major influence among the logical empiricists of the 1920s.

However, such economy of thought tends to economize away large parts of that which we usually count as important in science. It is difficult to see how it would be possible to conduct research without trying to understand. Science is a human activity, and as such it aims at human understanding. Possibly a community of robots could be programmed to conduct science without pursuing explanations, but this is not the type of science that we humans develop in order to gain insights about the world we live in, about ourselves and our options for action.

---

[107] Harré, Philosophies of Science, pp. 45-47, 81-82.

[108] Ernst Mach: *Die Principien der Wärmelehre*, 2nd ed., 1900, p. 437. Quoted in: Mario Bunge, *Causality and Modern Science*, 3rd ed, 1979, p. 284.

[109] Pearson, *The Grammar of Science*, 3rd ed (1911), p. v, quoted in Mario Bunge, *Causality and Modern Science*, 3rd ed, 1979, pp. 284-5.

7.2 Explanation and understanding

Explanations are common in everyday life. The answer to a "why" question is usually an explanation. Some examples:

> "Why did he bang the door?" – "He got angry when you slandered his brother."
>
> "Why did the vase break?" - "A stone fell down on it."
>
> "Why does a gas expand when it is heated?" - "The molecules move faster at higher temperatures, and this gives rise to more and more forceful collisions that make them move away from each other."

Explanations show how something is connected with, or follows from, something else that we believe ourselves to understand better. Consider the first example above. We know that people can get angry when their friends are defamed, and we also know that some people slam doors when they are angry. We know this not least because we recognize some of these patterns from ourselves. Explanations of human behaviour usually take this form: we put ourselves in their position and try to enter into their thoughts and feelings. This has been described as using oneself as a model for understanding other persons.[110] It is usually in this way we try to understand why historical figures behaved as they did, or what emotions a poet endeavours to convey.

Another important principle of explanation can be described as understanding "by doing" or by thinking through an hypothetical series of actions. Many everyday explanations have this form. The best way to understand a conjurer's trick is to either perform it yourself or at least think through how you would perform it. Many explanations of natural phenomena consist in dividing them into a series of consecutive steps in the same way as we divide a complex task into a series of actions that we know how to perform. Nature, the human body and even the human mind are often explained by saying that they function as machines. We believe that we understand natural or social phenomena when we know a "mechanism". The concept of a mechanism is difficult to capture, but it is related to that of a machine whose different elements interact in a planned way so that the machine as a whole achieves the intended outcome. In biology a mechanism is usually a series of comparatively simple events that in combination give rise to the larger, more complex process that we want to explain.[111] Hence, reflexes can be explained by describing step by step what happens in the nervous system when a reflex is elicited.

In the third example above the explanation is taken from a scientific model (the kinetic gas model). This is a common way to use scientific models. However, all scientific models do not provide us with explanations. Some models give us accurate descriptions without actually increasing our understanding. One example of this is Kepler's so-called third law, according to which the orbital period of the planets stands in proportion to the 3/2 power of their longest

[110] Arthur Ripstein, "Explanation and Empathy", *Review of Metaphysics* 40:465-482, 1987.
[111] William Bechtel and Adele Abrahamsen, "Explanation: a mechanist alternative", *Studies in History and Philosophy of Biology and Biomedical Sciences* 36:421-441, 2005.

distance from the sun. This mathematical model summarizes data in an excellent way, but it does not increase our understanding of planet movements. Such understanding, however, was achieved when Kepler's third law was shown to follow from Newton's mechanics.[112]

What principles of explanation are accepted in science is not given once and for all, but has changed as science has evolved. The oldest natural science employed essentially the same principles of explanation as those of everyday thought. As science developed, some of these principles have been abandoned, others have been modified, and new ones have been added. In this way, the explanatory principles of natural science deviated more and more from those of everyday life.

One example of this is the attitude of physicists to action at a distance. For long, physicists maintained that bodies can only influence each other if they directly touch each other. This idea has some support in our everyday experiences. Nevertheless it turned out to be too limiting to allow the construction of models that account for our observations of the physical world. The study of magnets gave rise to the concept of a field (first indicated by Robert Norman[113]), and finally Newton took the large step of introducing action at distance (gravity) as a central explanatory principle in mechanics.

The repertoire of accepted explanatory principles has changed in the social sciences as well. Still in the 18th and 19th centuries, social scientists often referred to the needs and strivings of peoples and nations. In the 20th century, such explanations were largely abandoned, and replaced by the ideal of methodological individualism. This means that intentions and intentional actions are only ascribed to individual persons. Nations and social classes do not then intend or act in any other sense than as an abbreviated description of the intensions and actions of individuals.[114]

In a longer perspective, the explanatory principles in science have changed substantially. It might be tempting to conclude that the choice of such principles is arbitrary. However, the changes in explanatory principles have not taken place in an arbitrary way or according to fashion. Instead, explanatory principles have been rejected or accepted according to their ability to provide us with a coherent picture of what we have known, at each point in time, about the world we live in. Everyday experiences and concepts have been the starting-point, but they have been adjusted as we have acquired more and improved knowledge.

7.3 Abandoned principles of explanation

Two important examples of rejected explanatory principles are functional explanations and fate explanations.

In everyday life, when talking about things created by humans, we can often explain their properties by referring to the purpose that they serve:

> "Why does this spoon have such a long handle?" – "In order to reach deep into jam jars."

[112] Herbert Simon, "Discovering Explanations", *Minds and Machines* 8:7-37, 1998.
[113] Harré, Great Scientific Experiments, p. 48-49.
[114] Raimo Tuomela, "Methodological Individualism and Explanation", *Philosophy of Science* 57:133-140, 1990.

Explanations of this type of are called functional. Such explanations are highly relevant in discussions of humans activities and things that humans have created. As should be expected, functional explanations are much used in technological science. But functional explanations have also been employed in much wider contexts. Many pre-scientific explanations were functional in spite of not concerning things created by humans. The question "why is there a sun?" could be answered by saying that humans need its light and heat. Such an explanation seems to presuppose some intelligence that makes things serviceable for us.

Functional explanations in their original form have largely been eliminated from non-technological science. However in some cases they have been retained as abbreviations of other, underlying, explanations. This applies in particular to explanations that refer to evolutionary processes. If asked "why does the giraffe have a long throat?" it is reasonable to answer: "In order to reach the leaves of the acacia tree." This is a functional explanation: but it should be understood as an abbreviation of a non-functional explanation, namely: "Natural selection was favourable to individuals who were tall enough to reach the leaves high up in the acacia trees. These individuals had higher survival rates and therefore the genetic trait of long-neckedness has become over-represented among the off-spring."

The other type of explanations that have been eliminated from science are those that refer to personal fate. Belief in fate was much stronger in ancient times than what it is today. A person who believed in fate could refer to it in explanations of in principle everything that happens in the life of a human being. If one person's house but not that of her neighbours was struck by lightning, there had to be a special explanation why this happened to her and not to them. Every person was supposed to have a pre-determined fate that she could not do much about, or as Sophocles said:

> For unknown and passing great Is the mystery of Fate;
> There is naught that can protect from her nor save,
> Not wealth nor armed powers, Not walled City towers,
> Nor the black ships that battle on the wave.[115]

The ancients' belief in fate was part of a mythological and religious worldview. Sometimes similar ways of thinking persist even today, albeit often without a clear religious basis. The person who draws a price in a lottery often wants to believe that there was a special explanation why just she won.

In modern science fate explanations have largely been replaced by references to chance and randomness. If a loose brick falls down and hurts me when I pass a building, someone who believes in fate will search for an explanation of why I was hit, such as a curse. A proponent of modern science would see this just as a matter of chance. That I of all persons who passed by got the brick on my head is, according to this view, nothing that has to be explained at all.

As this example shows, science has not only changed our principles of explanation; it has also changed our conceptions of what we should at all try to explain. Previous generations of scientists often looked for explanations of phenomena that we today regard as random effects

[115] Sophocles, The Antigone, transl. Gilbert Murray, London: George Allen & Unwin 1951,IV:1, strophe 950,p. 62.

in no need explanation. Hence, Kepler put much effort into explaining why the number of the planets (those that were known at the time) was exactly six. Today, the exact number of planets does not belong to that which astronomers at all try to explain. It is an important insight in modern natural science that everything does not need to be explained.

7.4 Reductions

Scientific explanations often take the form of explaining complex phenomena in terms of simpler ones. Such an explanation can be called a reduction.

In the 19th and 20th century a large number of important reductions were put forward and became generally accepted. The kinetic gas theory explains the properties of gases in terms of the mechanical movements of the molecules that they consist of. The laws of optics have been derived from the wave theory of light, and later from quantum mechanics. Many chemical phenomena have been shown to have physical explanations. Chemical discoveries, in particular the structure of DNA (1953) have substantially increased our understanding of biology. Summing it all up, a large number of successful reductions have brought the different natural sciences, physics, chemistry, geosciences and biology, much closer to each other than what they were before. There has also been some, albeit limited success in explaining psychology and social science in biological terms. Science as a whole has a much more integrated corpus of knowledge than ever before.

There is an old dream of reducing all the sciences to physics. The social sciences would then be reduced to psychology, psychology to biology, biology to chemistry and chemistry to physics. Are we on the way to such a development?

There are good reasons to believe that we are not. Already the reduction from chemistry to physics has severe limitations. The physical account of how particles interact are fully valid in chemistry as well. Indeed, chemists have no need for any other forces or other fundamental principles than those described by physicists. But it does not follow from this that the concepts used in physics are sufficient for chemistry.[116] Chemists use many concepts that cannot be defined in physical terms. This applies for instance to the concepts of substitution and functional groups. They both have essential roles in chemical explanations, but neither of them can be defined in a reasonable way in physical terms. Similarly, many of the concepts that are central to our understanding of biology cannot be reduced to chemistry - such as the concept of a species. Science is a human activity, and in each of the sciences concepts are developed that facilitate human understanding of its particular subject-matter. Our cognitive capacities are limited, and we cannot understand complex chemical reactions in terms of physical interactions, or biological evolution in chemical terms. Therefore, the dream of reducing all science to physics is futile.

[116] Joachim Schummer, "Towards a philosophy of chemistry*", Journal for General Philosophy of Science* 28:307-336, 1997, p. 307.

8 Finding causes

In most branches of science, causes and causal connections are sought for. Although "cause" is a central concept in science, it is very difficult to explain. Attempts to define it in terms of simpler concepts have failed. Many have thought of it as an almost mystical concept.

8.1 Cause as exceptionless repetition

Radical empiricists have tried to get rid of the complications in the concept of causality. They have claimed that science must have a precise and well-defined concept of causality, even if this requires large deviations from conventional views of causality. The most well-known proponent of this view was David Hume (1711 – 1776). He considered causal connections to be purely mental constructions. According to him, that *A* causes *B* means that *A* regularly precedes *B*. When we have experienced the series *A*–and–then-*B* many times, this creates a mental habit of seeing them as connected with each other. According to Hume and other empiricists, this is all that needs to be included in a scientific concept of causes. "*A* causes *B*" then means neither more nor less than "if *A* then always *B*". Following Hume, Auguste Comte (1798 – 1857) claimed that we should not look for causes in the traditional sense, but only for "their effective laws, that is, their invariable relations of succession and similitude".[117] Another empiricist, Hans Reichenbach (1891 – 1953), defined causality as "exceptionless repetition" of a series of two events.[118]

It is easy to show that this definition brings us far away from what we normally mean by a cause. Soon after my alarm clock rings in the morning my neighbour leaves in her car. From this it does not follow that my alarm clock has caused her to leave the house.

Many attempts have been made to supplement the concept of exceptionless repetition, for instance with different combinations of clauses requiring that causes are necessary and sufficient conditions for their effects. These attempts have not, however, led to any definition that corresponds to the concept of causality as we know it.[119]

An important conclusion can however be drawn from these attempts at definitions: A cause cannot take place after its effect. Either the cause comes first, or the cause and the effect are simultaneous. As one example of the former, a large number of persons recently died in a famine in East Africa. The shortage of food came before the deaths. As an example of the latter, the reddish-brown colour on my axe was caused by rust. The colouring came at the same time as the rust that it was caused by.

8.2 The concept of cause is anthropomorphic

An important clue to understanding the concept of causality can be gained by considering its application to human behaviour. In our daily lives, we perform actions that change the world

[117] Comte, Cours de philosophie positive, Ie leçon, vol I, p.3, 1830. Quoted in Mario Bunge, *Causality and Modern Science*, 3rd ed, 1979, p. 68.
[118] Reichenbach, The Rise of Scientific Philosophy, 1951, p.158. Quoted in Mario Bunge, *Causality and Modern Science*, 3rd ed, 1979, p. 42-43.
[119] Michael Scriven, "The Logic of Cause", *Theory and Decision* 2:49-66, 1971.

outside of our bodies. Hence, by moving your hands in a certain way you can break off a twig. In examples like this the concept of a cause is uncomplicated. The bodily movements are the *cause* that give rise to a change, the *effect*, in the world.

The concepts that we use to describe the world are strongly influenced by our own interactions with it. The notion of causality is no exception; it appears to be modelled after how our own actions influence the surrounding world. This concept is in other words anthropomorphic (shaped in human form). When we describe connections in nature in terms of cause and effect we incorporate them in a pattern that is based on experiences from our own actions.

A simple example can illustrate this. We can say that a piece of iron is red because it is hot (has a temperature above 550°C). Why do we not instead say that the object is hot because it is red?[120] The reason for this is closely connected with the anthropomorphic nature of our notion of causality. We can heat a piece of iron, and in that way make it red. In contrast, there is no way to make a piece of iron hot by making it red. If this had been possible we would perhaps have looked differently at the cause-effect relationship. Let us make a thought experiment.

> In a computer game we enter a world where the laws of physics are different from those of our own world. In this world there is a sort of radiation called G-rays. Most materials take on a red colour if they are irradiated with G-rays. To irradiate an object with G-rays is called to redden it.
>
> In this world almost all objects have the same temperature. There are only a few materials that can change their temperature. One of them is iron. Contrary to most other materials, iron becomes hot if you redden it.
>
> The following is a conversation between two persons playing the computer game.
>
> "Why did the iron rod become hot?"
> "Because it was reddened. Iron becomes hot when it is reddened."
>
> In this virtual world it is possible and reasonable to say that iron becomes hot when you make it red but not the other way around.

The anthropomorphic character of the concept of a cause does not prevent us from using it about events that could not have been affected by human hand. We use it for instance about events that took place far before there were humans, and we do not hesitate to talk about causes in connection with both elementary particles and galaxies. In the two latter cases, we move far away from the sphere of human experiences. In such cases the application of the concept of a cause often becomes more uncertain.

---

[120] The example is modified from Douglas Gasking, "Causation and recipes", Mind 64:479-487, 1955. See also Huw Price, Agency and Causal Asymmetry, Mind 101:501-520, 1992.

It is illuminating to compare the anthropomorphic analysis of causality with Hume's analysis. According to both conceptions our concept of causes should be based on empirical experiences. According to Hume they should be based on experiences that we can make as passive observers of external sequences of events. According to the anthropomorphic analysis they are based on our experiences as active human beings who interact with the surrounding world.

8.3 Everything does not have a cause

When we describe a regular connection between two types of events (either in nature or in human societies) we have a strong tendency to do this in causal terms. Cause-effect relationships have been used since very long as a general description model: We think of everything as having a cause. Already in an Indian religious source that is more than two and a half millennia old (the Upanishad) it is claimed that all events are parts of event chains that lead back to one single uncaused physical event and cause.[121]

For modern science, this attitude is not acceptable. We now know that many regularities in nature are difficult or impossible to describe in terms of causes. It is in many cases more accurate to talk of interactions between two or more phenomena, where each of them influences all the others, than to assign to one of them the role of a cause. We can for instance explain the movements of planets and other bodies in the solar system in terms of their gravitational interaction. It would be highly misleading to assign to some part of this system the role of a cause of what happens in the rest of the system.[122]

The concept of a cause often contributes to making science more intelligible. In particular in physics, causality often has a role in intuitive accounts of natural phenomena, whereas it is dispensed with in the more exact and precise descriptions that are achieved with mathematical models. It does not follow from this, however, that the concept of a cause is superfluous. It seems to be indispensable in science for the same reason that explanations are an indispensable part of science.

8.4 Determining causes

The best way to determine a cause-effect relationship is usually an experiment that is performed both with and without the proposed cause. If we want to know if a certain chemical substance causes cracks in plastic tubes, we should preferably make an experiment where plastic tubes are exposed to the chemical and then compared to otherwise similar tubes that have not been exposed to the chemical.

But as we saw in Chapter 3 it is many cases not possible to carry through such experiments. If we want to know if a substance causes myocardial infarction in humans, this cannot be done by experimenting on humans. If humans have already been exposed to the substance we can instead search for exposed persons and compare their health status to that of unexposed persons. This, however, is an error-prone procedure. We cannot know if there are

[121] Radhakrishnan (1931), Indian philosophy, vol I, kap v; quoted in Mario Bunge, Causality and modern science, 3rd ed, Dover, p. 247.
[122] Mario Bunge, *Causality and Modern Science*, 3rd ed, 1979, pp. 149-150.

The post hoc mistake

Already Aristotle pointed out that we humans are too eager to interpret correlations or coincidences as cause-effect relationships. He warned against the fallacy that is known as *post hoc ergo propter hoc*, "after this, thus because of this". The mistake consists in believing that since B follows after A, B is caused by A.

Medical articles in tabloids provide many examples of this fallacy. A typical article describes how some person has regained health after a particular treatment. Contrary to what is usually implied in such articles, such a report does not allow any conclusion about the effects of the treatment. As Aristotle told us, "after" is not the same as "because of". If a sufficient number of patients receive a treatment that does not help against their disease, some of them will improve for reasons unrelated to the treatment. This will happen even if the treatment tends to aggravate the disease.

any other factors than the substance itself that differ between exposed and unexposed persons. Perhaps there is a common underlying factor behind both exposure and disease?

This is exactly what happened in the beginning of the 1980s when it was discovered that many AIDS patients had used the drug amyl nitrate. Both researchers and government officials believed that the drug could be a causal factor that contributed to the disease. It turned out, however, that the correlation was not due to such a causal relationship. Instead it depended on a common, underlying factor. Persons with a very active sexual life often used this drug, and these persons also run a much increased risk of contracting AIDS.[123]

Another example is an investigation showing that persons in Great Britain who were born under certain zodiac signs were overrepresented in intellectual professions. A closer analysis of the material revealed that the results depended on the cold winters in badly heated English homes. Parents who planned a pregnancy preferred their babies not to be born in the winter. Planned pregnancies were more common in the upper social classes, and it was in these classes that the children were educated for an intellectual profession.[124]

It is important to observe that the strength of a statistical correlation does not determine if there is also a causal connection. There is for instance a strong statistical correlation between the viscosity of asphalt and the prevalence of polio in different geographical areas. However, this does not mean that polio is caused by soft asphalt. Instead, the two phenomena have a common, underlying cause: High temperatures make asphalt softer and polio more common.[125] It is in practice often difficult to distinguish between cause-effect relationships and correlations that depend on common underlying factors. Two basic rules should be applied in all analyses of causality. First: always look actively for common underlying factors that can be alternatives to direct causality. Secondly: Be suspicious against proposed cause-effect relationships that are not supported by known causal mechanisms.

[123] Harris Pastides, An epidemiological perspective on environmenal health indicators, World Health Statistics Quarterly 48:140-143, 1995. S Greenland & H Brenner, Correcting for non-differential misclassification in ecological analyses, Applied statistics

[124] Geoffrey A Dean, IW Kelly, James Rotton och DH Saklofske "The Guardian Astrology Study: A Critique and Reanalysis", *Skeptical Inquirer* 9(4):327-338, 1985, p. 331.

[125] Alcock, *Parapsychology: Science or Magic*?, p. 98.

8.5 Connections between several causal factors

We have discussed simple cases of causality in which an effect results from a single causal factor. Actual causality is often much more complicated, and there may be several causal factors that combine to give rise to an effect. One example of this is the set of causal factors that can contribute to a myocardial infarction – smoking, obesity, lack of exercise, genetic predisposition, dietary habits, etc.

If we want to know to what degree myocardial infarctions are caused by obesity, we can compare the prevalence of this disease in the general population (with its mixture of obese and non-obese persons) with its prevalence among non-obese people. Suppose that the latter group has 15% lower frequency of myocardial infarctions than the former. We can then conclude that 15% of the infarctions would not have occurred if it were not for obesity. In other words, 15% of the myocardial infarctions are attributable to obesity.

It is often believed that every disease should have a set of causal factors, such that the sum of their contributions, expressed as percentages, is 100. This, however, is not the case. Often several causal factors have to be present for a disease to manifest itself. Then the removal of only one of these factors will be sufficient to prevent the disease. In such cases, the sum of percentages can be higher than 100.[126]

This can also be shown with an example from another area, namely traffic safety.[127] Suppose that we have carefully investigated the causes of death on a dangerous road by comparing it to other roads that have been rebuilt in order to improve safety. We therefore know that 20% of the deaths on this road can be eliminated with improvements of the roads. We also know that 40% can be eliminated by reducing speed limits to 40 mph. It would then be tempting to say that 20% of the deaths are caused by the road itself, 40% by too high velocities and 40% by other factors. However, this is completely wrong. We can in all probability not remove 60% of the deaths by both improving the road and reducing the speed limits, for the simple reason that some deaths could have been prevented with either of these two measures.

This becomes clearer when we consider additional causal factors. In this example, we may assume that 50% of the deaths can be eliminated by eliminating drunk driving, 90% by introducing speed bumps on the road that reduce the speed to 20 mph, 98% by assigning to each driver a personal police officer, etc. The sum of these percentages can thus be far higher than 100. There is no straight-forward way to reduce the complex interactions between different causal factors to a distribution that adds up to 100 percent.

For similar reasons it is not possible to determine how large part of human intelligence is inherited respectively depends on the environment. By comparing identical twins that have grown up in different families we can find out the effects of the environmental differences that these twins were exposed to. However, these results do not tell us the size of the contribution of "the environment" to a person's intelligence, although they are very often interpreted in this way. If we could instead compare identical twins who had been exposed to

---

[126] Murray, Christopher JL and Alan D Lopez 1999, "On the Comparable Quantification of Health Risks: Lessons from the Global Burden of Disease Study", *Epidemiology* 10:594–605.

[127] Smith, Kirk R, Carlos F Corvalán och Tord Kjellström 1999, "How Much Global Ill Health Is Attributable to Environmental Factors?", *Epidemiology* 10:573–584, p. 577.

much larger differences in their social and intellectual environments, then the results would (with the same erroneous interpretation) have shown that intelligence is environmentally determined to a much higher degree.

Hence, it is misleading to at all assign percentages of contribution to inheritance and environment. Instead we should investigate the complex interactions between genetics and environment that determine our mental capacities. We should not take it for granted that these processes are always best described with the concept of a cause.

9 Science, values, and world views

Values enter science in many ways, for instance in the choice of areas for research and in our judgments of what research is ethically defensible. Here we will consider two aspects of the highly contested relationship between science and values. One of these concerns how values influence science, more precisely the role of values in determining which scientific hypothesis should be accepted respectively rejected. The other concerns how science influences values, more precisely its role in the evolution of our worldviews and out views of human existence.

9.1 Scientific decision-making

An essential part of the research process consists in deciding which hypotheses to accept and which to reject. In these decisions two major types of mistakes are possible: to accept an erroneous hypothesis and to reject a correct one. The relative seriousness of the two types of error can be crucial for the decision.

In an article published in 1953 the American philosopher of science Richard Rudner (1921 – 1979) put focus on the role in science of decision-making about hypotheses. He argued against the common view in those days that a strict division between facts and values is possible, and that science should be kept free from values. His main argument was that a decision to accept or reject a scientific hypothesis must take into account not only the available empirical evidence, but also the seriousness of the two types of mistakes:

> "But if this is so then clearly the scientist as scientist does make value judgments. For, since no scientific hypothesis is ever completely verified, in accepting a hypothesis the scientist must make the decision that the evidence is *sufficiently* strong or that the probability is *sufficiently* high to warrant the acceptance of the hypothesis. Obviously our decision regarding the evidence and respecting how strong is 'strong enough', is going to be a function of the *importance*, in the typically ethical sense, of making a mistake in accepting or rejecting the hypothesis... *How sure we need to be before we accept a hypothesis will depend on how serious a mistake would be*." [128]

The acceptance or rejection of scientific hypotheses for instance about positive and negative effects of vaccines and pharmaceuticals has large practical impact. In such cases, said Rudner, it is unavoidable that ethical values will have an influence scientists' judgments of the seriousness of the two types of error and therefore also on which hypothesis they accept or reject.

In a comment to Rudner's article, Carl Hempel (1905 – 1997) attempted to paint a more nuanced picture of the value-dependence of science. Hempel conceded that the choice between accepting and not accepting a hypothesis depends on values. However these were not ethical values, but another type of values that he called epistemic or scientific values.

[128] Richard Rudner, "The scientist qua scientist makes value judgments", *Philosophy of Science* 20:1-6, 1953, p. 2.

According to Hempel such values "should reflect the value or disvalue which the different outcomes have from the point of view of pure scientific research rather than the practical advantages or disadvantages that might result from the application of an accepted hypothesis, according as the latter is true or false."[129] According to Hempel epistemic values reflect the need of truthfulness, simplicity, explanatory power and other useful assets of a scientific theory. Paraphrasing Rudner, Hempel maintained that "[*t*]*he scientist qua scientist does indeed make value judgments*, but they are not of a *moral* kind; rather, they reflect the value to basic research of constructing sound and information-rich accounts of the world; and these are what I would call *epistemic* values."[130]

The notion of epistemic value has had considerable impact, and it has been further developed by other researchers.[131] However, many of these have combined Rudner's and Hempel's proposals and maintained that both epistemic and moral values have a role in scientific hypothesis testing.[132] Most researchers would probably require stronger evidence for an hypothesis that supports prejudices that they abhor than for an hypothesis that has the opposite implications. Arguably, the requirements of evidence will also in practice be stricter if powerful economic or political interests are at stake.

The burden of proof issue has been much discussed in relation to environmental policies. Action against suspected environmental dangers has often been proposed although full scientific evidence of the danger is not at hand. A decision can for instance be made to treat a chemical substance with caution because there are science-based suspicions that it is dangerous to health, although these suspicions do not amount to full scientific evidence. The use of lower criteria of evidence for practical action than for scientific proof has been called the precautionary principle.[133]

9.2 Distinguishing between facts and values

It has sometimes been claimed that facts and (non-epistemic) values are so intertwined, at least in some branches of science, that it is not worth trying to distinguish between them. This is not so. Even if we cannot separate facts and values completely from each other, it is often desirable to separate them as far as possible. This is one of the most efficient means that we have to disentangle complex issues and to make rational discussions of controversies possible.

That we endeavour to distinguish between facts and values does not mean that we believe them to exist as two types of entities that can at least in principle be completely distinguished between. Neither facts nor values "exist" in the sense that stars and flowers exist. Their mode of existence is that both factual aspects and value aspects are present in the statements we make about the world. Both in science and in everyday life, is often useful to develop each of these aspects separately, i.e. to distinguish between facts and values.

---

[129] Carl G Hempel, "Inductive inconsistencies", *Synthese* 12:439-469, 1960, p. 465.
[130] Carl G Hempel, "Turns in the evolution of the problem of induction", *Synthese* 46:389-404, 1981,p. 398.
[131] Isaac Levi, "On the seriousness of mistakes", *Philosophy of Science* 29:47-65, 1962. John C Harsanyi, "Bayesian decision theory, subjective and objective probabilities, and acceptance of empirical hypotheses", *Synthese* 57:341-365, 1983.
[132] Robert Feleppa, "Epistemic utility and theory acceptance: Comments on Hempel", *Synthese* 46:413-420, 1981.
[133] Sven Ove Hansson, "Adjusting Scientific Practices to the Precautionary Principle" Human and Ecological Risk Assessment, 5:909-921, 1999. Per Sandin, Martin Peterson, Sven Ove Hansson, Christina Rudén, and André Juthe, "Five Charges Against the Precautionary Principle", Journal of Risk Research, 5:287-299,2002.

The usefulness of this distinction depends on the context. Some branches of science are closely connected with moral values that are so generally accepted that it does not seem worth while to separate them. This applies in particular to medical science. It would hardly be helpful to interrupt a discussion about analgesics to point out that it is based on the value assumption that it is better if patients feel less rather than more pain. However in some cases scientists tend to treat their value assumptions as less controversial than what they really are. This applies in particular to the social sciences. Hence, economists tend to assume that more production of goods and services is always better than less production. This is a premise that has sometimes been questioned on moral or environmental grounds. It can therefore be useful to state this assumption as clearly as possible and to clarify to what extent the obtained results depend on it.

As a general rule, if values that influence science are potentially controversial, then they should be specified as far as possible, and their role in the research should be clarified.

9.3 Science and worldviews

> "In vain did the wise god separate portions of the earth from one another with the hostile ocean, if impious vessels then sail across the untouchable waves. The human race, daring to attempt everything, straight away undertook the forbidden deed."[134]

This complaint about technology, in this case ship-building, was written by Horace (65-8 BC). It is a contribution to a debate that has been on-going for more than two thousand years, namely the debate about the value or disvalue of the new tools that science and technology have made available. Science has provided us with printing and gunpowder, with vaccines and biological warfare, with antibiotics and environmental pollutants, with the Internet and military command centres. What attitude we take to this dramatic, two-sided development of human abilities depends on our optimism or pessimism, not so much about technology but about humanity itself. This discussion has usually been concerned with science as techne, as provider of new means to manipulate the world. It is important to observe that science as episteme, as provider of knowledge and interpretations of the world, also has a more direct influence on our worldviews, without mediation from practical applications. Science has, in the long run, forcefully undermined old moral values and existential conceptions, and contributed to the development of new ones.

This does not mean that science in itself contains new values that are imposed on society. The ethical implications of new science depends on the ideological environment in which it is received. Heliocentric astronomy would probably not have had the same impact on the worldview if it had not been for the hegemony of the Church in worldview issues. The effects of Darwin's evolutionary theory, relativity theory etc. depended in similar ways on the social, religious, and political environments in which they were received.[135]

The effects of science on our worldviews are unpredictable, if for no other reason because they depend on the unpredictable outcomes of scientific research. But there is a general trend

---

[134] Horace (Q Horatius Flaccus) pp 4-5 in Carmen III, Book I, Horace, Complete Works, Boston 1954. Quoted on p. 32 in William McBride "The progress of technology and the philosophical myth of progress", Philosophy and the History of Science, a Taiwanese Journal 1(1) 31-58, 1992.

[135] Loren R Graham, Between Science and Values, New York 1981.

in the impact that science has had up to now on our worldviews: It tends to remove ourselves from the central place that we previously had in our picture of the world we live in.

Pre-scientific worldviews gave humans a unique role and place in the universe. Such a viewpoint is called anthropocentrism. Expressions of this view can be found for instance in medieval theology. The Church father St. Augustine (354 – 430) said that the wild animals had been created for the needs of humanity – to punish, test and instruct us.[136] The influential theologian Petrus Lombardus (1100 – 1160) wrote:

> "Just as man was made for God, to serve him, so the Universe was made for man, i.e. to serve him. Therefore man is placed in the middle of the Universe, so that he can both serve and be served."[137]

A large number of scientific discoveries have made this form of anthropocentrism untenable. The sun does not revolve around the earth, but the other way around. The sun has in its turn been shown to be one of about 500 000 millions stars in the Milky Way, which is in its turn one of about 1 000 millions galaxies. Our knowledge of biological evolution has to some extent bereft humanity of its special position among the creatures on earth.

In addition, human behaviour has become the object of the same type of explanations that we apply to the world around us. Biology and psychology offer far-reaching explanations of human passions and human values. Even if some of these explanations have been controversial, the problem arises: With what force can our values be upheld, if they can be explained in the same way that we explain the structure of molecules or the behaviour of ants?

The philosopher Friedrich Nietzsche (1844 – 1900) expressed the effects of science on our worldview as follows:

> "Isn't it the case that since Copernicus the self-diminution of human beings and their will to self-diminution have made inexorable progress? Alas, the faith in their dignity, their uniqueness, their irreplaceable position in the chain of being has gone. The human being has become an animal, not a metaphorical animal, but absolutely and unconditionally—the one who in his earlier faith was almost God… All scientific knowledge … is nowadays keen to talk human beings out of the respect they used to have for themselves, as if that was nothing more than a bizarre arrogance about themselves."[138]

Instead of rejecting science for its lack of anthropocentrism we should carefully distinguish between descriptive and ethical anthropocentrism. It is the former type of anthropocentrism

[136] Quoted in Bergen Evens, The Natural History of Nonsense, NY 1946, p. 143.
[137] Quoted in Sven Ove Hansson, Vetenskap och ovetenskap, Stockholm, Tiden 1995, p. 71.

[138] Friedrich Nietzsche, Zur Genealogie der Moral III:25, Gesammelte Werke, München 1925, 15:440-441. Translation by Ian Johnston from http://brothersjuddblog.com/archives/2003/11/since_copernicus.html.

that we have discussed here. It puts man at the centre of its description of the world, and assigns to her a special role in the universe. Ethical anthropocentrism, on the other hand, puts man at the centre of its views on what ought to be done, building ethics from human welfare and human strivings. Ethical anthropocentrism can also be called humanism. There is no necessary connection between descriptive and ethical anthropocentrism. In the end, no moral view can be derived from our descriptions or explanations of how the world is constructed.

The rejection of descriptive anthropocentrism emanates from our unique ability to search for knowledge and to reflect on ourselves. Science is based on these abilities and is therefore to a very high degree a human project. It is no doubt one of most courageous and most marvellous projects that humanity has undertaken.