Philosophy for the humanities

Victor Gijsbers

22nd November 2021

Contents

Chapter 1

Foreword

This book is still very much a work in progress. You can send remarks of any sort – from spelling errors and translation issues to content related questions and suggestions – to my e-mail address V.Gijsbers@hum.leidenuniv.nl. Please do! (Since the translation of the first three chapters is very recent, I suspect that there are still many errors.)

One important terminological note. I will use the words 'science' and 'sciences' in order to talk about all the fields commonly studied in academia: the natural sciences, the social sciences and the humanities. There is a tendency in English to use the word in such a way that the humanities do not fall under it, a tendency that is not shared by – for instance – the Dutch 'wetenschap' or the German 'Wissenschaft'. Giving in to this tendency would leave me without a good word for science in general; and it would also prejudge the question of how sharp the divide between the humanities and the natural and social sciences actually is. Hence my choice.

Chapter 2

Methodology: an introduction

In our society, science is seen as a reliable way of gaining knowledge. But what is it that makes science reliable? The most straightforward and common answer is that scientists have a special way of developing and testing theories, a way called the *scientific method*, which works better than any of the alternatives. If we want to understand the strength of science, then we must get to know this method. The attempt to describe the scientific method is known as *scientific methodology*.

In this chapter we shall introduce some basic concepts and ideas that will enable us to think about the scientific method in a fruitful way. We will learn, amongst other things, what the standard view of science is; how we can argue rationally; and why it turns out to be hard to figure out how scientists come to their conclusions, or to show that these conclusions are objective and justified. In the next two chapters we will then dive more deeply into the questions surrounding scientific method by looking at some of the answers developed by philosophers of science in the twentieth century.

2.1 Discovery and justification

Which method do we use to gain scientific knowledge? This question can be interpreted, and thus answered, in a number of ways. It therefore makes sense to first clarify what it is exactly that we want to know. Here are three different interpretations:

1. How do new scientific theories and concepts come into being?

2. When are we justified in believing a scientific theory?

3. Under what circumstances does the scientific community accept proposed scientific theories?

The answers to these questions are to be found in very different places. The first question is mostly a *psychological* question that deals with the workings of creativity; the second question is a question of *logic* and *epistemology* which deals with the grounds of rational belief; and the third question is a *sociological* question about how ideas spread within social groups.

We will focus mainly on the second question. Why? We could simply say that logic (the study of reasoning) and epistemology (the study of knowledge) are parts of philosophy, while psychology and sociology are not, and thus that it is simply the duty of the author of a philosophy course book to focus on the second question rather than on the others. But that would be a rather superficial way of reasoning. More important than sticking to our academic pigeonhole is the following: we want, in this and the subsequent chapters, to figure out why we think that scientific claims are *reliable*. To meaningfully assess the reliability of a theory, we have to be able to explain what it means to have good reasons for believing or disbelieving that theory. Hence, the epistemological question is of great importance for us.

How the theory got into being (the question of creativity), on the other hand, is not particularly relevant when you are trying to judge if a theory is believable. It is often said that the chemist Kekulé got the idea that benzene molecules are shaped like rings when he had a dream about a snake that bit its own tail. If true, this is a fun fact; but it is irrelevant when we wish to judge the believability of the theory. A chemist who believes Kekulé's theory because it was revealed to Kekulé in a dream, is not operating scientifically; and the scientist who rejects the theory because the inspiration for it came from a dream is doing something equally unscientific. In judging the validity of a theory, we don't care about the way the scientist came up with it. The only thing we ought to be concerned with is how good the arguments and the empirical evidence are by which the theory is supported.

To clarify this point, philosophers of science have introduced a distinction between the *context of discovery* and the *context of justification*. The former contains everything relevant to the origins and conception of a scientific theory; the latter contains everything relevant to judging, as objectively as possible, whether or not a theory is believable – regardless of how it was conceived. Especially in the first half of the 20^{th} century, most philosophers of science agreed that they should only be concerned with the context of justification. Everything is permitted in the conception of theories; coincidence and subjective factors can play very large roles, without this harming the quality of science in any way. The difference between good and bad science is to be found somewhere else, namely in the right – logical and objective – way of judging whether the theory is believable or not.

In this we shall, for now, follow the philosophical tradition. However, the distinction between the subjective and objective aspects of science is not as clear-cut as it appears at first glance. This is well illustrated by the third, sociological question. Imagine a research study that shows that scientists always and only accept theories devised by researchers who occupy high positions in the academic hierarchy. Would that (imagined) fact about the behavior of scientists not also undermine the believability of all existing scientific theories? Perhaps it is true that strictly speaking the only thing of real importance is the objective relationship between the evidence and the theory; a theory does not become less probable just because certain scientists support it for the wrong reasons. In practice, however, no one scientist ever has all the relevant evidence. We have to base our judgments on those of other and earlier scientists. And this means that it is very important for the reliability of scientific theories that the scientific community functions well.

Here we have discovered an important tension in our thinking about the scientific method. On the one hand, the aim of methodology is to tell us how scientists *should* act, which ways of scientific reasoning are *correct*. We want to know about the *norms* of science. Methodology is a *normative* discipline, a discipline that tells us how things ought to happen. That this doesn't always correspond to actual scientific practice won't surprise anyone. It is certainly part of the scientific method that you are not allowed to make up your own data and then present them as if they were the result of empirical investigation. Nevertheless, there have been scientists – like the Dutch psychologist Diederik Stapel, the German physicist Jan Hendrik Schön and the American historian Michael A. Bellesiles – who did exactly this. The fact that people sometimes engage in unallowable behaviour doesn't imply that such behaviour is part of the scientific method.

On the other hand, methodology should not diverge too much from scientific practice. If I were to tell you a beautiful story about how to justify scientific theories, but this story had nothing to do with how scientists *actually* act, then it is very hard to argue that what I have described to you is the scientific method. Methodology is thus not only normative, but also *descriptive*: it has to give an accurate account of how proper scientific research is done in practice.

Considering these two sides of methodology – the normative and the descriptive – together, scientific method emerges as an *ideal*. It tells us how science should be done; and scientists attempt to approach this ideal in practice, and manage to do so often enough to make science as a whole fairly reliable.

The existence of these two sides also means that we have to test every proposal regarding the scientific method in two different ways. Does the proposed method lead to reliable knowledge? If the answer is no, then we are faced with a normative problem. Is the proposed method actually used? If not, there is a descriptive problem. It will become clear that it is not easy to fulfil both of these demands at the same time. However, if it were to become clear to us that these two demands cannot be reconciled at all, a very drastic conclusion would follow: namely, that our science does not lead to reliable knowledge! This would be quite tragic, especially since it would follow that the diploma you are trying to get is nothing but a worthless piece of paper. (Considered in terms of skills and expertise, at least. That bit of paper may still help you to get a job and make some serious money – alhtough, if that is your goal, humanities might not have been the best choice.) So let us hope that that is not our conclusion at the end of these chapters.

2.2 Empirical content

What are the standards that a scientific theory must live up to for it to be a good theory? Before we look at some of the more the complicated criteria, it is good to first consider several of the more basic ones.

A simple first criterion: a theory should not contradict itself. The following theory about the origins of the Finnish and Hungarian languages would not gain much support:

> The Hungarian language originated in the 12^{th} century from Finnish, when a group Finns settled in the area around Budapest. The Finnish language was born out Hungarian when, in the 18^{th} century, a group of Hungarians settled in the area around Helsinki.

We do not have to conduct any research to understand that this theory is incorrect. The theory *contradicts* itself. It is simply not possible that this theory is true; and it is therefore not a good theory.

The opposite of a contradiction (a statement that cannot be true) is a *tautology* (a statement that cannot be false). You may think that tautologies will be much desired in science, because what is better than a statement that is necessarily true? The problem with tautologies, however, is that they are completely uninformative, as is shown in the following example:

> The Hungarian language is derived from Finnish, or the Hungarian language is not derived from Finnish.

We can, without conducting any research, conclude that this is correct. But it is exactly for this reason that the theory is utterly uninteresting. A scientific theory should tell us something about the world; in this case, about the origins of the Hungarian language. But the theory mentioned above tells us absolutely nothing.

Here is a useful way of thinking about this topic. You can imagine all kinds of possible worlds. Our world is a possible world; but you can also imagine a world that is exactly the same as ours apart from the fact that the shirt I am wearing while writing this sentence is green instead of purple; a world in which everything is exactly the same up until 1593, but in which Ingram Frizer on the 30^{th} of May of that year stabs William Shakespeare to death instead of Christopher Marlowe; a world in which, during the age of the dinosaurs, the Earth is completely destroyed by an enormous comet; and so forth.

We can then ask ourselves about any statement in which worlds it is true and which it is false. “Shakespeare is the most performed playwright” is true in our world, but false in all the worlds where he is stabbed to death at the start of his career. If we use this way of thinking about statements, we can see what is special about contradictions and tautologies. A contradiction is false in every possible world – there is simply no imaginable scenario in which it is true. On the other hand, a tautology is a statement which is true in every possible world – there is no imaginable scenario in which it is false.

Why is this a problem for scientific theories? Because, as scientists, we want to say something about the world. In order to do so, we have to make a distinction between the way our world turns out to be and other ways that it could have been but isn’t. Contradictions and tautologies are precisely the kinds of statement that do not make a distinction between the world as it is and other imaginable scenarios, and therefore these statements can never be good scientific theories.

We want theories that do make such a distinction; and we shall say that the better a theory is at telling us how our world is different from other worlds, the more *empirical content* that theory possesses. It is not easy to quantify empirical content, but we can intuitively understand what is meant by the term. A theory has more empirical content when it tries to describe the exact nature of our world in more detail. Put differently: a theory has more empirical content when it excludes more possible worlds, that is to say, when there are more imaginable scenarios of which the theory claims that they are false. If I describe our world in more detail, then more alternative scenarios will be rejected because they do not fit my description; and the more alternative scenarios I can exclude, the clearer our picture of the world becomes.

In the next paragraph we will ask the question why we speak of ‘empirical’ content. But before we do so, let us first consider the example of two theories, one of which has more empirical content than the other:

1. The Roman empire became substantially weakened in the third century AD.

2. The Roman empire became substantially weakened in the third century AD due to epidemics, including the Plague of Cyprian.

The second theory has more empirical content than the first: it attempts to tell us what happened in much more detail. In other words, some scenarios are excluded by the second theory but not by the first – for example, scenarios in which the Roman empire did become weaker in the third century, but because of bad economic governance.[1]

Which of the two theories will we, as scientists, consider more interesting as an object of research? Clearly the second; we want to know more details, give more explanations, learn more about the world in which we live. It is nice to know that the Roman empire became weakened in the third century; but it is only when we try to investigate the causes of its weakening that things become really interesting. (For the curious: deadly epidemics and the subsequent manpower shortages in agriculture and the military are a plausible cause of the weakening of the Roman empire in this period; but the importance of these causes compared to other causes is still a topic of debate between historians.)

The amount of empirical content of a theory does not depend on its *truth*. Suppose that it was bad economic governance – like the minting of coins of progressively lower purity – that was responsible for the weakening of the Roman Empire. In that case, the first theory is true and the second is false. Yet the second theory still has more empirical content. Indeed, the more empirical content a theory has, the greater the probability that it is incorrect. That is precisely what we should expect: the more detail you give, the more likely it is that some of the things you claim will turn out to be wrong.

A special case of lack of empirical content are theories that aim to explain a phenomenon, but in fact do no more than give it a fancy new name. In the third intermezzo of his play *Le malade imaginaire* (1673), the French playwright Moliére ridiculed the medical sciences of his day precisely because they use such contentless explanations. During an exam for medical students (which is taken in a mixture of poor Latin and French), a student is asked to explain why opium makes people fall asleep. He answers:

> À quoi respondeo,
> Quia est in eo
> Virtus dormitiva.
> Cujus est natura
> Sensus assoupire.

which could be translated as:

> To that I answer:
> because opium
> has a sleep inducing power;

[1] Please note that our definition of empirical content is not precise enough to compare any two theories. For instance, which theory has more empirical content: "the Roman empire became substantially weaker in the third century", or "the Cyprian plague was probably the smallpox virus"? We cannot answer this question. One may doubt whether it is very useful to try to come up with a more precise definition of empirical content; we shall, in any case, not attempt to do so.

the nature of which is
that it makes the senses slumber.

The committee thinks the answer is brilliant; but the spectator of course understands how much the student explained with this theory: nothing.[2]

Scientists are therefore not only looking for theories that are correct, but also for theories that are interesting and contentful. A correct theory without content is easy to come up with: every tautology will serve. If you only aim for content and not for truthfulness, your life will also be very simple: you can simply imagine some very detailed account of the world without regard for its truth. Only if you want to strike a balance between content and truth, to arrive at a content-rich theory which is also probably true, then research becomes a serious and complicated exercise. And only then are you truly doing science.

2.3 A naive view of empirical science

In the previous paragraph we spoke about *empirical* content. As in the phrases 'empirical science' and 'empirical research', the word 'empirical' indicates a relation to observations made with our senses – especially sight, but on occassion also hearing, touch, smell and taste. We may, for instance, make a distinction between empirical research and a literature review. Empirical research involves generating new data (for instance, through experiments or archival research) whereas in a literature review, we review the empirical findings of previous scientists. But a philosopher would be likely to point out that both kinds of research lead to results that are based on observations; in the first case, our own observations, in the second case, the observations of others. In philosophy, we would call any knowledge gained this way *empirical knowledge*.

The opposite of empirical knowledge is knowledge gained through pure thought. Is such knowledge possible? To be sure! For example, to figure out that two plus twenty-eight equals thirty, we have to do no experiments. We don't even have to open our eyes. We can just see that it is true by thinking about it. This kind of knowledge is not empirical. It is also true in all possible worlds: after all, one cannot imagine a world where two plus twenty-eight does not equal thirty. Mathematical truths are tautologies.[3]

So although there is non-empirical knowledge, if we want to know how our world differs from other possible worlds, we will have to open our eyes

[2] The theory that opium has a sleep inducing power, does in and of itself have some empirical content, namely that people fall asleep when they smoke opium. The problem is that this empirical content is exactly the same as that of the fact that needed to be explained (namely, that people fall asleep when they use opium), and that the theory adds nothing to this.

[3] Not all philosophers agree with this claim, but a discussion about this theme would lead us too far astray.

and ears; we have to take an empirical approach. Since scientific theories are neither contradictions nor tautologies, they have to be tested with the help of our senses, and our science must be empirical. That is why it is logical to speak of the *empirical content* of theories, as we have done in the previous paragraph.

Lets return to the question with which we started this chapter, namely the question: what makes science reliable? Why, for example, are we inclined to believe an astronomer who tells us that there will be a visible solar eclipse next month, but not inclined to believe an astrologer who tells us that we will fall in love next month? Both theories have empirical content; they do not contradict themselves; and at first glance they seem about equally probable. (Both events are rare, but we tend to experience them a couple of times in our lives.)

What advantage does the theory of the astronomer have over the theory of the astrologer? Well, we suspect that the astrologer's theory is arbitrary and not based on observations; while we believe that astronomer's theory is based on centuries of carefully collected measurements. (If we find out that astrologer's have collected centuries of high-quality data about when people fall in love and have derived a great theory from these data, then we may have to reconsider our judgment and also believe the astrologer.) The reliability of science seems to come from proper usage of observations. Science works precisely because it is thoroughly empirical.

This is a standard claim about science and it is the basis of an influential and common view of science that can can be summarized as follows:

> Good science is based on careful and accurate observations. These can be observations of experiments in laboratories, of medieval documents from an archive, of artworks in museums – what sorts of things are observed depends on what is being researched.
>
> The beauty of observations is that they are *repeatable* and *objective*: different scientists will be able to reach agreement over the observed data. If I do not believe a statement made by a colleague, than I am able to repeat the observation (by doing the same experiment or going to the archives or the museum). Science is therefore not built on randomly selected ideas, but on a solid foundation of facts.
>
> Subsequently, scientific theories are built on this foundation. Theories that do not fit with the observations are rejected; theories that are confirmed by the observations are accepted. Often different theories are compatible with the observations; in such cases more research is required, until we are able to show which of the theories is correct. This can be a long process, but ultimately scientists can reach consensus about the right way of describing the world.

> As long as we observe the world carefully and as long as we are intellectually honest in comparing theories with observations, science can and will be an objective process that slowly moves towards consensus and will ultimately lead us to the truth about the world. This is why science is reliable.

Is this view correct or is it naive? To some extent it is both. A naive view is not necessarily an incorrect one, and there is some truth to this description of science. We shall nonetheless see that, as soon as we try to work out this view in more detail, we are confronted with all kinds of complicated problems; and that the claim that science is an objective process that moves towards consensus and truth must be made much more nuanced if it is to be at all believable.

2.4 Example: Brecht's *Leben des Galilei*

The naive view of science that was laid out in the previous paragraph is, as I said, influential. We can find a beautiful example of this in Berthold Brecht's play *Leben des Galilei* (1939). In the fourth scene of the play, Galilei (1564–1642) wants to use his self-made telescope to show the moons of Jupiter to his new patron, the young duke Cosimo II de' Medici (1590–1621). The scientists who are present, however, do not believe that Jupiter can have moons, since according to their hero Aristotle (384–322 BC) all objects in the universe revolve around the earth; objects that revolve around Jupiter are thus impossibile. Galilei does not have the patience to deal with their supposedly rational arguments. He attempts to get the duke to simply look through his telescope.

> **Galilei:** Wie, wenn Eure Hoheit die sowohl unmöglichen als auch unnötigen Sterne nun durch dieses Fernrohr wahrnehmen würden?
>
> **Der Mathematiker:** Man könnte versucht sein zu antworten, daß Ihr Rohr, etwas zeigend, was nicht sein kann, ein nicht sehr verläßliches Rohr sein müßte, nicht?
>
> **Galilei:** Was meinen Sie damit?
>
> **Der Mathematiker:** Es wäre doch viel förderlicher, Herr Galilei, wenn Sie uns die Gründe nennten, die Sie zu der Annahme bewegen, daß in der höchsten Sphäre des unveränderlichen Himmels Gestirne freischwebend in Bewegung sein können.
>
> **Der Philosoph:** Gründe, Herr Galilei, Gründe!

Galilei: Die Gründe? Wenn ein Blick auf die Gestirne selber und meine Notierungen das Phänomen zeigen? Mein Herr, der Disput wird abgeschmackt.

Der Mathematiker: Wenn man sicher wäre, daß Sie sich nicht noch mehr erregten, könnte man sagen, daß, was in Ihrem Rohr ist und was am Himmel ist, zweierlei sein kann.[4]

Regardless of what Galileo attempts, the scientists stubbornly refuse to look through his telescope, and they, in the end, are able to lead the duke away without Galilei having been given the chance to show his discovery.

The dramatic potential of the scene was probably more important to Brecht than historical accuracy. But he illustrates quite nicely the naive idea that scientists can always reach consensus, as long as they are willing to take observations serious. The mathematician and the philosopher that Brecht describes are clearly not willing to do so. They refuse to make any observations at all and instead cling to their theories simply because they were defended by Aristotle (as becomes clear at a later point in the scene). It seems very simple: Galilei is the good scientist here, and his adversaries are dogmatic fools.

Yet the arguments of the mathematician in the cited passage are not without merit. If Galileo's telescope shows things that to the best of our knowledge cannot exist, does it not simply follow that there is something wrong with the telescope? Is it not possible that what we observe through the telescope is not an accurate representation of what is actually going on in the universe? Let's not forget that in the time of Galileo the telescope was a new instrument that had not yet proven its worth in practice, and could barely be supported by the available theoretical knowledge. It was quite natural to wonder whether or not this new instrument was reliable.

[4] Translation:

Galilei: What, when your highness can see these both impossible and unneeded stars right now through this telescope?

The mathematician: One might be tempted to answer that your telescope, showing something that cannot exist, must be a rather unreliable instrument, might one not?

Galilei: What do you mean?

The mathematician: It would be much more useful, sir, when you would tell us about the reasons you have for supposing that stars can move around freely in the highest spheres of the unalterable heavens.

The philosopher: Reasons, mister Galilei, reasons!

Galilei: Reasons? When one look at the stars themselves and my compiled observations just show you the phenomenon? My lord, the dispute is becoming tasteless.

The mathematician: If one could be certain that this wouldn't agitate you even more, one might say that what is in your telescope and what is in the heavens can be two different things.

And then, what would duke Cosimo have seen if he had looked through the telescope? Moons that nicely circle Jupiter and thereby prove that Galilei is right? No. He would have seen some vague light spots that would very slowly (taking days) change their forms and positions. The conclusion that these were moons revolving around Jupiter was an interpretation that Galilei arrived at based on many observations, theories and calculations; it certainly could not be proven with just a look through the telescope.

We may therefore suspect that describing what makes good empirical science is not an easy task. Persistently refusing to look through a telescope is not very scientific, but just believing everything you see trough an unfamiliar instrument does not seem to be scientific either. It is the task of the methodologist to attempt to clarify the situation by giving a more detailed analysis of how science works. This is what we start with in the next paragraph.

2.5 Two basic questions of empiricism

Presenting a somewhat naive but very influential view of science in section 2.3, I wrote:

> Good science is based on careful and accurate observations. ... Theories that do not fit with the observations are rejected; theories that are confirmed by the observations are accepted.

These claims are not particularly controversial. Nobody can doubt the importance of observations for the acceptance of theories in the sciences: if empirical research comes up with data that don't fit my theory, the theory is in trouble; but if the data do fit my theory, people will be attracted to it. But things become complicated quite quickly as soon as we ask ourselves what it means for observations to 'fit' or 'confirm' a theory. To get a feel for the problem, let us consider three examples of theoretical claims from across the sciences:

1. The Earth is approximately 4.54 billion years old.

2. The rise of capitalism in Europe was linked to the protestant belief in predestination, because people came to believe that financial success was a sign of being destined for Heaven.

3. There is a significant stylistic break between the modernist literature and postmodernist literature.

Now suppose that we wanted to empirically test these theories. Questions would arise immediately. How do you even attempt to measure the age of the Earth? What evidence could prove a causal link between two movements in the history of thought? How can we determine style and breaks in style?

Each of these questions will lead to many others, and none of these theories can be easily 'proved' or 'disproved'. The difficulty stems, of course, from the fact that these theories talk about things that are not *directly observable*. We can't look at the Earth and just see its age any more than we can pick up some books and hear or smell whether there is a historically significant stylistic break between them.

In order to understand empirical science better, we can make a distinction between two kinds of claims. First, there are claims that can be tested directly by doing a single observation. An example might be "It is currently raining," since that is a statement I can test simply by looking out the window. Second, there are claims that cannot be tested directly, but only indirectly. An example would be: "It rains more in the Netherlands than it does in Spain." The only way to find out whether this statement is true is by recording the amount of rain in the Netherlands and Spain over a long period. So while this *is* a statement that I can test empirically, I can only conclude that it is true or false based on first finding out about the truth of other statements – statements that *are* directly observable.

If we want to clarify the relationship between science and observation, we must ask ourselves two questions, which I shall call the two *basic questions of empiricism*:

1. Which statements can we test directly?

2. How do we get from such statements to statements that cannot be tested directly?

The first question is the question of the *empirical basis*. It asks what the statements are for which we can perform direct empirical tests. They must form the basis on which all our empirical knowledge of the world is based. The second question we can call the question of *confirmation*: when and to what extent are other statements supported by statements from the empirical basis? (We say that a statement is *confirmed* by observations when these observations make the statement more likely to be true; and that it is disconfirmed by observations when they make the statement less likely to be true.) Only by answering both questions can we arrive at a coherent story of how we get from observation to theory.

In subsequent chapters we will dive more deeply into the question of confirmation. In this paragraph we will discuss the question about the empirical basis. Our goal is not so much to arrive at an exact answer to the question, but rather to bring out the difficulties that are associated with it.

As an example of a directly testable statement I used: "It is currently raining." But is that actually directly testable? It is certainly possible to claim that I cannot immediately see that it is raining, but that I come to this conclusion on the basus more elementary observations. I look out of the window and see drops of water falling down. But that these drops are coming from the clouds, and thus are rain, is something I cannot directly see from where I am sitting; it is a conclusion based on previous experiences, including not having any experiences of my upstairs neighbour hanging out of the window with a big watering can.

Come to think of it, do I really see that the drops are drops of water? At most I can see that they are made of a colorless fluid. So the statement that I can test directly is something like: "At this moment, colorless drops of fluid are falling past my window." But even in that statement we can, with a little inventiveness, recognize theoretical assumptions and interpretations.

I am obviously not giving this example to draw the true but somewhat insipid conclusion that "we can never know anything for certain," let alone to suggest that I can never be justified in believing that it is raining outside. But this example can help us to gain insight into how observation works. It helps us see that observation is rarely (maybe never) completely free of theoretical presuppositions: even if we think that we are only noting down the things we see and hear, we still make all kinds of assumptions that are not directly part of the observation itself. We call this phenomenon the *theory-ladenness of observations*.

What does this theory-ladenness mean for scientific practice? That is hard to say. If it turned out that scientists who support different theories also have totally different interpretations of the observations because of these theories, then theory-ladenness would be a very important phenomenon. In that case, observations would not help us to determine which theory is the correct one (because supporters of different theories would accept different observation statements); and in that case the naive view of science would be totally wrong. In the scene of Brecht we saw something like this: the mathematician did not accept Galilei's observations, because they where, based on his theoretical perspective, clearly unreliable. Galilei, at the same time, rejected the theoretical perspective of the mathematician precisely because it did not correspond with his observations. This seems to be a situation in which observation is powerless to create consensus among scientists.

On the other side of the spectrum, it could also turn out that in practice we easily agree on observation statements, and thus that theory-ladenness is of little practical importance. This would fit the naive view of science very well. But even if we have a more nuanced view of science, we might like to draw such a conclusion. After all, it would be a little disappointing if it turned out that scientists could never reach consensus about anything, regardless of how much research they were to do.

It is therefore hardly surprising that, especially in the first half of the twentieth century, many philosophers became very interested in showing that one only needs a few simple concepts to formulate statements in the empirical basis. These concepts would then form a common ground for all scientists. One popular idea was that you only really need statements of the form "I see a bit of dark blue in the top left corner of my visual field;

just to the right of that spot I see a bit of bright green." Statements like "it is currently raining" would then somehow be translated into these more directly observable and thus less theory-laden statements.

We will not go into the detailed proposals of how translation would work, nor will we discuss the problems with these proposals. But it is easy to see that in the characterization of the empirical basis, we need to strike a balance between two things. On the one hand: if we allow statements that contain a lot interpretation, we run the risk that our observations become so theory-laden that scientists cannot come to a consensus. On the other hand: if we only allow very basic statements such as "I see a bit of dark blue in the top left corner of my visual field," then we quickly become alienated from scientific practice. After all, scientists never write down such statements, and would not really know what to do with them if they read them in somebody else's work. As methodologists we would have to put in a lot of effort to connect these basic statements to scientific theories, and to show how complex theoretical ideas can be supported by such very basic observations (where in my field of vision do I have to see spots of which color before I am allowed to believe that the plague of Cyprian caused the fall of the Roman Empire?). In short: the smaller you make the role of theory-ladenness in the first basic question of empiricism, the harder it becomes to find an answer to the second basic question, the question about the relationship between the empirical basis and scientific theories that interest us. Hopefully we can strike a balance so that we can work with statements that are not too tainted by interpretation but that are still practically useful.

We now switch to the second basic question of empiricism: how do we get from observation to theory? As previously mentioned, we can also call this the question of confirmation.

2.6 Introduction to logic

Before we can develop a proper understanding of the different philosophical ideas about confirmation, we have to become familiar with some basic concepts of logic. Logic is the study of *argumentation.* An argument can be conceived of as a series of statements that begins with one or more *premises* and ends with a *conclusion.* (In more complex arguments, there might also be intermediary conclusions, but here we shall only consider very simple arguments.) Here is an example of an argument:

- William Shakespeare only wrote good plays.
- *Titus Andronicus* is a bad play.

- *Titus Andronicus* was not written by William Shakespeare.

Above the line are the two premises, or assumptions; beneath the line is the conclusion, the statement of which the argument tries to convince us. (You can see the line as a visual representation of the word 'thus'.)

Do we accept this argument? That depends on two things. Firstly, we can ask ourselves whether we agree with the premises. You might believe that *Titus Andronicus* is a magnificent play (perhaps because of a perverted preference for chopped off body parts). Or maybe you are convinced that Shakespeare produced a lot crap. In both cases you would reject the line of reasoning because you believe that *the premises are untrue*.

The second reason you can have to reject the argument, is that you may think that the *argument is invalid*. This would mean that, even if the premises were true, the conclusion nevertheless could not be drawn from it. A clear example of an invalid argument is this:

- William Shakespeare only wrote good plays.
- *Titus Andronicus* is a bad play.

- *Hamlet* is William Shakespeare's greatest work.

Even if you accept both premises, you can still disagree with the conclusion – because it doesn't follow from the premises. Whether an argument is valid or not does not depended on the truth of its conclusion: whether or not *Hamlet* is Shakespeare's greatest work, the argument here is simply invalid.

There are thus two reasons to reject an argument: the premises can be untrue, or the argument can be invalid. There is however a big and important difference between these two reasons. To determine whether the premises are true or false, you need to have knowledge about the topic of the argument – I must know something about Shakespeare and about *Titus Andronicus* to be able to determine the believability of the premises. But to determine whether or not the argument is valid, you do not need any knowledge of the topic. What logicians throughout the centuries have noticed is that the validity of an argument can be detected by only looking at the argument's *form*, without paying any attention to its *content*. For example, any argument with the following form is valid:

- Everything that has property X, also has property Y.
- a does not have property Y.

- a does not have property X.

It does not matter what we substitute for X, Y and a, the argument is always valid. We can recognize our original Shakespeare argument if for X we fill in "is a play written by Shakespeare", for Y "is a good play" and for a "*Titus Andronicus*". For then we get:

- Everything that is a play written by Shakespeare, is a good play.

- *Titus Andronicus* is not a good play.

- *Titus Andronicus* is not a play written by William Shakespeare.

and that is the same as what we had before, albeit written our a little more circuitously in order to clarify the form of the argument. Without knowing anything about Shakespeare, we can see that this argument is valid. (But of course not whether the premises are true.) This means that we can study which arguments are valid and which are invalid without paying attention to the content of those arguments, but only by looking at their form; this purely formal study of arguments is precisely what we call *logic*.

This is important for the philosophy of science because it suggests that we could say something about the scientific method – namely how we can draw theoretical conclusions from our observations – without taking into account the exact topic of a theory. A general theory of confirmation valid for all the sciences should be possible.

Before we can understand the problems that the attempt to arrive at such a theory of confirmation will raise, we must introduce one more distinction from logic: the distinction between deductive and inductive arguments. An argument is *deductive* when the conclusion necessarily follows from the premises. If the premises are true, than it must be the case that the conclusion is also true. The argument given above about Shakespeare is deductive: if Shakespeare only wrote good plays, and if *Titus Andronicus* is a bad play, then it simply *cannot* be the case that Shakespeare wrote it. Valid deductive arguments are truth preserving: if the premises are true, than the conclusions must also be true. This is of course a useful characteristic, since it means that we cannot make *new* mistakes when we use valid deductive arguments.[5]

An *inductive* argument is an argument that is not deductive, but where the premises do to some degree support the conclusion. (There are different definitions of inductive argument, but this is the one we will use in this book. By the way, if you happen to know the term 'mathematical induction', be warned: confusingly enough, it is a form of deduction.) It is therefore possible that the premises are true and nevertheless lead to a false conclusion; but the premises do give us some reason to believe the conclusion. Here is an example of an inductive argument:

- The written sources we have from ancient Athens are in Greek.
- There is no textual evidence that people in ancient Athens used a different language for writing than for speech.

[5] We should note that deductive arguments do not necessarily preserve falsehood: a true conclusion can sometimes 'accidentally' follow from false premises. For example: "Shakespeare only wrote good plays. *Medea* is a bad play. So Shakespeare did not write *Medea*." The premises are both false, but the conclusion is true, since *Medea* is a play by Euripides. (Or by Seneca.)

- The people in ancient Athens spoke Greek.

It is hard to hold anything against this line of reasoning. If someone were to ask us why we believe that people in ancient Athens spoke Greek, this is one of the arguments we could give. But it is not a deductive argument: the conclusion does not necessarily follow from the premises. It *could* be that everybody in ancient Athens thought it was so normal to write Greek and speak Hittite that they never bothered to write that fact down; and that we, due to a lack of sources, wrongfully believe that the people in ancient Athens spoke Greek. This is of course highly improbable, but it is compatible with the premises of the argument.

Deductive arguments show that *if* the premises are true, than the conclusion *must* be true as well. Inductive reasoning shows that *if* the premises are true, the conclusion is *likely* to be true – as long as no counterarguments could be brought forward.

How about these counterarguments? A characteristic of inductive arguments is that they can contradict each other, even if all their premises are true. Take the following two arguments. (1) "Victor is a Dutch man. Most Dutch men do not like Shakira. So Victor does not like Shakira." (2) "Victor owns four Shakira albums. Most people who own four albums from one artist like that artist. So Victor likes Shakira." All four premises in these two arguments happen to be true, and still the conclusions are exact and polar opposites! With deductive arguments this would never be possible. (If all the premises of two deductive arguments are true, then both the conclusions must be true as well, and that cannot be the case if they contradict one another.) This is why we often have to "weigh" inductive arguments against each other. There are arguments for both sides of a case, and we will have to determine which arguments are the strongest.[6]

Inductive arguments are thus less straightforward than deductive reasoning. They only give us probability (and then: how much probability?) and they can contradict each other. Still, induction is very important to science. It is after all almost never the case that a theory can be proven, without a shadow of doubt, by the available evidence.

Imagine a scientist making a *general statement*. A linguist who claims that Dutch is a verb second language (a language in which the finite verb is placed in the second position in a sentence), cannot have studied all the sentences of the Dutch language to prove this claim. He has only looked at a limited selection of sentences, and has drawn a conclusion on the basis of this incomplete evidence. He thus uses inductive reasoning. In a similar fashion, the physicist who claims that magnetic poles always attract their opposites (north attracts south and vice versa) cannot have studied all the magnets in the universe – this too must be a case of induction. General statements

[6] For the record, I do like Shakira.

can never be deduced from observations, because it is always possible (even though it might be highly unlikely) that the statement is refuted when we make more observations.

But induction is often also necessary when we are not dealing with general statements; for instance because we may not have any observations that are completely conclusive. Did Shakespeare write *Titus Andronicus*? We might have been able to deduce this from our observations if we had been standing next to him when he (or someone else, as the case might be) stared into the middle distance and murmured:

> "Speak, gentle niece, what stern ungentle hands
> Hath lopped and hewed and made thy body bare
> Of her two branches?"

and then entrusted this poetical gem to paper with gracious strokes of his quill. But alas, this is not the case; and so we must, on the basis of incomplete evidence and with the help of inductive reasoning, judge whether or not the theory is probable. (It is currently commonly accepted that Shakespeare wrote the play, but in previous centuries many critics rejected this idea precisely because they thought the play was too bad, too bloody and too banal for "the Bard".)

Keeping these concepts in mind we shall, in the next section, consider an example of the use reasoning in the humanities.

2.7 Example: *sergentes* in Todi

In order to get a better picture of what a scientific argument can look like, and how this relates to observations, let's look at an example of recently conducted research within the humanities. In "An Italian city-state geared for war: urban knights and the cavallata of Todi" (*Journal of Medieval History*, 2013, vol. 39, no. 2, pp. 240–253) Peter Hoppenbrouwers examines the Italian city state of Todi around the year 1300. It was important for the city states of the divided and war-hungry Italy of that time to be able to field relatively large amounts of cavalry. Cavalry had traditionally been composed of noblemen who had a duty to follow their lord into battle, and thus had to possess an expensive warhorse. The Italian city states, in which the nobility and those were not of noble descent worked closely together, had no feudal structure in place. This meant that a new way of mustering cavalry had to be found. For this purpose a group of people, the so-called *cavallata*, were defined and they were ordered to possess and maintain one or more warhorses. With the further development of this system, it became a type of inheritable tax burden that was not necessarily tied to an individual, but for example to a family; and which was thus no longer tied to actually riding these horses into battle, as was customary in the feudal system. At the same time, it was no longer only the nobility who were supposed to maintain warhorses.

A historian who wants to understand these military and social developments in more detail has to use sources that weren't written for the purpose of explaining precisely which changes took place and for what reasons. After all, you cannot expect that there is a document in a thirteenth century archive in which someone for whom all these developments are completely obvious meticulously explains them for the benefit of a researcher eight centuries later. Instead, you will have to rely on data you do have – which may have more to do with the judiciary system and tax collecting – in order to reconstruct the development in which you are interested.

Hoppenbrouwers's article deals in part with a group of people who are called the *sergentes* in our archival sources. The exact role and nature of these people is not immediately clear. The historian Getulio Ceci, who pulished published a standard work about medieval Todi in 1897, believed that the *sergentes* served in the army as infantry. According to him, the sources suggest that the *sergentes* would normally act as a type of armed police in the city and the surrounding *castelli*; and that in the army they fulfilled the role of a modern sergeant, which is to say that they would lead small groups of foot soldiers. Hoppenbrouwers argues that Ceci was mistaken, and that the *sergentes* actually made up part of the cavalry force. It is obviously important for a historian to know who of two is right, since otherwise we cannot understand the sources that mention *sergentes* properly, and as a result we may misunderstand the development of the armies of Italian city states. Let us therefore take a look at a couple of the arguments that Hoppenbrouwers gives for his thesis. (By zooming in on this specific point we do lose sight of large parts of his article, but for our current purposes this is not a problem.)

First, Hoppenbrouwers refers to a document from 1288 which defines the fines to be imposed on those who did not show up for their military service. The fines are different for foot soldiers from the countryside, foot soldiers from the city, *milites* (horsemen), and *sergentes*. He concludes that

> since the sergeants were mentioned after the *milites*, and not after the *pedites* [foot soldiers], they were considered to belong more to the cavalry than to the infantry (p. 247).

If we reconstruct this argument in somewhat more detail, we end up with the following:

> • If somebody makes a list of the form 'A, B, C, D,' in which C evidently belongs to a different category from A and B, than it is more probable that D fits in with C than that D fits in with A and B.
> • In the sources we find a list of the form 'foot soldiers from the

countryside, foot soldiers from the city, *milites*, *sergentes*'.

- The *sergentes* belong with the *milites*, not with the foot soldiers, and thus were part of the cavalry.

This is inductive reasoning: it is clearly not inconceivable that the list was, for example, written in an illogical way, or that it was based on a different system of ordering than we suppose. But the premises do support the conclusion to some extent; and, moreover, the premises are in themselves believable. The first premise does fit well with what we know of how people generally write and think, while the second premise is an observation done by Hoppenbrouwers (or by another historian who delved into the archive).

Now it is the case that we can strengthen the conclusion of an inductive argument by giving more arguments that lead to the same conclusion; at least, we can do this if these arguments are all believable and independent of each other. Hoppenbrouwer indeed gives more arguments than just the one based on the list. Here is a another:

> In addition a provision of the important Statute of Todi of 1275 [...] states that *sergentes* had to receive the same pay as *milites* [...] These sources are discussed in more detail below, but at this point it is clear that they all indicate that in the communal army of Todi *sergentes* were part of the cavalry, not the infantry. (p. 247.)

When we write out this argument in more detail, it starts as follows:

- Soldiers who are held in higher esteem are generally paid more.
- The sources show us that the *milites* and the *sergentes* were paid the same amount.

- *Milites* and *sergentes* were esteemed about equally.

We can use the conclusion of the first argument as a premise in the second:

- Cavalry is held in higher esteem than infantry.
- *Milites* (cavalry) en *sergentes* were esteemed about equally.
- All soldiers are either cavalry or infantry.

- The *sergentes* were cavalry.

With that we have reached the desired conclusion again, but by using different sources and assumptions than the argument which was based on the order of the list.

We should notice several things. First, that arguments are rarely written out in full detail in the scientific literature. It would be time consuming to do that, and since many assumptions and steps of reasoning are bound to be clear to the careful observer, it would also be boring for both the reader and the writer. It is however the case that arguments can often be worked out in different ways. For example, I have worked out Hoppenbrouwers's argument in terms of *esteem*; but you could also argue that cavalry got paid more than infantry because they are more *effective*, or because of the higher *costs* associated with maintenance of a horse. Perhaps it is immediately clear for a scholar of medieval history which of these arguments is best; or perhaps they are all valid, and that is why it is not important to make clear which exact line of reasoning has been followed.

A second thing we might notice is that it is not always perfectly clear whether an argument is meant to be deductive or inductive; and that it is often possible to change one type into another. This is an inductive argument:

- Soldiers who are held in higher esteem are generally paid more.
- The sources show us that the *milites* and the *sergentes* were paid the same amount.

- *Milites* and *sergentes* were esteemed about equally.

since the fact that something is true "generally", obviously leaves open the possibility that this case happens to be different. I could also have reconstructed the argument slightly differently, with "always" instead of "generally":

- Soldiers who are held in higher esteem are always paid more.
- The sources show us that the *milites* and the *sergentes* were paid the same amount.

- *Milites* and *sergentes* were esteemed about equally.

This is a deductive argument. If the premises are true, then the conclusion must be true as well. The flipside of this is that the first premise has become less probable. It is likely that higher status *generally* goes hand in hand with higher pay, but one can doubt whetyher this is *always* the case. It is for example imaginable that noblemen were sometimes expected to fight without compensation because their position imposed this duty on them. ('Noblesse oblige.')

When you read an argument as a deductive argument, the argument becomes stronger, because the conclusion follows from the premises with necessity, not just with probability. On the other hand, the premises themselves become less probable, as we have seen in the example above. Since we are often unable to justify the premises needed to set up a deductive argument, we are often forced to reason inductively. We shall return to this phenomenon in a later chapter, when we discuss the use of laws in the study of history. ("Soldiers who are held in higher esteem are always paid more" is an example of a law: a general claim is made about how two things are always and invariably connected.)

A third things we notice is that arguments can form 'chains'. First we argued that *milites* and *sergentes* were held in similar esteem and then we used that as a premise for the next argument. When we link arguments in this way, very complex lines of reasoning can be constructed and it can end up being quite tricky to figure out what the premises are. The example that we have just talked about takes up only two sentences in Hoppenbrouwers's article; if we were to analyse the entire article, we would end up with some very complex chains of reasoning. For instance, Hoppenbrouwers combines the conclusion that *sergentes* were cavalry with, among other things, detailed statistics of levied taxes and ideas about what percentage of households could afford a horse, to ultimately show that the population of rural Todi was about twenty-five percent larger than previously assumed by historians.

Let us imagine what will happen to this scientific insight in the future. A historian could use this new estimate of the population of Todi in a study about population growth of Italy. The conclusions of this study follow could be used by a scholar who wishes to explain why the cultural changes that we call the 'Renaissance' started in Italy before spreading to the rest of Europe. This theory about the Renaissance would then be based (to some small extent) on the fact that a manuscript in Todi states that *milites* and *sergentes* should be paid equally – but this third historian is probably unaware of this. And what may happen in the future has already happened very often in the past: the presuppositions that Hoppenbrouwers makes in the article, for example about the development of social and military structures in Italian states throughout the ages, go back to an immense number of sources which no scholar could ever take into account in their entirety, or even trace back to the historiographical literature. To put it differently, a fourth conclusion that we can draw is that the relationship between our observations and the theories that we develop on their basis is often highly indirect and tremendously complex.

In this process every scientist is forced to base himself on what his predecessors have observed and the conclusions they drew from it; which means that science is for a large part not based on what we have observed and have come up with ourselves, but on the trust we bestow upon *other scientists*. We will examine the roles of trust and criticism in science, and how these have to keep each other in balance, at length in a later chapter.

All these points are actually pointing towards the same conclusion: that it will not be easy to figure out how believable a scientific theory really is. In order to do that we would first have to give a detailed reconstruction of the arguments in its favour, even though these are rarely given in detail and even though such a reconstruction is often a matter of interpretation. On top of that, the arguments are often very complex and we reach the initial observations only a great number of steps, many of which the scientists themselves are not explicitly aware of.

We should therefore not expect too much from a theory of confirmation, at any rate not that it provides us with an equation that we can simply apply to any scientific statement to calculate whether or not it should be accepted. The ability to properly judge a scientific theory is perhaps more an art than a science; in any case it is something for which great affinity with and experience in the field is necessary, so that you know what assumptions are reasonable under which circumstances; and it always includes a lot of uncertainties, if only because you cannot check the work of all the scientists on who's ideas you are building.

2.8 Conclusion

In this chapter we have formulated the question of the scientific method: under what circumstances is a scientific theory believable? The naive empiricist view is that theories are believable because they are based on observations; and that using this objective basis of observations will allow scientists to reach consensus and the truth.

In working out this standard view we have made a distinction between the empirical basis (statements which can be tested directly by observations) and theoretical statements. Theoretical statements have to be connected to the empirical basis by deductive and (most often) inductive arguments, so that they can be confirmed or disconfirmed by our observations.

Clarifying the scientific method thus consist in determining what statements are part of the empirical basis, and explaining how confirmation works. But as soon as we start to think about this we run into trouble. Observations turn out to be theory-dependent. That makes it difficult to hold that they form a complete objective basis that will always lead to consensus. And scientific arguments turn out to be complex and often extremely indirect; so complex and indirect, in fact, that in practice we can rarely find out on exactly which observations our theories are based.

That is no reason for methodologists to throw in the towel. But it invites us to be somewhat modest. We can probably be happy if we are able to succeed in clarifying the general structure of proper scientific thought. A simple checklist with which we can determine whether or not an academic article or book is good or bad is something we cannot expect – for that science is simply too complicated and subtle.

In the following chapters we shall delve deeper into ideas about confirmation by following the way they developed in the philosophy of science of the 20^{th} century. Ultimately this will bring us to a reevaluation of both the idea that statements have a wel-determined empirical content and the idea that science is objective and leads to consensus. The naive view of science will turn out to be naive indeed.

Chapter 3

Methodology: confirmation en falsification

The standard view of science that we described in section 2.3 tells us that scientists start out from a shared basis of objective observations. They then test their theories against these observations. Because scientists can agree about the basis, and also about how theories should be tested, they are able to reach consensus. Furthermore, we will steadily come closer to the best theory, that is, the theory that fits best with all the observations we could possibly make.

If we wanted to defend this view of science – but also if we wanted to attack it – it would be necessary to say something about how scientific tests work. How do we use our observations to check whether our theories are on the right track? Several answers to this question are possible.

In this chapter we will first look at the attempt to give a theory of *confirmation*, that is, a theory that tells us when and to what extent scientific theories are supported by the observations. Then we will discuss Karl Popper's highly influential, but in the end unsuccessful, attempt to avoid the problems surrounding confirmation by arguing that science is not about supporting theories at all, but about refuting them; not about confirmation, but about *falsification*.

3.1 Confirmation by instances

'Confirmation' is the technical term for what happens when observations support a theory; 'disconfirmation' is what happens when the observations undermine the theory. Which observations support or undermine a theory, an to what extent, is that a *theory of confirmation* wants to tell us. In this section we will look at a very simple form of confirmation. Even when we confine ourselves to this simple form – the confirmation of a general statement by the 'instances' of that statement – the problems of confirmation will become apparent.

With a *general statement* I will mean a statement that tells us something about all the members of a particular kind. Examples are "all plays of Shakespeare are good," "all speakers of Dutch use the word *boom* to refer to a tree," and "no Roman emperor before Constantine the Great was a Christian." Sometimes it is not immediately clear that a statement is a general statement. "Constantine the Great was the first Christian emperor of the Roman Empire" doesn't look like a general statement, but is in fact identical to the claim that all emperors before Constantine were non-Christians, which *is* a general statement.[1]

The *instances* of a general statement are the specific cases that the statement talks about. If the general statement is "all plays of Shakespeare are good," then some of the specific instances are "*Hamlet* is good" en "*As You Like It* is good."

It seems evident that we can confirm a general statement by its instances. The more instances we see, the more believable the general statement becomes. As we see more and more plays of Shakespeare, all of which turn out to be good, we will also become more and more confident about the claim that all of his plays are good. If we perform an experiment ten times and get the same result every time, we will slowly come to the conclusion that this result will happen every time. And so on. Reasoning from a few observations to the general statement that these observations are instances of, is something that we do quite often and that is very important in science.

We have thus found a rule for confirmation: as we observe more instance, the general statement becomes more believable.[2] But this rule is not very useful if it cannot be made more precise. For it doesn't tell us *how* believable a general statement is, given certain observations. And at the end of the day, what we are really interested in is deciding where any given theory should be placed on the spectrum ranging from 'utterly speculative' to 'based on very solid evidence.'

A serious theory of confirmation, then, will attempt to answer the following question: how many instances must we observe before we are allowed to have confidence in a general claim. If we could give a clear answer to that question, then we'll have really achieved something. For then we would be able to take any general statement made by scientists, compare it to the evidence, and decide whether there are enough positive observations for us

[1] With a little ingenuity, it may be possible to rewrite any claim as a general statement, though sometimes the end result will be very unnatural. For instance, "Victor is a philosopher" is identical to "All not-philosophers are not Victor."

[2] Even this rule is not unproblematic, but we won't go into that topic in this text. If you want to know more, you can look for information about Carl Hempel's raven paradox, Nelson Goodman's 'new riddle of induction,' or even the classical problem of induction associated with David Hume. Or you can consider this example: would the general statement that all human beings are less than 4 metres tall be confirmed by observing somebody who is 3,98 metres tall?

to believe the theory. Scientific consensus could always be easily reached. But, alas – as soon as we consider this question in more detail, we will come to understand that a simpel answer to it cannot be expected.

Suppose that you are a linguist interested in determining the grammatical rules of the English language. For this purpose it will be useful to know the truth or falsity of general statements of the form "all competent speakers of English hold that X is a correct English sentence." To see whether such statements are true, it will be useful to take competent speakers of English, show them certain sentences, and ask them whether or not they believe them to be correct sentences.

Suppose you come to be and show me the sentence: "Trees tend to be greener than cows." You ask me whether this sentence is grammatically correct. I tell you that it is. Wouldn't that be enough justification for you to conclude that all competent speakers of English will hold that this sentence is a correct English sentence? One single observation has a lot of weight in these circumstances.

Of course, I'm not saying that you can immediately be one hundred percent certain that all competent speakers of English will agree with me, or that further research is unnecessary. There are after all sentences about which competent speakers disagree; for instance, "You are better than me." (Some would insist that the correct sentence is "You are better than I." Other think that is excessively formal.) But in most cases competent speakers of a language *do* agree about which sentences are correct and which are incorrect. So when I claim that a sentence is correct, you immediately have good – though not conclusive – evidence that all competent speakers of English will share my judgement.

But in other cases a single observation is absolutely insufficient for allowing us to believe a general statement. Suppose that I see Picasso's *Femme aux Bras Croisés*, a painting made in 1902 that is mostly blue. Can I conclude from this observation that all paintings of Picasso are mostly blue? Of course not. Even less would I be justified in concluding that all paintings made in 1902 are mostly blue, or that all paintings ever made are mostly blue. I would have to see lots and lots of blue paintings, made by different painters working in all kinds of periods and cultures, before that last conclusion were to become even slightly probable.

Apparently, we cannot expect a simple answer to the question how many instances you have to see before you are justified in believing a general statement. In some circumstances, you would need to see many; in other circumstance, only a few. What we can do, however, is enumerate several of the factors that determine whether we require a lot of evidence or only a little for any given general statement. Here are four important ones:

1. *The expected amount of uniformity.* We know that some things tend to resemble each other: for example, the judgements of native speakers

about whether or not a certain sentence is grammatically correct. We also know that other things tend to show little uniformity: painters tend to paint pictures with different dominant colours, often change their style during their career, and attempt to produce works that differ from those of their predecessors and contemporaries. In addition, we expect works of art from different cultures and from different periods in history to be radically different. So before we see any particular evidence, we are already convinced that there will be more uniformity among grammatical judgements than among artistic colour choices, which means that we will require less evidence when talking about the former.

2. *The scope of the general statement.* In a sense this is the same point as the previous one, but it will be useful to also formulate it in a different way. The broader the scope of a general statement – that is, the more things the statement is about – the more evidence we will want to see before believing it. A statement about all paintings made by Picasso in the period 1901–1904 is easier to believe than a statement about all paintings made by Picasso irrespective of time period; and if somebody makes a statement about all paintings ever made, we will be even more sceptical and require even more evidence. Obviously, this is because we expect less uniformity when the scope of a statement broadens: it is more probable that Picasso always used the same dominant colour than that all painters in the history of the world used the same dominant colour; and it is even more probable that Picasso did this during a small period in his artistic development. (The period 1901–1904 is indeed known as Picasso's 'blue period'.)

3. *How representative the instances are.* It is not just the scope of the general statement that is important, but also how representative we believe that the observed instances are. If the linguist happens to interview only people living in Australia, it is possible that she ends up studying the grammar of Australian English rather than that of English in general. If she wants to draw conclusions about English grammar in general, she will need to make sure to investigate people from different countries and regions; and also people of different ages, levels of education, and social backgrounds. If an art historian only considers works from the period between 1901–1904, we won't be tempted to draw conclusions about the entire oeuvre of Picasso. For that, we would demand observations from different phases of his artistic development. In the same way a historian who makes claims on the military organisation of Europe in the high middle ages will be criticised if all his data are from Scotland. So we do not only care about the amount of evidence, but also about how diverse it is, and whether

we believe that this diversity of the evidence represents the diversity of our subject.

4. *The probability of the general statement.* Another important consideration is how probable we believe the general statement to be before the evidence comes in. Some statements seem quite probable, and we need little evidence to be convinced of their truth; while others are very implausible and we would need a lot of evidence before we came to accept them. Suppose that an archaeologist wants to convince us that the prehistoric Bell-Beaker culture was entirely pacifist by showing that no skeletons from that period show any evidence of a violent death. Presumably, we would consider this theory as rather improbable – very few, maybe no, truly pacifistic cultures are known – and we would need a lot of evidence before we started to believe the theory. On the other hand, suppose that the archaeologist launches the claim that the Bell-Beaker culture didn't have the technology required to make and place metal artificial hips, and wants to convince us this by showing us skeletons that do not have metal hips. Then we would be convinced rather quickly – indeed, we would believe the entire investigation to be pointless, since we already believe the theory before have seen any evidence at all!

3.2 Confirmation and theory-ladenness

The four factors we just enumerated have something in common: in each case, we use theoretical convictions that we already have in order to determine how much the observations support the theory under consideration.

How much uniformity do we expect? That depends on the ideas we already have about the subject under investigation. We believe that colour use will be less uniform than judgements about grammaticality, because we know that a painter can choose to use a different colour whenever he feels like it, while we also know that learning a language consists to a large extent in learning to use it in the same way that others do.

How representative is the evidence? To determine that, we already need to have an idea of the kind of variation that tends to happen in the domain under investigation. In the case of Picasso's paintings, we expect variation through time, since we know that the output of artists can often be divided into recognisable time periods. In the case of grammatical sentences, we expect geographical variation, since we know that words and idiom are often restricted to a specific region.

How probable is the general statement before we have done our observations? That too obviously depends on our earlier convictions; for instance, our conviction that neolithic cultures were unable to perform difficult feats of surgery.

In other words: when we determine how much support certain observations give to a theory, we don't do that in a theoretical vacuum. On the contrary, we use convictions that we already have. This means that the relation between observation and theory is more complex than the standard view of science suggested. It is not the case that we have a theory, we test it against the observational evidence, and then we know how probably the theory is. In order to test a theory against the evidence, we are always already using a whole set of theoretical convictions that we need to determine how relevant the observations are and how much support they give to the theory. Testing a theory always depends on theories we have accepted earlier; confirmation always takes place within a *theoretical framework*.

We started with the naive idea that theories could be derived from the observations using induction. That is, we thought that reasoning in the sciences goes roughly like this:

- We have made 50 observations of an A having property B.
- We have never observed a counterexample.

- All As have property B.

where I have chosen the number 50 more or less at random, and where the argument is obviously inductive. All the premises in this argument describe observations that we have done or not done. If the theory-ladenness of observation (see section 2.5) turns out not to be a big deal, then this type of argument would give us a relatively simple route from observation to theory.

But as soon as we start thinking in more detail about how confirmation works in actual practice, we understand that the kind of reasoning that goes on in science is more like this:

- We have made 50 observations of an A having property B.
- We have never observed a counterexample.
- It is likely that As are quite uniform with respect to property B.
- The observations we have done are representative for all As.
- That all As have property B is already a relatively plausible hypothesis.

- All A have property B.

and that is an argument in which many of the premises do not simply describe our observations. Instead, they are part of a theoretical framework within which I can test my theory.

The standard view of science told us that we start out with observations, and then use these to test our theories. It now appears that this view is mistaken, since all testing presupposes theories. But this raises an uncomfortable question: where did we get those theories? Perhaps they too were tested on the basis of earlier theories – but then, what about the first theories we ever used for this purpose? Did we just make them up? Wouldn't that undermine the reliability of science? That would be a rather unwelcome conclusion.

The idea that scientists can reach consensus also starts to look problematic. If theories can be tested against the evidence directly, then different scientists should be able to reach the same theoretical conclusions, provided that they can agree about the observations. That doesn't seem a particularly steep hurdle. But if theories can only be tested using other theories, then reaching consensus might be impossible. If you and I have different theoretical commitments, then we could also arrive at different judgements about new theories – perhaps because we have a different appraisal of the initial plausibility of the theory, or a different judgement about how representative the evidence is. And that means that it is possible that acquiring more evidence doesn't bring our scientific ideas closer together, but instead drives them apart.

Do such unbridgeable differences of opinion really occur in the sciences? If they do, what does that mean for the objectivity and reliability of science? Can we think of a way to bridge the apparently unbridgeable, or are we doomed to remain trapped in our own theoretical frameworks?

These questions will receive much attention in the next chapter. We will see that in trying to answer them, important elements of the standard view must be given up. But first we will look at an influential attempt at circumventing the problems of induction: Karl Popper's falsificationism.

3.3 Falsifiability

In George Eliot's (pseudonym of Mary Ann Evans; 1819–1880) beautiful novel *Middlemarch* (1872), we meet a certain Mr. Casaubon. This rather dusty scholar has devoted his life to writing a book titled *Key to All Mythologies*, in which he will explain the origin and meaning of all myths from all cultures using a grand overarching system. The reader starts to have doubts about the quality of this work rather soon, but halfway through *Middlemarch*, in chapter 48, we finally hear from the omniscient narrator how worthless Casaubon's ideas really are. It is noteworthy that the problem is *not*, according to the narrator, that his theories are *false*:

> Doubtless a vigorous error vigorously pursued has kept the embryos of truth a-breathing: the quest of gold being at the same time a questioning of substances, the body of chemistry is prepared for its soul, and Lavoisier is born.

There is, the narrator suggests, nothing wrong with passionately developing a theory that will, in the end, turn out to be false. The attempts of the alchemists to turn base metals into gold were a failure, but they were also a success, since they led to modern chemistry and the discoveries of the great chemist Lavoisier. So what *is* wrong with Casaubon's *Key to All Mythologies*?

> But Mr. Casaubon's theory of the elements which made the seed of all tradition was not likely to bruise itself unawares against discoveries: it floated among flexible conjectures no more solid than those etymologies which seemed strong because of likeness in sound until it was shown that likeness in sound made them impossible: it was a method of interpretation which was not tested by the necessity of forming anything which had sharper collisions than an elaborate notion of Gog and Magog: it was as free from interruption as a plan for threading the stars together.

The phrases that Eliot is using here – "not likely to bruise itself," "flexible conjectures," "not tested," "free from interruption" – all indicate one thing: Casaubon's interpretations are so flexible that they can never come into conflict with the evidence. Nothing could possibly show that his theory is wrong. Whatever may be discovered about ancient myths, he will always be able to fit these discoveries into his system. And it is exactly that, the utter invulnerability of his theory, that makes it completely worthless. The quest of the alchemists failed, but that failure taught us something; Mr. Casaubon's quest cannot fail, but for precisely this reason it is doomed to be fruitless.

Eliot thus presents an idea which would make the Austrian philosopher of science Karl Popper (1902–1994) famous half a century later: the idea that a good scientific theory has to be *falsifiable*. A theory is falsifiable just in case the theory can be tested in such a way that it might turn out to be false; in other words, when we can think of some way in which the theory could be disproven. When a theory is disproven, we say that it has been *falsified*.

Popper develops the idea of falsifiability in an attempt to clarify the distinction between real science and pseudo-science. In section 2.3, I asked what makes the theory of an astronomer (whom we would normally call a scientist) better than the theory of an astrologer (whom we would normally call a pseudo-scientist). According to Popper, the difference is not, as the naive view of science would suggest, that the astronomer's theory is based on careful observations and the astrologer's theory is not. The difference is that the astronomer makes predictions that are so precise that her theory might be falsified, whereas the astrologer makes predictions that are so vague that his theory can never be falsified. For instance, the astronomer might use a theory about the solar system to calculate that there will be a solar eclipse on the 23^{rd} of September at precisely 14:15. This prediction allows us to test the theory: if the eclipse doesn't occur, then the theory must be false. So the astronomer's theory is falsifiable.

What does the astrologer predict? In order to find out, I went to the website of the Dutch newspaper *De Telegraaf* and opened my horoscope for that particular day.[3] I read:

> You are more than usually motivated to make a positive change concerning your health and fitness. Distance yourself from people who want to exploit you. Say no to responsibilities that weigh too heavily.[4]

Are these falsifiable prediction? The last two sentences are words of advice, not predictions, which means that they cannot be tested. Furthermore, they are trivial: of course you should say no to something that weighs *too* heavily and of course you should avoid being exploited. The first sentence does have the form of a prediction; but it is a prediction that is extremely hard to test. How do you measure motivation? Not by checking whether I actually act on the motivation: is it possible that I am *more than usually* motivated to do a work-out today, but not motivated *enough* to actually do it. And if I spend the entire day lying in bed guzzling beer and eating candy, it is still possible that I was more than usually motivated to make a positive change concerning my health... but that, alas, I was also more than usually motivated to indulge in my sinful habits of laziness and gluttony, and that the latter motivation trumped the former. Whatever I do, it will never falsify the prediction. The theory of the astrologer – if these predictions were the result of a theory, and not just of random imaginations – is not falsifiable.

Of course, few people believe that horoscopes are serious science. But Popper believes that within the sciences people also use theories that turn out, on closer analysis, to be unfalsifiable. When he develops his ideas in the 1930's, the examples that are foremost in his mind are Freudian psychoanalysis and the political and economic theories of Marx.

What these theories have in common is that they introduce a framework of concepts and ideas that allow you to explain observed events – the actions, thoughts and dreams of a patient in the case of Freud, the big events in world politics and the world economy in the case of Marx. At first sight these theories appear to be very powerful. Freud can trace back all kinds of behaviour, not matter how varied, to sexual tendencies that were repressed in childhood. Many different phenomena can be explained within the system. But, Popper tells us, this power is only apparent; or, rather, the theory

[3]What does it say about our society that the horoscope is in the section of the newspaper called 'Woman'? That there is such a section at all? And –most damningly, in my opinion – that there is no section called 'Man'?

[4]Telegraaf.nl, 23 September 2013. My translation.

is *too* powerful. Whatever behaviour somebody exhibits, I can *always* fit it into psychoanalytic theory – and that is precisely the theory's weakness.

When someone indulges in excessive sexual activities, the explanation is that his "id," the part of the human mind where the bodily desires are located, is not sufficiently reigned in by his "superego," the repressive and censoring part of the mind. If someone is extremely uninterested in sexuality, the explanation is that his superego has become too strong and inhibits the id to a very high degree. If somebody engages in sexuality to a normal degree, that is because the id and the superego are in balance. It doesn't matter what behaviour somebody engages in: Freud's theory will always give me the concepts to explain it.

But that, Popper tells us, is exactly why these are not real explanations. A good scientific theory makes predictions that are testable; and Freud's theory never makes testable predictions. Psychoanalysis looks like a science, but it's not really a science. Psychoanalysts can keep producing new theories without any risk of being proven wrong; their work is "free from interruption," to use Eliot's words. While Freud was undoubtedly much smarter and much more original than poor Mr. Casaubon, neither of them was a scientist. Their theories are pseudo-scientific.

To determine whether Popper's analysis of Freud is justified, we would need to discuss psychoanalysis in a much deeper and more thorough way. We won't do that, although it would be quite interesting.[5] Instead, we will first take a look at the logic of falsification; and then we will ask ourselves whether Popper's scientific method can be really applied in the sciences. At the end of our discussion of falsification, I will come back to the important supposition that we saw Popper making, namely, that there is something wrong with sciences that do not make predictions.

3.4 The logic of falsification

According to Popper, good scientific theories are always falsifiable: they could be disproven by certain observations. In addition, he believes that good scientists actively search for falsifications. A good scientist does not attempt to *prove* her theory, since you can always find proof if you are looking for it. No, a good scientist tries to *disprove* her theory. Scientists are always critical; they never accept a theory as the truth, but search for ways to undermine existing theories; and precisely that is what scientific progress consists in.

I argued in section 2.6 that scientists almost always use induction, not

[5] Freud is a very engaging writer who has had a massive impact on the way we think about psychology. How much of his work is of lasting scientific value is a deeply contentious issue, with many contemporary psychologists having quite negative opinions about Freud. Unfortunately, many discussions about this topic are more ideological than insightful.

deduction. The argument was that scientific theories always make claims that go beyond the current observations, which means that the current observations cannot conclusively prove their truth. So scientists have to use induction if they want to move from observations to theories.

Popper is more or less unique among philosophers of science in disagreeing with this position. He argues for the opposite conclusion, namely, that scientists only use deduction. We can understand why he would say that by thinking about the logic of falsification.

In the process of falsification, the following steps are taken: first, an observation is derived from a theory; then, it is observed that the theoretically expected observation does not fit with what is really observed; and, finally, it is concluded that the theory must be false. As a logical argument, it looks like this:

- If theory T is true, we should observe W.
- We do not observe W.

- Theory T is false.

A famous example from the humanities is the falsification of the theory that the *Donatio Constantini* was signed by emperor Constantine the Great in the 4^{th} century. In this document, Constantine tells the reader that he donates half of the Roman Empire to the Church. As you will understand, it was a document that the Church liked to refer to whenever it wanted to assert its worldly power.

There had been doubts about the authenticity of the *Donatio* for a long time, but in the 15^{th} century philologists like Cusanus and Valla gave very strong arguments that these doubts were justified and that the document was a medieval forgery. They gave arguments like this:

- If the *Donatio Constantini* was written by Constantine the Great, then it will not contain words that were not in use in the 4^{th} century.
- The *Donatio Constantini* does contain such words.

- So the *Donatio Constantini* was not written by Constantine the Great.

This is a falsification of the claim that the *Donatio Constantini* was written by Constantine. What we should notice is that this argument is – at least at first sight – a *deductive* argument. The theory makes a prediction; the prediction is not in accordance with our observations; and so the theory has to be false. If the premises are true, the conclusion must follow. There is no other possibility.

When we think back to earlier parts of this chapter, we will understand why Popper is very happy with this result. In order to confirm a theory, we need to use induction, but we are then faced with the problem that we need a theoretical framework to determine how much support the evidence gives to the theory. As we have seen, this leads us to the weird conclusion that we always already need to accept theories before we can test theories.

But to falsify a theory, we only need deduction, and that is much simpler. If my theory predicts that a certain word will not appear in a certain text, and the word does appear in that text, than my theory must be false. I don't need a theoretical framework to draw that conclusion. In other words, when scientists are engaged in falsification, they *can* test theories directly against the evidence. And this means that they can reach consensus: consensus about which theories have been falsified.

If we accept Popper's thesis about the nature of science – his *falsificationism* – then we can entirely circumvent the problems developed in section 3.2. That sounds almost too good to be true.

3.5 How critical are scientists?

In the last section we saw that, according to Popper, scientists should be easily able to reach consensus about which theories have been falsified. But suppose that we have already attempted to falsify a theory many times, but that the theory has withstood all these attempt; does that mean that scientists should be allowed to conclude that the theory is probably true?

Popper emphasises again and again that this is not the case and that scientists do not in fact do this. To conclude from failed attempts at falsification that the theory is probably true, would be a kind of confirmation; and one of Popper's central claims is that scientists never engage in confirmation, but are only interested in falsification. This is not a claim he could easily give up: it is essential to his methodological position. After all, if scientists use confirmation, then they need a theoretical framework, which would reintroduce all the problems of section 3.2. In that case Popper's theory would be of little help to us.

According to Popper, scientists never claim that a theory is true or probably true. He believes that scientists make only two kinds of judgements about scientific theories: either the theory has been falsified, and is therefore untrue; or it has not been falsified, which means that we do not know whether it is true or false. We can never go further than that.[6]

[6] Popper does introduce a technical term, 'corroborated,' to speak about theories that have been tested many times but have never been falsified. On the one hand, he consistently holds that corroboration is very different from confirmation; on the other hand, he claims that it is rational to make predictions using theories that have been well corroborated. Why that would be rational, if no confirmation has taken place, has always remained mysterious for non-Popperians; I will not go into it. A very good article on this subject is Wesley Salmon, "Rational Prediction," *British Journal for the Philosophy of Science* 32 (1981), pp. 115-125.

Is is truly the case that scientists never claim that their theory is well-supported and probably true? Does our scientific knowledge consist of nothing but the knowledge that certain theories are *not* true? At the very least, that is not how scientists talk. A historian might tell us that Kievan Rus' was conquered by the Mongols under Batu Khan in the years 1237–1240. He will think of this as an uncontroversial claim, as something that we know more or less for certain; hence not as a claim of which we know no more than that it ha not yet been falsified.

If we consider what a scientific education in a particular field consists in, we notice that an important part of it is learning about many theories that are almost universally accepted by those working in that particular field. It is not the case that we are only taught which theories have been falsified; indeed, frequently only little attention is paid to theories that are no longer accepted.

When scientists make practical predictions, they also generally seem to be quite certain. The astronomer we talked about earlier, who uses her theories of the solar system to predict an eclipse, does not think of this prediction as particularly doubtful. She will be quite convinced of its truth. And would anybody enter an aeroplane of which the designer wants to say no more than that the theory that it can fly "has not yet been falsified"?

Even in science – which Popper tells us is an enterprise based on unrelenting criticism – many theories are held to be so secure that nobody believes that we should actively test them. A biologist trying to get funding for research that attempts to falsify the theory of evolution, will not be able to get any money. (At least not from the usual sources.) An art historian who suggests doing research into whether the painter Karel Appel really lived in the 20^{th} century, or whether he might instead have lived in the 14^{th}, will be laughed away. Some ideas have been established by so much evidence that we cannot seriously doubt them.

But if that is true, then Popper's falsificationism cannot be correct as a description of science. Apparently, scientists are not only critical; apparently, they do believe that their theories can be confirmed by the evidence. As far as the descriptive task of the methodologist (see section 2.1) is concerned, Popper seems to fail.

He might claim that his method nevertheless describes how scientists *should* act. If they don't live up to the ideal of falsificationism, well, so much the worse for them. Of course, this response only works if falsificationism is a coherent ideal – and we will see in the next section that it is not.

3.6 The logic of falsification revisited

In section 3.4 I presented the following reconstruction of an argument that falsified a theory:

• If the *Donatio Constantini* was written by Constantine the Great, then it will not contain words that were not in use in the 4^{th} century.
• The *Donatio Constantini* does contain such words.

• So the *Donatio Constantini* was not written by Constantine the Great.

But is this reconstruction correct? Haven't we neglected to mention several crucial presuppositions? To investigate this, let us attempt to list all the presuppositions that I will have to make in order to conclude that the *Donatio Constantini* was not written by Constantine the Great.

I will have to start by presupposing that certain Latin words – for instance, to remain historically accurate, the word 'feudum' – were not in use in the 4^{th} century. Then I will have to presuppose that the document in front of me is in fact the *Donatio Constantini.* I will also have to presuppose that Constantine the Great lived in the 4^{th} century; if he didn't, it doesn't matter which words were in use at that time. I will also have to presuppose that I have correctly deciphered the manuscript, that, for example, this piece of handwriting here really says 'feudum'. And I will have to presuppose that that piece of handwriting has always been part of this manuscript and has not been added later – after all, if someone changes to the *Donatio* centuries after it was originally written, that wouldn't mean that it had not in essence been the work of Constantine.

Perhaps I need to presuppose even more, but let's stop here. The philologist's argument, then, is as follows:

• The word feudum was not in use in the 4^{th} century.
• Constantine the Great lived in the 4^{th} century.
• The document I am looking at is the *Donatio Constantini.*
• This document is essentially in its original state; no additions have been made after it was written.
• I can correctly decipher the handwriting.
• If all of the above are true and the *Donatio Constantini* has been written by Constantine the Great, then it cannot be the case that I see the word 'feudum' in this document.
• I do see the word 'feudum' in this document.

• So the *Donatio Constantini* was not written by Constantine the Great.

This is still a deductive argument. But something has happened that means trouble for Popper. By developing the argument, we have seen that many premises had to be added. And not all of those premises describe direct observations; on the contrary, they are theoretical claims. You cannot simply observe that the word 'feudum' did not appear in the 4^{th} century, or that Constantine lived at that time, and so on. These are all theoretical claims that will have to be derived from more basic observations.

Popper could adopt one of two strategies. Either he claims that the philologist cannot know whether all of these assumptions are true; or he claims that the philologist can know whether they are true (or at least more likely to be true than to be false).

Suppose he chooses the first option: the theoretical presuppositions are simply hypotheses of which we do not know whether they are or not. In that case the philologist could not use the argument given above. From the fact that he sees the word 'feudum', he can conclude that *one* of his assumptions must be wrong. But he cannot know which one. Perhaps the *Donatio* was not written by Constantine. But perhaps one of the other presuppositions is false: this is not the right document, or the word 'feudum' actually was in use during the 4^{th} century, or Constantine lived later than we thought, and so on.

When I observe something that doesn't fit my expectations, then at least one of the hypotheses on which I base my expectations must be wrong. But the observation cannot tell me which of the hypotheses is to blame. Since almost every expectation is based on a large amount of hypotheses – as our example shows – this means that I'm generally in the situation where I know *that* one of my hypotheses is false, but that I do not know *which* one it is. And that means that I cannot falsify the theory I'm testing. After all, the fact that my expectation was wrong could be the fault of the theory, but it could also be the fault of one of my other hypotheses.

In brief: if Popper chooses to say that the philologist cannot know that his assumptions are true (or at least true with high probability), then he must also admit that the philologist cannot falsify his theory. There is no way to know whether the theory is mistaken or whether one of the other assumptions is.

But if Popper makes the opposite choice and tells us that the philologist can know that his assumptions are true, then he has an even bigger problem. As we remarked before, most of these assumptions are not statements that can be directly tested by observations; they are theoretical claims. And it is essential to Popper's methodology that scientists never accept theoretical claims as true, but always only try to falsify them.

For instance, how could the philologist know that Constantine the Great lived in the 4^{th} century? Certainly not by just observing that this is the case, since he cannot use time travel to go back to the 4^{th} century and see whether Constantine is anywhere to be found. It is a theoretical claim that has to be based on observations indirectly – perhaps observations of ancient documents and inscriptions.[7] But this means that in order to falsify his

[7] And further observations that help to establish the relation between ancient Roman

theory, the philologist will first have to establish his other assumption by using induction.

It seems, then, that falsification is only possible if we already have a theoretical framework of assumptions, since a theory alone is not enough to derive any observations. If the scientists doesn't have good evidence for the truth of the framework, he won't be able to falsify anything. But if he does have good evidence for its truth, well, that means that he is using induction to establish the truth of theories – which means that we have returned to the first half of this chapter. If so, Popper has not escaped from the problems of confirmation after all.

The fact that even Popper, perhaps the philosopher of science most eager to do without theoretical frameworks, cannot in the end come up with a methodology that actually does without them, is – or so I would like to suggest – strong confirmation for the claim that we need theoretical frameworks to do science.

3.7 Should theories be predictive?

Let us discuss one final criticism that may be levelled against Popper's falsificationism and that is especially relevant for the humanities. Popper claims that unfalsifiable theories aren't scientific but pseudo-scientific. This means that a theory which does not make predictions is automatically pseudo-scientific. If I don't make predictions, then nobody can ever make an observation that disproves my theory.

But is prediction really the aim of all scientific endeavours? Suppose that I want to develop a theory about the nature of the tragic in Shakespeare – I want to explain what all Shakespearean tragedies have in common and what Shakespeare's idea of the tragic is. And suppose that I want my theory to make prediction. Obviously, those predictions can only be about Shakespeare's tragedies, so, let us say, about *Romeo and Juliet*, *Julius Caesar*, *Hamlet*, *Othello*, *King Lear*, *Macbeth*, *Antony & Cleopatra* and *Coriolanus*.[8] How should I proceed?

Well, I obviously can't start by reading all the tragedies and then constructing a theory that fits them all. Once I've done that, no test is possible. I can't make predictions if I've already used all the evidence there is in the construction of my theory.

---

consular dating and our modern Anno Domini or Common Era dating, which is a 6^{th} century invention.

[8] Debate is possible about which of Shakespeare's plays should be counted as tragedies. Since our judgement about this issue undoubtedly depends on our theory of the tragic, an extra difficulty emerges. If one of Shakespeare's plays doesn't fit my theory of Shakespearean tragedy, is that because my theory is wrong, or because that play is not a real tragedy?

If I want to do Popperian science, I will have to do the following. First, I read only some of Shakespeare's tragedies; based on this reading, I construct a theory about the nature of the tragic in Shakespeare; from that theory I deduce predictions about the plays I haven't yet read; and then I test the theory by checking whether my predictions are correct. Only by following this procedure will I be able to make the kind of risky prediction that Popper wants me to make. Only in this way can I be a real scientists.

Evidently, this is ludicrous. The idea that a theory about the tragic in Shakespeare is more scientific if it is based on only a few of his tragedies, and less scientific if it is based on all of them, makes no sense. No Shakespearean scholar will adopt this procedure; just as no art historian interested in Rembrandt will attempt to remain ignorant of part of his oeuvre; and just as no historian will ignore part of the archival material so she cal later test her theory against it.

Are all of these sciences pseudo-sciences? It appears that Popper has failed to make an important distinction: a distinction between *predictive* sciences and *interpretative sciences*. Sometimes, scientists want to predict our future observations. But quite often, a scientist only wants to understand the material that he has already collected. We already have Shakespeare's tragedies; no new ones will appear; which means that we do not have to make predictions about them. But that doesn't imply that all our questions about them have been answered. Although we could, in theory, make predictions about authors who are still alive, scientists are not especially interested in this either. Interpreting the existing novels of Arnon Grunberg or Thomas Pynchon is something that serious scholars actually do; predicting what their next novels will be about is an activity more suited to the pub. There are many things that we do not immediately understand and that we want to think about in a scientific manner.

A. C. Bradley writes in his classic work *Shakespearean Tragedy*:

> We remain confronted with the inexplicable fact, or the no less inexplicable appearance, of a world travailing for perfection, but bringing to birth, together with glorious good, an evil which it is able to overcome only by self-torture and self-waste. And this fact or appearance is tragedy.[9]

This isn't meant as a prediction about Shakespearean tragedies. No, the ideas is that we can better *understand* Shakespeare, and perhaps can *appreciate* him more as well, when we read or see the tragedies with Bradley's theory in the back of our minds. That is where its value lies.

Such science, which seeks *understanding*, does not fit in Popper's vision; which gives us an extra reason to believe that his method cannot be adopted by scientists, certainly not in the humanities.

---

[9] Original edition of 1904/1905. These are the final two sentences of the first chapter.

3.8 Conclusion

In this chapter we tried to think through the idea that scientific theories are based on observations. When we consider confirmation, it turns out that it is only possible within a theoretical framework. This suggests that the relation between theory and observation is not simple: observations can only be used when we already have theories that can be used to interpret them. If we claim that "our theories are based on our observations", we are guilty of gross oversimplification.

Karl Popper's proposed alternative, falsificationism, is not, in the end, a real alternative. Falsifications also require a theoretical framework. When we test theories, whether through confirmation or through falsification, we always already have a background of theories ready.

In the next two chapters, we will explore the idea of theoretical frameworks in much more depth.

Chapter 4

Methodology: Kuhn and paradigms

4.1 A role for the history of science

In section 2.1, we made a distinction between the normative and the descriptive side of methodology. The scientific method is a *norm*; it tells us how science *ought* to work. The methodologist is thus looking for the ideal way for scientists to behave, even though we know that scientists will not always adhere to this ideal in practice.

At the same time, the norms discovered by the methodologist have to reflect actual scientific practice to a high degree. If the practice of scientists is completely different from the method given by the methodologist, then the latter may be telling us a nice story, but it is not very plausible that that story is the scientific method. So a methodologist always has to keep the *descriptive* part of his job in mind.

Since there are two sides to the scientific method, we can approach the project of methodology in two corresponding ways. We can start by thinking normatively about the right ways of reasoning, and then go on to see how such reasoning is used in scientific practice. This approach, which we could call the 'logical approach', has been followed by us in the two previous chapters. But we could also do the opposite. We could start by describing scientific practice, and then go on to distil normative methodological insights from that description. Here the starting point would be a historical or sociological description of the activities of successful scientists. I will call this the 'historical approach'.

In this chapter we will follow the historical approach. We will focus on the work of Thomas Kuhn (1922–1996), probably the most influential author about science of the 20^{th} century. Although Kuhn was very interested in philosophy and wrote many philosophical articles, his professional training was as a historian of science. Central to his work, then, is not some philosophical theory about logical argumentation, but the history of (natural) science. Kuhn observes patterns in the historical record that followers of the logical approach had not seen; and he believes that understanding these patterns is precisely what we need to understand how science actually works, and how it should work.

The logical approach automatically raises the question to what extent the proposed logical rules agree with actual scientific practice. The historical approach obviously raises the opposite question. How do we know that patterns in the history of science reflect the scientific method, rather than being just random occurrences that might equally well not have happened? And why would such historical patterns have any normative import? In other words, why would we want to behave just like the scientists of the past? Kuhn knows that it is not evident that history will teach us useful lessons for the present. And yet the opening sentence of his most important book, *The Structure of Scientific Revolutions*, is hardly modest about the importance of studying history:

> History, if viewed as a repository of more than anecdote or chronology, could produce a decisive transformation in the image of science by which we are now possessed. (p.1)

For Kuhn, the crucial insight we can draw from history is that science does not have the same properties – and does not use the same method – at every moment of its existence. Instead, the sciences develop in a more or less cyclic pattern in which dogmatic and revolutionary phases alternate. This would mean that the philosophers we discussed in the past two chapters, who tried to determine 'the' scientific method, where on the wrong track. There are different methods, one belonging to each of the different phases. And in fact, Kuhn will proceed to tell us, there are many more methods than that.

But before we look at Kuhn's theory of the phases of scientific development, it will be useful to look at his answer to a seemingly simpler question; the question, namely, how young scientists are taught to distinguish between good and bad science.

4.2 Exemplars

How is a young scientist taught to recognise good science? An assumption that we made in the previous two chapters was that the scientific method takes the form of *rules*, to be precise, rules that indicate which theories you are or are not allowed to accept given that you have made certain observations. According to these assumptions, the young scientist would need to learn these rules during her studies, until finally she has mastered the entire scientific method.

But this theory of scientific education faces two major problems. The first of these is that it is, as we have seen, extremely hard to define the exact rules of scientific reasoning. But if philosophers of science and the scientists themselves are unable to formulate the rules, then how do they teach them to aspiring researchers? To teach theoretical knowledge to someone else, it is surely necessary to first make that knowledge explicit and express it verbally. So it appears that we simply *cannot* learn the scientific method in this way.

The second problem is that the kind of methodological rules that we have looked at in the previous chapters can only help us with a very limited part of scientific research. We might be able to use the rules of induction to find out whether Hoppenbrouwers's arguments, which we discussed in section 2.7, are persuasive. But where did Hoppenbrouwer get the idea to investigate the social and military role of the *sergentes*, and not a different topic – for instance, what colour of horses the *sergentes* preferred? Why did he decide that to answer his question, he needed to go all the way to Todi and look at archival records? Why did he look at tax records? What could possibly give you the idea that tax records will allow you to answer research questions that are not, at least at first sight, about taxes?

Someone who believes in a strict distinction between the *context of discovery* and the *context of justification* (section 2.1), might be tempted to brush aside such questions as irrelevant. Where a researcher gets the idea to investigate a certain topic in a certain way is not important; what we should care about is whether the data justify the conclusion. But in scientific practice, such questions are very important indeed. If a young scientists was taught nothing but a method of inductive reasoning, he or she would have no idea how to start doing good and relevant research. It is an essential part of any scientific education that the students learns how to formulate good research questions; how to gather relevant data; how to use those data to draw conclusions that are as interesting as possible; and that the student is taught a theoretical framework in which all of this can be fruitfully done.

It seems, then, that the scientific method is much broader than just a set of rules for correct reasoning, and that it cannot be taught just by providing such rules. How can it be taught? Kuhn claims that scientific training happens primarily through acquainting students with 'exemplars', exemplary examples.[1]

Exemplary examples are examples of successful scientific research that scientists become acquainted with during their education, and that are presented to them as good examples to follow. According to Kuhn, learning to emulate examples is the most important way in which the scientific method is learned. After reading a large number of recent books and articles

[1] In *The Structure of Scientific Revolutions*, Kuhn calls these exemplary examples 'paradigms'. Since he uses the same term in the same book with a very different meaning – we will talk about it later on in this chapter – this has generated a lot of confusion. In his later works, Kuhn changes his terminology and introduces the term 'exemplar', which I will be using here.

in the field of history, the budding historian has developed a good sense of what counts as a serious historical research question and what kind of data you could try to gather in order to answer such a question. The historian may not be able to formulate explicit 'rules' for good research questions; but he will immediately understand that serious historians are going to be more interested in the social and military structures of medieval Todi than in the favourite colour of horses at that time. He will also know from many examples that tax records and other financial archives can yield valuable knowledge about many non-financial aspects of life.

If Kuhn is right, then the way students learn to do science is more like the way children learn to speak a language than like the way adults learn a theory. Children only rarely hear about the explicit rules of their language – not many parents will tell theirs babies about declension and conjugation, though some very enthusiastic linguists may try – but instead are bombarded with examples of correct speech. By emulating those examples, they become fluent in the language.

The idea that criteria can take the form of examples rather than of rules is not new. Kuhn points out the similarities between his own idea and the claim of the famous philosopher Ludwig Wittgenstein (1889 – 1951) that many words in our language get their meaning through family resemblance. Take the word 'game'. If someone were to try to formulate the rules that a thing has to follow in order to be a game, he is unlikely to succeed. Is a game an activity in which you try to win? Not necessarily, since there are games in which you cannot win (such as many collaborative roleplaying games) or in which it is impossible to win (such as the computer game *Space Invaders*, which becomes harder and harder until the player finally loses). Perhaps a game is something with a board and pieces? No, because domino and bridge are games as well. Etcetera – any rule you can think of will have exceptions. And yet we have little trouble recognising games. How is that possible?

Well, Wittgenstein tells us, to determine whether or not something is a game, we don't use strict rules; instead, we judge whether it resembles anything that we have already learned to call a game. Is curling a game? It resembles *boules*, which is a game, so curling is a game too. Is *Catan* a game? In some respects, it very much resembles *Monopoly*, which is a game, so *Catan* is a game too. And so on. All speakers of the English language have learned to associate approximately the same examples with the word 'game', which means that we can generally reach agreement about whether something is a game or not – although there will always be borderline cases. (Is Russian roulette a game?)

A scientific education works in the same way, according to Kuhn. By seeing many examples of good research in linguistics, the aspiring linguist learns to distinguish good from bad research. If we ask her to give arguments for her judgement about any specific piece of research, she will probably be able to give an explanation for her judgement, but not by invoking strict general rules.

If we think about scientific training in this way, the problems we identified disappear as snow in the sun. If scientists never explicitly use rules, and if there may not be any strict rules, then it is not very surprising that we can't formulate them. And exemplars form a much broader basis than rules about correct reasoning: they can also teach us what makes for a good research question, how we should gather data, and so forth.

However, an important and difficult question arises. If there are no rules that separate good from bad research, then how can we know which examples of science are the *exemplary* examples – the examples worth following? How do we decide which research is good and which is bad? Is the term 'good' in this context simply synonymous with 'as we do it in our current scientific community'? That would lead to a radically different view of science than the usual one. It would seem to undermine the normative status of the scientific method. Before we can understand Kuhn's position on these issues, however, we first have to look at his theory about the historical development of the sciences.

4.3 Paradigms and normal science

Kuhn's project is based on investigations into the history of science. In that history, he claims to four stages or phases in which scientific fields (such as linguistics or biology) or subfields (such as phonetics or ecology) can be.

These phases differ from each other in the role played in that phase by a *paradigm*. A paradigm is a collection of concepts, theories, techniques, questions, research methods (and whatever else may be important in scientific practice) that a specific scientific community accepts as unproblematic. The paradigm encompasses all the instruments, literal and figurative, that scientists in this community use in their daily work without having critical doubts about whether they really are the best possible instruments – just as a baker will simply use his oven without worrying whether our current techniques for building ovens are optimal.

The four phases that Kuhn distinguishes are:

1. In the phase of **pre-paradigmatic science**, no paradigm has been developed yet. Scientists do not agree which concepts, theories, and so on, should be used when doing research.

2. In the phase of **normal science**, all research in the discipline is done within the shared paradigm.

3. In the phase of **crisis**, scientists have doubts about the value of their paradigm, even though they still work within it. They are interested in developing alternatives.

4. In the phase of **scientific revolutions**, one paradigm is changed for another.

The phases follow each other in a set pattern. Every scientific discipline starts out in a pre-paradigmatic phase (phase 1), which is always followed by a phase of normal science (phase 2). Normal science leads to a period of crisis (phase 3). This crisis can be solved in two ways: either the crisis is resolved, or the crisis leads to a scientific revolution (phase 4). In both cases science returns to a period of normal science (phase 2); the difference between the two ways of getting out of a crisis lies only in the occurrence or non-occurrence of a revolution. A science that has risen above the pre-paradigmatic phase (phase 1) will never return to it: pre-paradigmatic science is a one time phenomenon, though this phase can last for a very long time.

The history of every scientific discipline has to start with a pre-paradigmatic phase. After all, all the concepts, theories and research methods still have to be developed, so there cannot be consensus about them in the scientific community. In this phase, everything must still be invented or discovered by the scientists; not only do they still have to do all the research, but they also still have to find out *how* to do that research. Which phenomena should a scientist investigate? What instruments, if any, should be used to measure them? What concepts should be used to describe them? What questions should scientists try to answer? Under what circumstances can we speak of a scientific success? The scientists will have to answer all those questions, and it is unlikely that they will immediately agree with each other about the answers.

As an example, let us look at two important figures in early Western historiography, the Greek historians Herodotus and Thucydides. The *Histories* of Herodotus and the *History of the Peloponnesian War* of Thucydides are in some respects very similar: both works describe an important war in what was then recent history, and they are evidently part of the same scientific field. Nevertheless, there are several big differences in the way approach history. Herodotus's interests are wide-ranging: he mixes his descriptions of political and military events with long geographical and ethnological treatises. His research method consists mostly in the gathering of oral reports; if these contradict each other, Herodotus tends to write them all down, often indicating his sources, and often without offering a judgement about their truth. Partly because of this, his book contains many legendary and otherwise unbelievable anecdotes. Finally, Herodotus believes that studying history teaches us important lessons of a moral nature.

Thucydides, who lived slightly later, disapproved of the method of his older contemporary. Where the *Histories* ranges far and wide, the *History of the Peloponnesian War* has a strict focus on military and political events. While Thucydides uses orally transmitted reports, just like Herodotus, he never tells us what his sources were, and he never gives us competing versions. Instead, he narrates the events in the way he believes them to have happened. He attempts to use only reports by eyewitnesses, which automatically means that he can only write about very recent history. An important part of his book are the many speeches given by important persons; Thucydides has written these himself, in an attempt to show what they must have thought and said. He dislikes moral lessons.

Evidently, Herodotus and Thucydides disagree about several fundamental questions. What should a historical work be about? How should you treat your sources? Is it legitimate to try to piece together the more distant past? Are you allowed to write down versions of the events that you yourself consider unlikely to be true? Are you allowed to put speeches you have written yourself into the mouth of the historical figures? Is the aim of historiography to draw moral lessons from history, or merely to tell us what happened? It is not the case that Thucydides is scientific and Herodotus is not, or the other way around. Rather, they have different ideas about how to do historiography. This, then, is an example of a pre-paradigmatic period in science: scientists are not yet in agreement about the basic framework within which their research has to be done.

Now Kuhn believes that lacking a paradigm is very unsatisfactory for scientists. Instead of being able to focus on their research, they have to spend much of their time critically thinking about the framework within which they do their research. They have to invest energy into debates – often quite hostile and unproductive – with other scientists who have chosen different frameworks. Furthermore, they will have trouble using the results of other scientists in their own research: if somebody has used different methods or measuring instruments from the ones you yourself use, it is always doubtful whether you can incorporate their results in your own studies. Even communication between scientists is difficult. As long as there are no agreed upon concepts, scientists will often encounter miscommunication and lack of mutual understanding.

All these problems of pre-paradigmatic science are solved as soon as a scientific community develops a paradigm. Such a paradigm will often come into existence around one or more scientists whose work is seen as an exemplary example and whose successes inspire others to follow in their footsteps. Once scientists share a paradigm, they can work together in a way which was previously impossible. They agree about the questions they want to answer; they are generally able to reach a broad consensus about the correct answers to those questions; and they can view themselves are engaged in the collaborative building of a bigger and bigger tower of knowledge about the aspect of the world they investigate. Normal science is therefore a period characterised by the achievement of concrete results that are widely accepted within the community.

According to Kuhn, normal science is the most common form of science, and it therefore exerts a large influence on our ideas about scientific method. We believe that scientists can achieve consensus and that the scientific community can develop concrete answers to important questions. Scientists do not constantly question the methods and central theories of their field. A biologist will not easily be tempted to wonder whether DNA really exists; a historian won't be racked by doubts about the scientific consensus that archival sources are generally more reliable than oral traditions; and a linguist will hardly feel the need to consider whether sentences really have something we can call grammatical structure. These are basic ideas that everybody in the field accepts. You have seen them again and again during your education. It would also be very unproductive to take a sceptical attitude towards the elements of your paradigm, since life is simply too short to critically investigate all the results of earlier science. And anyway, how would you investigate them? They themselves form the standards by which you decide whether a piece of research is good or bad.

So normal science, science within a paradigm, is science in which a certain amount of things is simply accepted and not the subject of discussion. At first sight, this may seem strange. Isn't science famous for being critical? Shouldn't it be anathema to science to accept any idea in a dogmatic fashion? Was Popper really totally wrong, when he claimed that science is all about testing and attempting to falsify theories? Well, Kuhn would say, he was indeed mostly mistaken – at least when we are talking about periods of normal science. A normal scientist is not in the business of falsifying theories, but of *solving puzzles*.

There are several ways in which scientific questions investigated within a paradigm can be fruitfully compared to puzzles. Puzzles are problems that we believe to have a solution; a solution, furthermore, that can be found within a framework of rules; and that we can recognise once it is found. A good crossword puzzle does not contain any incorrect clues; can be solved without adding new squares or putting more than one letter in a single square; and is solved if all our words fit the clues and fit together in the diagram. These same criteria hold for scientific problems in periods of normal science. We can formulate the problem clearly, because we have clear and well-defined concepts. We have a clear methodology that tells us which investigative moves are allowed and which are forbidden. (For instance, a historian is not allowed to solve a problem by fabricating evidence.) And scientists can generally agree whether a problem has or has not been solved.

A period of normal science, then, is a period during which scientists can solve puzzles without fielding any serious criticisms of the paradigm. Since more and more puzzles get solved, there is a clear sense of progress. And as long as scientists remain successful at solving puzzles, nobody will be much tempted to question the validity of the paradigm. But what if things don't go well? When scientists find themselves confronting seemingly unsolvable problems? Then, Kuhn tells us, we move to a new phase of science: a period of crisis.

4.4 Crisis and revolution

During periods of normal science, scientists rarely entertain critical thoughts about their paradigm. They have no reason to do so, since the paradigm allows them to solve problems efficiently. But, of course, scientists sometimes come across problems that they fail to solve. They fail to find an explanation for a particular phenomenon; or they make observations that do not fit their theoretical expectations. In such cases, Kuhn speaks of an *anomaly*, something that falls outside of the normal laws ('nomos' is the Greek word for law).

The existence of anomalies is not in itself a problem for a scientific field. It would be far too optimistic to expect that we can always solve every problem immediately. Where Popper wants us to reject our theory as soon as it doesn't fit the observations, Kuhn claims that scientists generally hold on to their theory and ignore its problems, at least for a time. After all, it is quite likely that we will be able to solve the problem at some later point, when we know more and have developed our theory further, without having to change anything about our paradigm.

But what *is* problematic, Kuhn tells us, is the situation where a field is faced with a growing number of anomalies; a situation that occurs when more and more new anomalies are found, while the existing ones are rarely if ever resolved. When that happens, scientists will start to worry that the framework within which they are doing science might no longer suffice. They start to doubt their own paradigm. Kuhn calls such a period of doubt a *crisis*.

During a time of crisis, scientists are much more interested in research that casts doubt on the reigning paradigm than they are during periods of normal science. Alternative concepts and methods get more attention; there is a proliferation of *out-of-the-box* thinking. The worse the crisis gets, the less scientists are still engaged in their normal activity of puzzle solving, and the more they are trying to find new frameworks within which they might pursue their investigations with more success.

No matter how bad the crisis becomes, scientists will never, according to Kuhn, give up their paradigm if there is no alternative available. Even a bad paradigm full of anomalies is better than no paradigm at all; since without a paradigm, the scientific community disintegrates, communication and co-operation become impossible, and science as we know it cannot continue. This means that scientists will never give up their central theories if they do not have a replacement. Obviously, this is radically different view of science from the one given by Popper, in which scientists are always ready to give up any theory. Kuhn even judges that abandoning your paradigm without having adopting a new one would be equivalent to "rejecting science itself."

But isn't that a problem? Wasn't the strong suit of Popper's account of science that Popperian scientists are never dogmatic? Kuhn develops an argument that attempts to show that we can only falsify central theories after a scientific community has held on to them for some time in a relatively dogmatic fashion. After all, if a theory gets into trouble, it is not immediately clear whether these problems might not be solved without giving up the theory. Someone who abandons his theories as soon as there is any trouble is being far too rash. Perhaps a little more research would have vindicated the theory. What happens during a period of normal science is precisely that scientists do their utmost to solve all the problems of their central theories. If they fail, if the anomalies continue to crop up despite their best efforts, only then do we have really good justification for thinking that there is something wrong with the theory itself. So for Kuhn falsification is not an alternative to normal science; it is normal science that in the end makes falsification (or something like it) possible.

There are two ways to get out of a crisis. It is possible that many of the anomalies are solved within the existing paradigm – an unexpected breakthrough might occur. In that case, scientists will regain confidence in the paradigm and we return to a period of normal science.

It is also possible that a new paradigm is developed that is in some respects more successful than the old one; and if these successes promise a bright future, then the scientific community can become convinced that the new paradigm might be better than the old one. If that happens, scientists will start abandoning the old way of doing science and embrace the new ways. The old paradigm is changed for the new one – Kuhn calls this a *paradigm shift* – and a new period of normal science begin.

Kuhn has another term for a paradigm shift in science: a *scientific revolution*. Just as in a political revolution one political system is changed for another, so in a scientific revolution one theoretical framework is changed for another. Science in the revolutionised field will look very different after the revolution, since a change in paradigm doesn't just mean that new theories are embraced, but also involves new concepts, methods, research questions, yes, even entirely new ideas about the nature and aim of science. Scientific revolutions are truly radical. But, as we will see in the next section, this raises some hard questions.

4.5 Incommensurability and progress

One of Kuhn's most influential and at the same time most controversial ideas is his idea that scientific paradigms are *incommensurable*. Literally, 'incommensurability' means 'not-together-measurable-ness'. A trivial example of two incommensurable entities would be a metre and a kilogram. What is more, a metre or a kilogram? That is a senseless question: one is a distance, the other a weight, and so you cannot compare them.

Kuhn believes that something a bit like that happens when we ask which of two scientific paradigms is better. To answer that question we would have to look which of the two paradigms better approaches our scientific ideals. But our ideals have no existence independently of our paradigms; our ideas about how to good science are part of our paradigm. When a scientific revolution happens and we discard one paradigm in favour of another, we don't just change our theories, but also the measures we use to determine what is good science. So how could we measure which paradigm is better?

Let us assume that there is, at first, a community of historians that follows the model of Herodotus; and then afterwards a community of historians that follows the model of Thucydides. Which of these two communities is doing better science? Well, the scholars following Thucydides would certainly claim to be the more scientific: after all, they use only reports by eye witnesses and they attempt to give their readers nothing but the truth. On the other hand, the scholars following Herodotus would point out that their opponents can't do serious historical research, because they are unable to look farther back then one lifetime; and also that their method makes them unable to do justice to the diversity of sources and perspectives that any historian is bound to encounter. Who is right in this debate?

Is anybody right? That would only be the case, Kuhn tells us, if there were to exist a set of neutral methodological rules 'above' or 'beyond' the paradigms. Then we could compare the paradigms to these neutral rules and discover which of the paradigms most closely approximates the rules. But Kuhn thinks this is impossible. Ideas about what should count as good science change through time, and there is no neutral point of view from which we can judge objectively about the different paradigms. Another to formulate the same point is this: we cannot step out of our own paradigm, and so we will always judge using our own standards. But that ensures that our judgement will not be neutral.

This sounds pretty sensible – but it leads to a pressing question. If there is no neutral way to compare paradigms with each other, then it would seem that we cannot say that one paradigm is better than another. But if we cannot say that, then it seems that we also can't say that changing from one paradigm to another was an instance of scientific progress. And if that's the case, doesn't that mean that we have to abandon the notion of scientific progress altogether?

The traditional view holds that scientist learn more and more about the world, which means that there is indeed scientific progress. If Kuhn is right, then that view is only correct within periods of normal science. In such periods, scientists are solving more and more puzzles, leading to an ever greater collection of solutions – a kind of progress. But during a scientific revolution the concepts and standards of science change, meaning that much of the old knowledge may no longer be relevant or even acceptable. Many of the problems that the old paradigm solved will no longer exist in the new paradigm, or will look very different, so that the old solutions have to be abandoned. A scientific revolution destroys a large part of the existing scientific edifice. There is, then, no scientific progress in the traditional sense of the world. And if, in addition, we cannot say that the new paradigm is better than the old paradigm, because they are incommensurable, then it seems that the very idea of scientific progress must be relinquished.

What must also be relinquished is the idea that choosing to follow a theory is a rational process. If two paradigms cannot be compared, then how could you rationally choose between them? Kuhn likens the move from one paradigm to another to a religious conversion. The convert will believe that she has made progress – otherwise she would not have converted – but is rarely capable of giving rational arguments for this belief.

The idea that science doesn't know progress and is at bottom an irrational process, has been hard to swallow for many. We generally believe that we know more about the world than the scientists of fifty or a hundred years ago; and much more than those of a thousand years ago. Is it possible to do justice to Kuhn's insight that the standards of science have changed throughout the ages, and to nevertheless believe that there is a form of scientific progress?

In his later work, Kuhn attempts to nuance his views by admitting several forms of continuity that are not threatened by scientific revolutions. He identifies several values that scientists have found important throughout history. We all want theories that are accurate, that explain as much as possible, that are not too complicated, that are consistent with the findings of other sciences, and that lead to new discoveries. Although all of these criteria are rather vague, and different scientists can interpret them in different ways and can emphasise one or the other, they still form a shared and neutral basis on which we can judge competing paradigms. So while we cannot claim that there are rational rules for choosing one paradigm over another, there is still some opportunity for rationality and argumentation – perhaps more than the phrase 'religious conversion' suggested. And if that is the case, perhaps we can also say that there is scientific progress. This progress will be far less uniform and much harder to measure that traditional theories about science have told us; but it is progress nonetheless.

4.6 Lessons from Kuhn's theory

If Kuhn's interpretation of the history of science is correct, we can draw the following lessons about the scientific method:

1. The scientific method is a not a set of logical rules that are the same everywhere. Different fields have different methods; and such methods

consist of a wide variety of things, including ideas about how to formulate research questions, which concepts to use in describing the world, and how to set up an experiment or investigation. We are taught this method mostly through being made acquainted with exemplars – examples that we are supposed to follow.

2. In addition, the scientific method is not timeless, but changes throughout history. Every scientific revolution rewrites the rules for the discipline in which it occurs.

3. Scientists are not always equally critical. During periods of normal science, they aren't critical at all about the framework in which they are working; they accept this framework dogmatically. Only at specific moments of history, during periods of crisis, is fundamental criticism something that scientists are interested in.

4. Scientific progress cannot be understood as a linear process during which we get to know more and more facts. Knowledge may be lost during revolutions – because it doesn't fit in the new paradigm – and it is not always easy or even possible to determine whether a new paradigm is 'objectively' better than an old one.

There can be no doubt that all of these points throw light on the practice of science. But here we have to ask the question that arose in section 4.1: why believe that Kuhn's description of the history of science has normative value? How do we get from the claim that this is how science has developed, to the claim that this is how science ought to develop – that, for instance, scientists are often justified in doing normal science within a dogmatically accepted paradigm? And what about the way in which Kuhn uses historical research in order to draw such sweeping methodological conclusions? Is that acceptable?

4.7 History and/or methodology

We now need to look in more detail at two further aspects of Kuhn's use of the history of science. Not only are they crucial to understanding and judging his theory of science, but they will also guide us towards important questions about historiography in general. These two aspects are:

1. Kuhn uses specific episodes from the history of science to draw general conclusions. That there are paradigms and revolutions is not, according to Kuhn, a contingent fact, but a historical necessity. Given the nature of science, its historical development *had* to be like this.

2. These general conclusions are not of a merely descriptive nature, but also have normative value. They give us insight into the method of sci-

ence, that is, into the processes through which science can *best* attain its goals.

Let us start with the second point. Isn't it strange that more insight into the history of science is supposed to teach us something about how science *should* be done? After all, the historian only sees how science *was* done. But how can we be certain that it was done in the right way? Many of us are not convinced, for instance, that *political* history shows us how *politics* should be done; why would science be different?

Now Kuhn can point at two special features of science. First, science is a social process that we tend to characterise in terms of progress, both in the short term (within paradigms) and in the long term (over the millenia). Whether politics has made much progress in the past twenty years is surely a question open to debate; you can even wonder whether there is such a thing as political progress at all, and if so, whether people could ever agree about what it was. But that science has made progress in the past twenty years seems undeniable. We know more and we are able to do more than two decades ago. Science is successful, and that suggests that there is something *right* about what scientists are doing. For that reason, it makes sense to assume that a description of scientific practice will have at least *some* normative value.

By itself, this argument isn't too strong. Even though science is often very successful, that doesn't mean that all the features of our scientific enterprise contribute to that success. For instance, it is a fact that scientists nowadays use English as the language for international communication; but it would be rather weird to conclude from this that the use of English is in part responsible for the success of science. French, Dutch and Japanese would probably have been equally useful. And, who knows, perhaps the formation of paradigms and the existence of periods of normal science, or the occurrence of scientific revolutions, might be among the features of science that do not contribute to its success? Perhaps they even slow science down?

The second – and more important – thing that Kuhn can point to, is that he doesn't just identify historical patterns, but that he also engages in philosophical reflection on whether and how these patterns might contribute to scientific progress. We saw this in the previous sections. For example, Kuhn argues that you cannot do serious research without a paradigm; and that tenaciously holding on to paradigms, as happens in periods of normal science, allows us to identify the really serious and deep problems that they have. Kuhn doesn't just embrace and celebrate any historical pattern that he identifies. He first thinks them through from a methodological perspective.

Kuhn thus endorses a profound collaboration between the historiography of science and the philosophy of science. The philosopher who tries to understand the scientific method without any historical knowledge will, according to Kuhn, have an utterly wrong conception of scientific practice. Knowledge about the actual development of science is necessary if we want to understand that development. The historian who keeps her distance from philosophical theories, on the other hand, will not be able to understand the difference between good and bad science, and will thus be blinded to an important distinction.

In addition – and this brings us to the first of the two aspects we identified at the beginning of this section – such a historian will not be able to see the difference between accidental events in the history of science and its necessary patterns. Even if something has *always* happened in history, that of course doesn't mean that it is a necessary pattern. (A historian in the late 19^{th} century could have recognised the universal pattern that almost all scientists are men. In the past one hundred years, we have learned that this patters is not necessary.) When Kuhn concludes that science *has to* evolve using his proposed scheme of pre-paradigmatic science, paradigmatic science, crisis and revolution, then this is not only because he finds this scheme everywhere in history; but also because he believes that he is able to show that this is exactly the pattern you would expect when solving specialised puzzles is the fundamental goal of a community. If the historian want to do more than just describe specific episodes from history, if she also wants to understand their specific place in the development of science, then she *has* to use overarching, philosophical frameworks within which those episodes can be interpreted. She will have to arrive at a conception of the goal of science, and to ponder which features of the episodes she describes do, and which do not, help scientists to achieve that goal.

I would say that all of this sounds plausible. But Kuhn's ideas about the necessity for collaboration of historians and philosophers of science have not been taken up with much enthousiasm by the historians. This is not out of an irrational hate of philosophy, but instead has to do with several deep convictions about the nature of historical research that most contemporary historians hold, convinctions that it will be useful for us to spell out.

For a long time, historians of science would look at history squarely from their own perspective. The term 'science' was supposed to apply only to those activities that led to what we today recognise as scientific knowledge. Good scientists were those who brought us closer to our current insights, while people who had very different ideas were thought to be bad scientists, or perhaps not scientists at all. In this way, the history of science became a 'success story' in which humanity managed to slowly free itself from the swamp of prejudice and ignorance, always struggling to get closer to the truth. The heroes of this story are just the people whose thinking was most like ours; the villains were people who engaged in alchemy and astrology, or who resisted theories that we now know to be true.

From the 1960's onward, historians started to take special care to avoid writing history in this way. They had several very good reasons for that. Someone who is always looking at the world through the glasses of the present will not be able to do justice to the past. For instance, the famous physicist Isaac Newton spent much time working on alchemy and the search for the philosopher's stone. The traditional historian of science will be tempted to say that these activities of Newton were not scientific, were just a personal excentricity, and are not worthy of detailed investigation. Newton's dabbling in alchemy just goes to show, he might add, that even a genius can fall victim to superstition.

The contemporary historian will point out, quite rightly, that in this way one becomes blind to what is really happening in history. At the end of the 17^{th} century, alchemy was not seen as a superstition. By taking our current distinction between 'serious science' and 'superstition' and projecting it on the past, we misunderstand history. Not only will the traditional historian fail to understand alchemy, but he will also misunderstand those parts of past science that he *does* believe to be worthy of study, since he fails to understand the worldview that the practitioners of those sciences had. To understand Newton's science, one must understand Newton; and to understand Newton, one must understand why for him there was no sharp distinction between physics and alchemy. So what the historian must do, is understand the past *in the terms that the past actors themselves would use*. And this of course means that we should not make judgements from our current standpoint.

Contemporary historians are also quite resistant to using the kind of universal scheme that Kuhn believes to have found in the history of science. The historian will point out that such schemes generally do more harm than good: once we have a scheme, we start interpreting the past in such a way that it will fit the scheme, and this blinds us to the unique features of the historical episodes. To strengthen this complaint, the historian can point to many famous authors – Vico, Hegel, Spengler, Toynbee – who came up with such universal historical schemes, all of which have now been rejected, not just because they were *false*, but because they were *unfruitful*. A historian using such schemes will invariably attempt to interpret the historical facts in such a way that they fit the scheme, and this makes her a worse historian. (We will revisit this idea in a future chapter.)

This means that there are good reasons to be wary of using any universal schemes or any of our contemporary value judgements when we are interpreting the historical development of science. It is easy to draw the conclusion that normative ideas should play no role at all for the historian. For what other norms could we be using than our current norms; and what are norms if not a universal scheme that all historical episodes have to be fitted into? This conclusion has become generally accepted by historians of science. The historian attempts to stay as close as possible to the factual study of specific episodes, and attempts to keep her distance from value judgements, and even from terms like 'true' and 'false'.

What does this mean for Kuhn's approach? Is it true that, by mixing history and philosophy of science, he is no longer able to do justice to the past? At first sight, this may seem to be the case. Kuhn is interested in the fact that the sciences, more than any other human enterprise, are characterised by progress. He then investigates history with the aim of answering both the question of what constitutes that progress, and of how that progress is achieved. This seems a typical example of the approach where we look at history through the glasses of the present.

But things are not quite that simple. Kuhn will point out that while he does indeed ask certain questions to the past, he has not thought up the answers in advance. He sees progress, but he wants to use historical analysis to find out what this progress consists in. He does make normative judgements, but history itself is his guide. What is more, the concept of paradigm is especially useful for understanding and doing justice to changes of ideas throughout history. Using that concept, Kuhn has an easy time explaining why it was not 'irrational' or 'unscientific' for Newton to engage with alchemy.

But shouldn't a historian simply keep their distance from all value judgements? If we want to take the past seriously, shouldn't we just let it speak for itself – as far as this is possible – and give no judgements at all?

It may be enlightening to develop the following analogy. Suppose that you and I have a discussion about how to do good science. But instead of exhanging arguments with you, what I do is the following: every time you say something I agree with, I applaud; and every time you say something I disagree with, I scowl. At the end of the 'discussion', I calculate how often you agreed with my views, and then I say: "Your score is 43%. You are quite a stupid person." In this scenario, in which I am constantly judging you by my standard, you will not feel that you have been taken seriously.

But now let us consider another scenario, in which I make no judgements at all. Every time you say something I disagree with, I say: "Aha, so that is your opinion. Interesting. But my opinion is different, namely . . . ". When you try to give arguments against my position, I refuse to engage with them. "No," I say instead, "my opinion is my opinion. You have a right to your opinion, I to mine. Don't disrespect my views." In this scenario, I make all discussion impossible by acting as if our points of view are completely subjective, as if there are no standards that we could use to judge them by. This too is not a real discussion; and once again you will not feel that you have been taken seriously.

So while it is undoubtedly good for a historian to be careful with value judgements about historical events, this does not automatically mean that those who do make judgement thereby fail to take the past seriously. Careful judgement *is* a way of taking something seriously. On top of that, Kuhn gives us a theoretical framework within which we can let the past speak for itself – we can understand the decision of past scientists in terms of the paradigm that they were working in – while also allowing us to talk *about* the past, namely by letting us recognise paradigms, crises and revolutions. Of course, it *might* be the case that these terms fail to do justice to the history of science; but that is an empirical question, to be determined by carefully studying the history of science, not a question of principle.

4.8 Conclusion

In the previous chapter, we concluded that theoretical frameworks play an important role in science. In this chapter we have seen Kuhn's development of this idea. He claims that paradigms are the historically determined frameworks within which a scientific community works without engaging in fundamental forms of criticism. Only at specific moments in history do scientists see the point of critical reflection on the framework itself and of attempts to develop new paradigms. This means that science is, on the whole, much more dogmatic than most earlier thinkers had thought.

Kuhn also suggests that the scientific method changes throughout history, even when we consider basic questions about the subject and purpose of science. Precisely because the standards that determine what counts as good science are themselves a part of these historical changes, it can be very difficult to say that science has made progress – after all, how can we measure progress if there is no overarching standard of quality? We have seen that this conclusion can be made more nuanced, but a general lesson we can learn from Kuhn's work is that the progress of science is less continuous and less orderly that earlier methodological ideas may have led us to hope and believe.

We have seen that Kuhn's ideas also raise questions about the nature of historical research. Can one use such research to draw the kind of genereal conclusions that Kuhn wants to draw? Or should the historian limit herself to letting history speak for itself, in all its individuality? These are questions that we will return to in later chapters.

Chapter 5

Interlude: Michel Foucault

5.1 Foucault's facelessness

It makes sense to discuss Michel Foucault (1926 – 1984) at this place in our book, just after Thomas Kuhn, since there are many resemblances between these two thinkers. Both worked extensively in the history of science, but became better known for the philosophical conclusions they drew from their historical studies. Both put a special emphasis on discontinuities in history, rejecting the idea that science can be understood as continuous progress towards the truth. Both coined a word – *paradigm* for Kuhn, *épistème* for Foucault – to indicate the frames of thought within which scientists work and which they rarely, if ever, question.

And yet it would have made quite as much sense to position Foucault almost anywhere else in this book. His development of the new historical methods that he called archaeology and genealogy would make him fit very well in a chapter on the nature of historical understanding. He is an important figure in the development of both structuralism and hermeneutics, two schools of thought we will meet with later. His theories about power, perhaps the most influential part of his writings, as well as his attempts to give a voice to marginalised groups like homosexuals, prisoners and the mentally ill, made him one of the central figures of postmodernism. So we must bear in mind that many of the themes we will see in this chapter will receive a fuller treatment only later on.

That Foucault is not easy to place is also due to the many changes in his interests and methods over the years, which made his thought develop into new and often unexpected directions. This makes him hard to pin down – and that is exactly the way he wanted it, because for Foucault, having a determinate and well-defined identity that is easy to pin down, is nothing less than a nightmare. He makes this clear in a famous passage from the introduction to *The Archaeology of Knowledge* (1969), which ends with an imagined exchange with a critic who blames Foucault for changing his ideas all the time. The critic says:

> 'Are you going to change yet again, shift your position according to the questions that are put to you, and say that the objections are not really directed at the place from which you are speaking? . . . Are you already preparing the way out that will enable you in your next book to spring up somewhere else and declare as you're now doing: no, no, I'm not where you are lying in wait for me, but over here, laughing at you?'

To which Foucault imagines replying:

> '. . . I am no doubt not the only one who writes in order to have no face. Do not ask who I am and do not ask me to remain the same: leave it to our bureaucrats and our police to see that our papers are in order. At least spare us their morality when we write.'[1]

Before we delve into Foucault's thoughts on the history and philosophy of science, it will be useful to discuss *why* he writes 'to have no face'. We will see that understanding this desire equals understanding what is perhaps the most fundamental and constant factor in Foucault's thinking.

Let us start with the part of Foucault's desire that is easiest to understand: he wants to be free from the bureaucratic systems and social norms that try to define our identity, to pin us down and turn us into something that is easy to classify, to understand, to predict and to control. Such systems and norms are everywhere. Your interests may be very diverse and personal, but once you go to university you are expected to choose a single subject – history, philosophy, English, Chinese, chemistry – and from that point on part of your identity is formed by the fact that you are, for example, a 'history student'. And you will always remain 'someone who studied history' and be judged accordingly.

Many other examples can be given. Your gender identity may be very complicated or ill-defined, but both your passport and the people around you expect you to be either 'male' or 'female' and behave according to the expectations we have of males and females. (And if you do not, then *that* becomes part of your identity – you will be spoken of as 'effeminate' or 'a tomboy', categories that also come with a lot of expectations.) Your job may be just a way to earn money to you, but because you are a consultant, or an academic, or a nurse, or a lawyer, you are expected to dress in a certain

[1] The translation by A. M. Sherridan Smith is quite loose here. The original French is: 'Plus d'un, comme moi sans doute, écrivent pour n'avoir plus de visage. Ne me demandez pas qui je suis et ne me dites pas de rester le même : c'est une morale d'état civil; elle régit nos papiers. Qu'elle nous laisse libre quand il s'agit d'écrire.' Rather than speaking of the bureaucrats and the police, Foucault is talking about the registry office – in Dutch, the 'burgerlijke stand.'

way, talk in a certain way, write in a certain way, even think and feel in a certain way. Or, to stay closer to Foucault's original quote, you may have published a book in which you defend thesis X, and then your colleagues start thinking of you as 'someone who defends X' and expect you to continue to do so. Everywhere in society we are constantly put into categories that we are then expected to live up to.

Revolting against the feeling that your identity is being defined by the expectations that people have of you, is of course a very common phenomenon. Most adolescents go through such a phase of revolt; as do, for instance, people in their mid-life crisis. Everywhere people are trying to escape from the systems that define them, in very diverse ways. One person may feel free when going on meditative retreats or walking to Santiago; another chooses to participate in drug-fuelled parties; while a third uses all her money to buy a fast sports car.

But there is one thing that almost all strategies of revolt have in common: they believe that underneath the surface of what society has made of us, there lies a 'true self' waiting to be discovered. On the one hand there is 'me'; on the other hand there is what society wants to make of me; and to be free means to escape from society's influence and just be the real me. When *Pink Floyd* protest against being "just another brick in the wall" by singing "we don't need no education, we don't need no thought control", that is the background theory: without the thought control that comes from society, we can just be our true selves. When *Metallica* complain that nameless forces "dedicate their lives to running all of his," they summarise the problem with "never me, never free." What we have here is a general theory that to be *free*, I've got to get rid of society and its expectations and just be *me*.

We now come to the more original and harder to understand part of Foucault's desire. For it turns out that althought Foucault wants to escape from the systems that try to define him, he does *not* believe that the answer is to just 'be yourself'. In fact, he believes that the idea of a 'true self' hiding beneath the surface of our lives is nothing but an illusion. And that is why the alternative to being what society wants me to be is not 'being me', but 'having no face.' The only alternative to having the identity provided by society is having no identity at all.

Why does Foucault believe this? Because he believes that everything that defines us – the way we think, the way we talk, the way we feel, the way we behave, the way we desire – depends on the social and linguistic structures that surround us. Those structures do not define exactly who I will be. After all, I can choose to be either a lawyer or a historian, either a mainstream guy or a goth, either politically left-wing or politically right-wing, either a positivist or a postmodernist. There are many choices I can make. But the structures that surround me define *which choices I can make* and *what the meaning of those choices is.* I can be mainstream and conform myself to society; or I can rebel against it. But in both cases I will have chosen one of the possibilities that our society itself makes possible. We can choose different positions *within* the structures that surround us, but we cannot get *outside* of them.

If I rebel against society by wearing black clothes and black make-up, getting a nose piercing my mother thinks is horrible and listening to extreme metal music, could this be considered the discovery of a true self that is independent of society? Well, no. This entire style of rebellion is already out there in society, waiting for me to adopt it. It is one of the options that society presents to me. And even if I were an innovator and developed new kinds of clothes, body art and music to express myself, all of them would only have the meaning they have because of their place in our society and their relations to other elements of our culture. Extreme metal music would have a very different meaning as a cultural phenomenon if it had been the traditional music of the Catholic Church; nose piercings would have a very different meaning if they had been used in the 1930's as a symbol of allegiance to the Nazi party; and so on. All the ways I can express myself are available only because of the way our society is already structured.

In fact, the entire project of self-expression is something that is only possible within a particular society. We in the 21^{st} century possess the idea of a unique individual self that can be revealed through artistic creation and lifestyle choices; but that conception didn't really exist before the 19^{th} century and may cease to exist again in the future. The very possibility of being a misunderstood genius is a possibility only because our society is structured in a certain way.

So what Foucault wants to point out is that we cannot look into ourselves and find something that is independent of the structures that surround us. The very words we use to describe ourselves are the words of a language we have been given; and we can't escape from that common language into a 'natural' or 'original' way of thinking that lies waiting for us in the very core of our being. We can't escape from society by finding our real face. Therefore, Foucault wants to be faceless; he wants to reject all solid identities and wants to be able to change all the time. But he doubtlessly knows that this too cannot be a real escape; for after all, if it is possible in our society to be faceless – and it probably isn't – then facelessness too is only one more possible option given to us by that society itself. We are always stuck in the structures that surround us. Even if we change those structures, we can only change them in ways that those structures themselves make possible.

5.2 Archaeology: history beyond the subject

As we have seen, Foucault criticises the idea of a true self that is independent from its context. But what does that have to do with the history of science? As it turns out, Foucault's approach to that topic is tightly linked with his criticism of the self. For just as Foucault thinks that we put far too much emphasis on the subject in thinking about ourselves, he also thinks we put far too much emphasis on the subject in thinking about the development of knowledge.

What does a history of a scientific discipline generally look like? It is mostly concerned with the specific *theories* that scientists defend, with the *arguments* that they use to defend them, with the *observations* they make, with the experimental and methodological *tools* that they develop and use. In other words, it is mostly concerned with what the scientists themselves write and think about. A Kuhnian paradigm, for instance, consists of precisely such things. This means that if we want to know what the paradigm is within which a group of scientists is working, all we need to do is ask them about the theories, methods, tools, and so on, which all of them use and consider to be reliable. People are generally *conscious* of their own paradigm; or, to say the same thing, paradigms are *accessible* to the subjects working within them.

The emphasis of traditional historians of science on the things that the scientists themselves are thinking and talking about has sometimes been criticised for being too *internalist.* An internalist history of science is a history which assumes that all changes in science are due to factors internal to science, factors like new observations and new arguments. But, the criticism goes, science is not isolated. It is also affected by *external* factors, such as economic, political and cultural changes. For instance, the Scientific Revolution didn't just happen because Galilei and some other scientists had good new ideas. The reason that they could have those ideas and that people listened to them, was that changing social conditions – such as the decreasing power of the nobles and the Church, and the rising power of a well-to-do middle class – led to an eagerness for radical ideas that challenged the traditional authorities. So, this line of thought goes, it is the social conditions that caused science to change. A history of science which emphasises such external factors is called an *externalist* history.

Foucault, however, believes that both an internalist and an externalist history of science are missing out on what might be the most interesting and profound level of analysis: the level of what Foucault calls *discourse.* This is the level of the actual texts and documents that the historian can study. According to Foucault, both internalist and externalist historians try to get away from the level of discourse. The internalist wants to find, behind the actual documents, the thoughts and feelings of the subjects who produced those documents. Beneath the texts, which are fragmented, scattered, and perhaps full of mistakes, we try to find the original lucid and coherent vision of the scientist. This means that we see discourse only as a clue to what we are *really* interested in: the consciousness of the subject.

The externalist, on the other hand, tries to show that whatever happens in these documents is just the effect of causes that take place at very different levels: the level of politics, of social organisation, of the economy, and so on. So here, too, the historian wants to get away from the texts themselves as soon as possible to focus on something else: the environment in which the texts were produced.

Foucault, on the other hand, wants to take discourse seriously. He believes we should study the texts themselves and discover the *rules* that they obey. Now, we know that texts obey grammatical rules that are studied by linguists; and we also know that most argumentative texts obey logical rules that are studies by logicians. But Foucault is interested in another kind of rule. He wants to take a historical field of knowledge – for instance, knowledge about language in 18^{th} century Europe, or knowledge about madness in the Renaissance – and he wants to find out what a piece of text had to be like in order to be accepted into that field of knowledge. So his basic question is: how did people have to write in order to be taken seriously?

It turns out that the rules that a text has to obey in order to be taken seriously can change radically in the course of history. As an example, let us take two fragments from texts that belong to the field of knowledge about animals (which today we would think of as a subfield of "biology"). First, a text from 1550, Girolamo Cardano's *De subtilitate*, which is quoted by Foucault in the *The Order of Things*:

> The rat of India is pernicious to the crocodile, since Nature has created them enemies; in such wise that when that violent reptile takes his pleasure in the sun, the rat lays an ambush for it of mortal subtlety; perceiving that the crocodile, lying unaware for delight, is sleeping with its jaws agape, it makes its way through them and slips down the wide throat into the crocodile's belly, gnawing through the entrails of which, it emerges at last from the slain beast's bowel. (p. 24)

Second, a 1995 article from the journal *Mammalian Genome*, titled "An integrated genetic linkage map of the laboratory rat", by Brown et al:

> **Abstract.** The laboratory rat, *Rattus norvegicus*, is a major model system for physiological and pathophysiological studies, and since 1966 more than 422,000 publications describe biological studies on the rat (NCBI/Medline). The rat is becoming an increasingly important genetic model for the study of specific diseases, as well as retaining its role as a major preclinical model system for pharmaceutical development.

And another citation from the same article, in which several huge tables of genetic data appear, of which this is a slightly abbreviated part:

D1Mgh18	CYSS	Cytostatin S	M75281
D1Mgh19	M26926	Olfactory marker prot. gene	M26926
D1Mgh20	ORFEP	ORF for P-glycoprotein	X61106

Even if we know nothing about rats, we can immediately see that the way you have to write about them in order to be taken seriously has changed immensely. It's not so much that our theories about rats have changed; but the texts are so different that they don't seem to be talking to each other, or about the same thing, at all. And much of that difference, Foucault would argue, can be traced back to the underlying rules that govern how texts containing knowledge about animals are to be constructed in these two different periods.

The 1550 text is written in what we can think of as everyday prose, although it was made more acceptable as a vehicle of knowledge by being written in Latin. The text states facts, and only facts; it is not necessary for the text to give evidence of those facts or to trace how the author came to know about them. It is full of normative, value-laden terms: 'pernicious,' 'violent,' 'enemies.' The general claim that it makes has the form of what Foucault calls an 'antipathy,' a link between two things in nature that hate each other. So we can see that one way to be taken seriously in 16^{th} century discourse about animals is to write a Latin prose text that presents value-laden claims about antipathies in nature.

The 20^{th} century text is completely different. It starts with an abstract that establishes the importance of the claims that are going to be made. It identifies its subject matter animal by giving a systematic scientific name, which means it puts itself into a context of scientific name giving that didn't exist in the 16^{th} century. It references its sources (indeed, it ends with a long bibliography). It contains tables to summarise a lot of empirical information, information that is presented in technical terms that cannot be understood by the lay person. And it is precisely because of this that it can be taken seriously: no article that fails to reference sources, use scientific nomenclature, and so on, would be taken seriously by contemporary biologists.

So when Foucault looks at the history of a field of knowledge, he finds intricate systems of rules that define how texts in that field can be written. Why is this interesting? Because, Foucault says, the rules that determine how 'serious' investigators can write (and talk) *limit what they can say.* It doesn't strictly *determine* what they can say: within the rules that govern Cardano's discourse, he can claim both that the rat hates the crocodile and that the rat loves the crocodile. But what he cannot do is give a table with empirical properties of the rat – for instance, a table in which 50 rats are listed, each with its weight, size and tail length – because tables of empirical data are not part of the rules that govern his discourse. If he *had*, somehow, anachronistically, had the idea of doing such measurements and writing down such a table, nobody would have taken him seriously. It would not have looked like a scientific document to the 16^{th} century reader. On the other hand, a modern biologist could not possible make the claim that Nature has created rats and crocodiles as enemies: such a claim doesn't fit the rules of our discourse, and we would not be able to take it seriously. In a sense, the modern biologist can't even *think* that way: it just wouldn't occur to her to have such a thought.

All the things, then, that the traditional historian of science is interested in – theories, arguments, measurements – are only possible because of the way that discourse in a certain period is structured. Beneath all the things that the subject thinks and writes, there is the level of the rules of discourse; rules that people implicitly follow, but are mostly not aware of. The study of these rules, which determine to a very large extent the science of any period, is what Foucault calls *archaeology*.

5.3 Epistemes and paradigms

Archaeology, then, is the study of the rules of serious discourse. Foucault calls a system of such rules a *discursive formation* or an *épistème*. (We will drop the accents from the word, as is common in the English literature.) An episteme is – at least in Foucault's original conception – a set of rules that govern *all* serious scientific discourse in a certain society and time period. So, for instance, we can speak of the current episteme of Western science; or of the episteme of Renaissance Europe. An episteme governs serious writing at such a deep level that it will be evident across the sciences: all knowledge in a certain period will conform to it. A change in episteme will therefore have effects in all the sciences; and Foucault indicates that this is one of the reasons he came up with the concept in the first place:

> [T]wo things in particular struck me: the suddenness and thoroughness with which certain sciences were sometimes reorganized; and the fact that at the same time similar changes occurred in apparently very different disciplines. (*The Order of Things*, p. xii.)

Now according to Foucault, there have been three epistemes in Europe in the past five centuries, and it will be useful to us to take a brief look at them. They are the epistemes of the Renaissance, of what Foucault calls the Classical Age, and of the Modern period.

1. In the *Renaissance*, knowledge claims take the form of statements that set out a *resemblance*, or *analogy*, or *sympathy* between two – often wildly different – things in the world.[2] Things in the world are treated as meaningful signs that point to other things in the world. So, for instance, the walnut and the human brain resemble each other in

[2] We should note that many scholars believe that Foucault paints a one-sided picture of the Renaissance, focussing too much on Neo-Platonic writers while neglecting the Aristotelian majority.

appearance, and this is taken to be a significant relationship between them; a relationship that lends credence to the idea that eating walnuts is good against headaches.

More complex resemblances between entire systems of things are treated as analogies: some Renaissance authors believe, for example, that there is an analogy between the known planets, the known metals, and the major organs of the human body:

Sun	Gold	Heart
Moon	Silver	Brain
Mercury	Mercury	Lungs
Venus	Copper	Kidneys
Mars	Iron	Gall bladder
Jupiter	Tin	Liver
Saturn	Lead	Spleen

The sun is of course radiant and yellow, and thus easily linked to gold; while its pre-eminent position in the Universe makes it akin to the heart, which is our most vital organ. Saturn – who in Greek and Roman mythology was a rather dubious character who ate his own children – was associated with dark and melancholy moods, and therefore linked to the dark and heavy metal lead, and to the spleen, which was supposed to create a 'black bile' that makes people melancholic.

Relations of sympathy and its opposite, antipathy, are also supposed to pervade the world; we already saw the example of the rat and the crocodile in section 5.2.

2. In the *Classical Age*, which for Foucault spans the 17^{th} and 18^{th} century, knowledge no longer took the form of statements of resemblance and analogy. Instead, the perfect form of knowledge was the *systematic table* in which all the things in the world could be classified.

   In order to make such a table, scientists had to analyse things into small parts: for instance, the biologist had to look at minute details of plants to find out about all the ways in which different plants can be different. He then had to set up a language, often an artificial and highly technical language, to describe those smallest details that are necessary for classification. Thus, the biologists had to come up with a vocabulary to describe, say, all the ways that leaves can be attached to the stem. He could then set up a table that contains all the ways a plant can be – it could have such-and-such leaves, attached in this or that way, with flowers like such, and so on – and attach a scientific name to each of the possibilities. In this way, scientists arrived at intricate systems of classification.

Foucault also points out that the 17^{th} and 18^{th} century see a lot of interest in the question of how to construct an ideal, perfect language. This makes sense, for if scientists want to have a true systematisation of everything, they need a language with words that exactly fit the order of the universe.

3. In the *Modern period*, knowledge takes the form of claims about how things fit together in time. From the early 19^{th} century on, scientists become interested in questions of origin and evolution, in laws of history, in relations of cause and effect. It is in the 19^{th} century, for instance, that biologists stop just classifying the species, but start developing theories of evolution. It is also in the 19^{th} century that linguists start developing historical theories about the development of languages. In general, when we moderns are confronted with something unexpected, we want to know how it came about; what caused it; what laws, if any laws can be found, would have predicted it.

Whether Foucault has characterised these epistemes in exactly the right way, or whether we might perhaps want to distinguish more or fewer of them, are of course questions any historian of science would have to ask herself. But it is easy to see that the basic idea of an episteme allows us to understand several features of science that are otherwise hard to make sense of.

First, Foucault's theory of epistemes allows us to understanding why some ideas are not just rejected by scientists – for that, we can often point to experiments, measurements, arguments, and all the other things that have a central place in a traditional history or theory of science – but why some ideas are just *not taken seriously at all*. We don't do astrological research. But the reason for that is not that we know in advance that all astrological theories are wrong; after all, you can always think of new ones that have not been disproved. We don't do astrological research because astrological claims do not fit in our episteme: they are simply the wrong kind of claim to appear in texts that can be taken seriously by scientists in our society.

Second, we can understand why certain ideas are taken seriously in one period but not in another. We can see immediately why astrology was taken absolutely seriously by people in the Renaissance, including many of the best astronomers of the time. If our knowledge of the world has the form of complex analogies, then the claim that the planets are analogous to and therefore connected to bodily organs, and thus have effects on health and personality, makes a a lot of sense. One can still *disagree* about whether astrological claims are true – the Catholic Church, for instance, was mostly opposed to astrology because it was thought to be incompatible with free will – but the claims are worth investigating and having an opinion about. But in our own time, where we are looking for causal mechanism, astrology cannot be taken seriously; for nobody, including the astrologers, has any coherent proposals for causal mechanisms that link the planets to our moods.

Third, we can understand why certain ideas are born in one time rather than another, even if the data needed for having the idea were already available. It had, for instance, been known for a long time that there were similarity relations between the different languages of the world; but only around the end of the 18^{th} and the beginning of the 19^{th} century, did these observations become the data for historical hypotheses about the evolution and history of languages. Foucault explains this by pointing out that historical hypotheses just didn't fit in the earlier epistemes.

Let us finish this discussion by pointing out some of the differences between Foucault's epistemes and Kuhn's paradigms. Both are frameworks that determine to a very large extent the ways that scientists do their research and the theories that they will come up with. But Kuhn's paradigms are, so to speak, closer to the surface: they consist of things that scientists are aware of, like methods and theories. An episteme, on the other hand, is so basic that we generally do not even realise that things could be otherwise; that we could take other things seriously, if only the rules of our discourse were different. Because epistemes are more basic, they are also larger, both in the number of disciplines they affect and in their duration. A paradigm is always the paradigm of a single scientific discipline; but an episteme is shared by many disciplines, maybe even by all. And where paradigms might change relatively often, an episteme generally lasts for centuries.

Of course, one can believe both in paradigms and epistemes: deep underlying structures of discourse can be combined with more changeable sets of assumptions that scientists at a given moment agree on. Foucault and Kuhn could both be right, but just be talking about different aspects or layers of science.

Some things are missing, however, from Foucault's story about epistemes; things that according to Foucault's analysis we moderns are especially interested in. First, Foucault's archaeology, which investigates the rules of discourse, can find out what the episteme of a certain period was, but it tells us nothing about how and why epistemes change. Second, archaeology also fails to address the question of the effects that epistemes have on society as a whole. This is not surprising: when developing his archaeology, Foucault chose to ignore all questions of causality and influence. But he came to see this as a weakness in his own methodology, an insight that led him to develop a new project that he called *genealogy*.

5.4 From archaeology to genealogy

The term 'genealogy' indicates, in its everyday use, the study of families and the tracing of lineages. But when Foucault uses this term to describe his own methodology, this is not what he means. Instead, he uses it to mean what the German philosopher Friedrich Nietzsche (1844–1900) meant with the term, namely, a historical story about the origin of our contemporary values; a story which, in the hands of both Nietzsche and Foucault, will show that those values came to be dominant because of irrational, weird and contingent causes.

Nietzsche's most famous genealogy is the one worked out in his book *The Genealogy of Morals*. Nietzsche starts by giving us two broad visions of morality. The Greeks and the Romans thought of people in terms of the opposition 'good' and 'bad'. Good people are strong, honourable, and brave; whereas bad people are weak, dishonourable and cowardly. But, Nietzsche tells us, the Christians thought of people in terms of the opposition 'good' and 'evil'. Good people are humble and harmless; while evil people are proud and dangerous. This means that what is good for a Roman and what is good for a Christian can be quite different things. According to the Graeco-Roman view, if someone hits you, you should stand up for yourself and hit him back; or at least you should demand some kind of compensation that will satisfy your honour. According to the Christian view, however, you should not hit back your assailant; in the famous words that the Gospel attributes to Jesus, you should instead "turn the other cheek."

How did it happen, Nietzsche wonders, that the originally dominant Graeco-Roman view gradually lost out against the new Christian view? Was it because Christians had brilliant arguments that convinced everyone? Or did they have a revelation from God that they could prove to be genuine? Neither. Instead, what happened was that these two systems of values fitted the psychology of very different groups in society, and the group that happened to be receptive to the Christian values turned out to be the majority.

The idea that a good person is someone who is strong and brave fits very well with the citizens of small city states, citizens who have to regularly fight for the survival of their city and therefore need to be strong and brave; which means that this idea was a good fit for the Mediterranean world before the Roman Empire. And even in the time of the Roman Empire, it was an attractive vision for the Roman elites. But, Nietzsche points out, in the time of the Empire, there were large classes of people who couldn't really strive for this type of goodness. Slaves, for instance, had no business being proud or brave; that would just result in their getting beaten by their masters, or worse. The same was true for poor people, trying to scrape together a living; and even for the rich among the subjected peoples in the Empire, whose attempts to stand up for themselves and get rid of the Roman occupiers were forcefully and bloodily suppressed. These people had little to be proud of, and it was prudent for them to be meek and humble and turn the other cheek when violence was done against them.

The groups that were low in the social hierarchy therefore had to choose: either they could hold on to a system of values under which they had to see themselves as bad and worthless people; or they could adopt a new system of values according to which they, rather than the Roman aristocrats and soldiers, were the good people. Of course, the latter was much more psychologically satisfying. So when Christianity came along, with its praise of humble and harmless people, large groups in society were ready to embrace it: not because of good arguments or divine revelations, but simply because these new values were good for their self-esteem. And perhaps an even stronger psychological motivation was the fact that Christianity promised revenge against the ruling Romans, a revenge that was made all the sweeter because it did not involve such dangerous things as revolt. All one had to do was wait for the Day of Judgement, and then all one's enemies would be condemned to eternally burn in Hell.

According to Nietzsche, then, Christian values became dominant in late Antiquity because there were many people whose self-esteem issues and whose secret lust for vengeance against their oppressors were well served by these values. Obviously, that is not the story that people had traditionally been telling about the rise of Christianity; and for a devout believer it might in fact be a rather shocking suggestion, since Nietzsche's story turns the historical origin of these values into something that is completely irrational (since there were no good arguments involved) and rather unethical (since it is not very praiseworthy to believe in Hell merely because you want to see other people burn). So Nietzsche's historical story serves as a kind of criticism of our current Christian value system, a criticism that works by dissolving its own traditional origin claims.

Now, Nietzsche is of course aware – and we should be aware of it too – that his genealogy doesn't show that Christian values are *wrong*. After all, people sometimes accept right ideas for wrong reasons. But genealogy does sow doubts about the status of our values; it dispels their obviousness; and it may make us wish to reconsider whether these values are really better than those that came before them. Nietzsche himself in fact advocates a return to Graeco-Roman morality, because he thinks these values will be far better at stimulating the emergence of artistic geniuses and philosophical free-thinkers. But that advocacy can be separated from the genealogy itself.

What about Foucault? Like Nietzsche, Foucault, in his genealogical works, searches for the origins of our values and ideas. Like Nietzsche, he concludes that these origins are not rational, but irrational and contingent. And again like Nietzsche, he uses this discovery to dispel the 'obvious rightness' that our values and ideas often have for us, because we are so used to them that we can't take any alternatives seriously. But Foucault's version of genealogy is not identical to Nietzsche's, because he is a very different kind of historian, as we can already guess from comparing Nietzsche's story to what we read about Foucault earlier in this chapter.

First, Foucault is opposed to Nietzsche's psychological interpretation of history. As we have seen, Foucault believes that historians have traditionally put too much emphasis on the subject, and Nietzsche's use of psychology as an explanation of changes in how people think is a clear example of that. Foucault is going to remain faithful to archaeology and the analysis of discourse; but in order to broaden his tool-kit, he is now going to talk, not just about texts, but also about social procedures and arrangements.[3] Second, whereas Nietzsche's historical stories are extremely broad, making sweeping claims about the hidden causes of massive social changes, Foucault is a much more careful and detailed historian. He is interested in the way that very specific and small changes in society – the invention of a new rifle, the development of a new architecture for prisons, the establishment of a new method of teaching, and so on – have changed the way that people think and write.

We will investigate, in the next two sections, the general theory that Foucault develops from his genealogical studies. According to this theory, many of the most important aspects of our culture have not been shaped to fit the beautiful ideals we often invoke to defend them, but to fit the necessities of modern power. For instance, the change from corporal punishment to modern prisons has not been made, Foucault believes, because we have become more humanitarian, but because prisons are a more effective way of controlling people than corporal punishment is. The changes in our educational system have not been made in order to give students more opportunities for growth, but because they turn education into a more effective tool of control. To understand and unmask our society, we thus need to focus on the concept of power.

There is an additional reason for this focus. Although a Foucauldian genealogical story will look very different from one written by Nietzsche, the basic aim remains the same: to show that certain aspects of our society as not based on the kinds of values we thought they were based on; and in doing so, to free us – as far as possible – of the immense power that these aspects exert over us. So it is not only in terms of content that Foucault's theory of society focuses on power, but also in terms of its aim. By understanding how power works, we can, perhaps, to some extent become free from it. Let us therefore talk about power.

5.5 Power and knowledge

What is power? According to traditional ways of thinking, power is wielded by either the state or by some individuals who are 'in power'. Furthermore, power is fundamentally *repressive*. The state's power is shown most clearly when it forbids certain modes of behaviour and punishes anyone who doesn't obey. For instance, the state tells us not to murder; and anyone who nev-

[3] In fact, he had already done so in some of his earliest texts, before he worked out archaeology as a coherent method. As always, Foucault's development as a thinker is complex and non-linear.

ertheless murders is incarcerated. The state tells us not to drive above a certain speed limit, and anyone who does will get a speeding ticket. In this way, power is used to repress behaviour that the state sees as undesirable.

To be sure, the state also sometimes tells us what to do, rather than what not to do. It tells us to pay taxes, to drive on the right (or left) side of the street, to send our children to school, and so on. But there is no deep difference between such positive commands and the negative commands mentioned above. First, because there is no deep difference between the (positive) command that I must pay taxes and the (negative) command that it is forbidden not to pay taxes. Second, and more importantly, because in both cases we think of power as something that is *external* to us and that is constantly trying to *impose its will* on us through explicit commands. Power takes the form of rules and punishments.

We will come to Foucault's criticism of this conception of power in a moment. But first, let us consider the relation that power has with what Foucault is most interested in: science, scholarship, knowledge in general. What does the traditional conception of power suggest about this relation?

The traditional conception will make us think about power and knowledge as two things that are fundamentally opposed to each other. After all, what do we need to acquire knowledge? We need the freedom to research any topic that we want to investigate, and the freedom to defend any conclusion we come to. In addition, we need to have an environment where people are not forced to adopt certain opinions based on threats or violence, but where everybody is allowed to make up their own mind based on rational arguments. As the German philosopher Jürgen Habermas (born in 1929) would say, truth and knowledge are best served by *power-free communication*.

If power intervenes in the domain of knowledge, then, it will generally be in order to repress certain views and distort the truth. A famous example is the way that the Catholic Church forbade certain scientific and philosophical books by putting them on the *Index Librorum Prohibitorum*, the 'index of forbidden books'; and how it sometimes persecuted and punished people whose opinions it did not agree with. Or, to take a more contemporary example, we can think of nations were freedom of the press is restricted and only certain political points of view are allowed to be published. In both cases we tend to say that the truth is suppressed by those in power. And we also tend to believe that those in power suppress knowledge because they are afraid of it.

This is, of course, a comfortable idea for scientists and intellectuals. When we dedicate our lives to investigating and understanding the world, to writing books, thinking up theories, collecting facts, arguing with each other; when, in other words, we are dedicating ourselves to knowledge; we are then automatically also dedicating ourselves to freedom. The truth will set humanity free; and humanity needs us, the intellectuals, to discover and disseminate that truth.

But, according to Foucault, this comfortable idea is a dangerous illusion. Rather than being opposed to power and being champions of freedom, intellectuals tend to be accomplices of power and destroyers of freedom. But in order to understand how this could be the case, we first have to abandon the traditional idea of power as something external that imposes its will on us. For while that might have been a somewhat plausible idea several centuries ago, in a world of kings whose relatively limited control over their subjects did often take the form of laws and punishments, it fails to capture how power functions in a modern society.

In order to illustrate that shift, let us look at two examples of the treatment of criminals that Foucault presents in his major genealogical study, *Discipline and Punish: The Birth of the Prison* (1975). First, Foucault describes the gruesome execution in 1757 of a man who assaulted the French king Louis XV. Here we have a traditional display of power: the law is broken in the worst way possible, by an assault on the monarch, and the powerful respond by punishing in the severest way imaginable. Second, Foucault presents a set of prison regulations from 1837, which consist of a precise time schedule indicating at what time the prisoners have to get up, how long they have to work, at what time they have to go to sleep, and so on. What has changed? The entire idea of punishment. In the first example, punishment is meant to, well, punish; and also to act as a deterrent to others who might want to transgress the law. But in the second example, the punishment is meant to *transform* the criminal and turn him into a productive member of society. The criminal is supposed to internalise the discipline of the prison, so that, once he is released, he will be able to hold a job and make a living without engaging in criminal activity.

This change in how society thinks about punishment is perhaps most beautifully exemplified by the *panopticon*, a new type of prison thought up by the English philosopher Jeremy Bentham (1747–1832). The panopticon is a circular building with the cells in a single ring on the outside and the guard sitting in the middle, in a position where he can look into every cell. Crucially, the building is built in such a way that while the guard can see the prisoners, the prisoners can't see the guard. Thus, never knowing whether they are being watched or not, they'll have to be on their best behaviour all the time; and their fear of being watched will essentially take over the job of the guard, who doesn't even have to be around in order to be effective. After a while, this idea of having to behave well will be so internalised that the prisoners will behave well of their own accord – even after their release.

Now it may seem that Louis XV had real power, while a state that merely reforms criminals has far less power over its citizens. But actually the opposite is true. Repressive power is very limited in its scope and application. The everyday behaviour of citizens in a traditional state is left mostly up to themselves, and only excesses are punished. And it couldn't be otherwise, since it would be extremely inefficient and costly to have a police force that is constantly watching everyone and making sure that they work hard, for instance. Modern power, on the other hand, while being more subtle, is also more 'powerful'. For once we have changed someone into the kind of human being we want him to be, we no longer need to watch him; instead, he will watch himself. If we can educate or discipline someone into being a 'normal', 'upright' and 'productive' citizen, then he will watch himself constantly and behave himself the way we want him to behave in every single aspect of his life. So although it may be more *humane* to discipline criminals than to execute them, it is also a way to have *more power* over them.

According to Foucault, then, power is no longer an external force that punishes transgressors;[4] power at its most important and most efficient is *internal* to people. Through our training and education, we are turned into a particular kind of person; and once we are that kind of person, no external force is needed any more. We believe that we are free, because we can do exactly what we want. But in fact, we are only free to be the kind of person that we have been turned into and to want what we have been taught to want. Thus, modern power limits our freedom in much more radical ways than traditional power ever could. It makes us, to quote a song by *Radiohead*, into someone who is:

> Fitter, happier, more productive,
> Comfortable,
> Not drinking too much,
> Regular exercise at the gym
> (Three days a week),
> ...
> At ease,
> Eating well
> (No more microwave dinners and saturated fats),
> A patient better driver
> A safer car
> Baby smiling in back seat
> ...
> Healthier and more productive
> A pig in a cage on antibiotics

If this is how modern power functions, it also means that power is no longer something that is wielded by someone. Nobody is outside of power; everybody is part of the society that turns us all into a specific type of person. There may be a sense in which a high-up politician or a rich CEO is more powerful than the rest of us; but in a fundamental way, they are just

[4] If it ever was. We can imagine, for instance, that the traditional religious ideas of Hell and Heaven worked in much the same way as Bentham's panopticon does: by giving us the idea that we are constantly being watched and judged by God, we start to constantly watch and judge ourselves.

as much formed by power as we are. The politician, for instance, has had to turn herself into a politician, something that requires immense amounts of discipline and self-control. We are all caught in the nets of power. Power cannot be localised, and precisely because of that it cannot be overthrown. Even if we topple the government, we will still be subject to the power that we have internalised.

We now return to the question of power and knowledge. Is knowledge opposed to power, as the traditional idea of power led us to believe? Absolutely not! Knowledge, Foucault points out, is one of the most important factors that cause us to discipline ourselves. The *Radiohead* song can serve as a good example. Why do I exercise three times a week? Why did I stop eating microwave dinners and saturated fats? Why is my baby in the back seat? Because of the medical and technical knowledge that is current in our society. I am bombarded by research that tells me that not exercising leads to cancer, that saturated fats lead to heart attacks, that fatal accidents happen more often when the baby is placed on the front seat. Immanuel Kant, writing at the end of the 18^{th} century, could tell us that I have lost my freedom when I have

> a book that thinks for me, a pastor who acts as my conscience,
> a physician who prescribes my diet,[5]

but for me, living in the 21^{st} century, the last phrase sounds almost bizarre. *Of course* I let physicians decide my diet; after all, they have knowledge about what is healthy that I myself lack. And of course I let technicians and statistician decide for me where I should put the car seat for my baby. After all, I want the best for my child, and these people know what is best.

Once we look for it, we can see this relation between knowledge and power everywhere. Because of their presumed medical knowledge, doctors get to decide if I must work or will get paid sick leave. Because of their presumed knowledge of human psychology, psychiatrists get to decide whether violent offenders can re-enter society or need to be confined in mental hospitals. Because of their presumed economic knowledge, economists get to decide to a very large extent the fiscal and financial policies of the state. But, again, this is not power wielded by the individual, for all these people have to make their choices in accordance to the current state of knowledge in their field, or otherwise lose their jobs and responsibilities. Everyone, the specialist as much as the non-specialist, is caught in the power structure of knowledge.

We can further illustrate this, as well as some of the details of Foucault's ideas about power and knowledge, if we look at modern education. As example we can take something close to home: the course that this book was written for.

[5] *Answering the Question: What Is Enlightenment?* (1784)

5.6 This course: a Foucauldian analysis

According to Foucault modern power works primarily by changing us into the kind of people that society wants us to be; it turns us, in other words, into 'docile bodies'. Foucault identifies three tools that have been developed for this purpose: *observation*, *normalising judgement* and *examination.*

As we have already seen in the example of the Panopticon prison, we can make sure people behave well simply by making them conscious that they are, or could be, *observed.* The Panopticon is a highly dramatic example, but modern society is full of other, simpler ones. Companies like to have open work spaces with few walls and much glass, because this ensures that all of their employees are constantly watching each other. They also like their IT department to have tools with which they can watch people's computer screens, to discourage their employees from spending private time on the internet. Teachers like sloping lecture halls so they can watch what all their students are doing. And our streets and other public places are filled with surveillance cameras that should discourage criminal activity by making potential criminals feel that they're being watched.

Normalising judgement simply means that we judge people according to pre-defined norms. Of course, this has always happened, but the amount and precision of norms have sky-rocketed in modern society. There was a time when a restaurant might just be judged as 'good' or 'average' or maybe 'better then that other place down the street.' But nowadays, a restaurant is not only judged in increasingly precise ways on the quality of its food and service (both by professional critics and by its customers on a whole array of websites and apps), but also according to large amounts of very specific rules having to do with cleanliness, administration, work circumstances, and so on. These rules together define what is 'normal' and what is 'abnormal,' where being abnormal often carries social repercussions. In the case of a restaurant that has an 'abnormally' designed kitchen, this might mean closure; in the case of a child who is 'abnormally' active or impulsive, this might mean being removed to a special school or being forced to take medication. In these ways, society exerts a normalising force on almost all our activities.

Observation and normalising judgement come together in the phenomenon, well-known to any student, of the *examination.* In an examination, the person examined is observed, is made highly visible, by having them answer a set of questions that reveal their skills and the extent of their knowledge. At the same time, the examination is used for a normalising judgement: we take the answers and compare them to a pre-given standard of correctness. This comparison leads to a precise grade that measures how 'good' the student is. This grade is then entered into an information system that records all the past performances of the student and uses that record to determine whether or not the student has attained the required level to be given an official title (like 'bachelor of arts').

Note the role that knowledge plays in the examination. On the one hand, knowledge is defined in terms of the ability to perform according to certain norms. You, the student, really *know* something about Foucault when you are able to correctly answer the questions that I will ask about Foucault during the examination. I, the teacher, set the standard for correct answers and judge you according to that standard. I then transfer the grade you attain to our central information system, where you will be rewarded (eventually, by being awarded your bachelor's) or punished (by having to resit the examination or retake the course) based on whether you have effectively turned yourself into the kind of knowledge reproducing person that society wants you to be. So I am the tool the system uses to turn you into a certain kind of *subject of knowledge* (someone who knows).

But at the same time you are turned into an *object of knowledge* (someone who is known). The information we store about your performances serves to increase the set of data that society has about you. We use that data to judge you, and to compare you to other students, your year group to other year groups, this course to other courses, and so on. So at the same time that you are turned into someone who knows, you are also turned into someone who is known. And the end result of the whole process is that you will have been turned into the kind of happy, productive citizen who has earned the title of BA.

As you can imagine, many have found Foucault's story about modern power and knowledge troubling, and have reacted to it with a wish to escape from the networks of power that determine so much of our lives. Earlier, I said that Foucault's genealogy is supposed to make it possible for us to criticise parts of our society – like our educational system – that would otherwise seem to make perfect sense and be almost inevitable. And this is certainly true. But can this kind of criticism help us to *escape* from modern power? If you decide to rebel and not learn the exam, you will just get a bad grade. If you don't turn up for the exam at all, this will also be duly noted in your file. If you decide to stop your education altogether, well, that just forces you to choose another of the roles that are possible in our society.

This brings us back to where we started this chapter: to the topic of facelessness. Whatever face you have, society will be able to measure and judge it. Only if you have no face at all could you escape from the networks of power and knowledge that surround us. But is it possible to have no face? Foucault himself wrote in order to have no face, but that hasn't stopped people like me from turning him into a book chapter and a bunch of multiple choice questions. What this course proves is that Foucault is as much an object of knowledge as anyone, as much caught up in the network of observation, normalising judgement and examination as we all are. True facelessness would consist in leaving behind no traces at all – and that is, in our modern societies, perhaps only possible by never having been born.

Chapter 6

Objectivity

Objectivity is often seen as the hallmark of good science and good scholarship. The work produced by the serious academic researcher is not supposed to be merely a presentation of her opinion, her political position, or her subjective taste. Using facts and arguments, she should rise beyond the merely subjective and ascend to the level of objective truths.

But objectivity is also often thought to be impossible. We are well aware that any interpretation we give, any theory we develop, will be inescapably influenced by our historical, cultural and social position. You and I simply see things differently from people at other times and people with different backgrounds. *Our* facts may not be *their* facts; *our* best arguments may fail to convince *them*. So clearly all our claims will be subjective. Even if there were such a thing as objective truth, we will never be able to reach it.

Both trains of thought – one pushing us towards objectivity, one pushing us towards subjectivity – exert a powerful influence on the self-understanding of the humanities. It will therefore be well worth our time to try to get more clarity about the nature and possibility of objectivity.

6.1 Descartes and the objective world

René Descartes (1596–1650) was a major figure of the Scientific Revolution of the 17^{th} century, not just because of his scientific and mathematical discoveries, but especially because he did more than perhaps anyone else to develop the modern conception of science. A key component of his philosophy is a strict distinction between *subject* and *object*. So if we want to understand where our modern ideas about scientific objectivity come from, Descartes's philosophy is the perfect place to start.

In his famous *Meditations on First Philosophy*, Descartes sets out to discover whether there is anything he can know for certain. Yes, he has learned much at school; but repeated disappointments have shown him that much of what he was taught wasn't really true. So he cannot simply trust what other people tell him. If he wants truly reliable knowledge, if he wants *certainty*, he will have to find it for himself.

There seems to be a single obvious source of reliable knowledge: sensation. I see and feel the table in front of me; so surely there really is a table? But Descartes is more pessimistic. Our senses sometimes deceive us: a stick in the water may look bent, even though it is in fact straight. Two objects may seem to have different colours, even though they are in fact the same colour. In dreams, or in fits of madness, or after taking certain recreational drugs, we even seem to see things that aren't there at all. Of course this doesn't mean that the senses are *always* wrong. But before we can trust them, we first need to find out *when* we can trust them. And that is not something the senses themselves can tell us.

So what *can* we know for certain? One certainty, Descartes points out, is that *I think*. Even when I find myself in a state of total doubt, even when I worry that everything I believe is wrong, I can still be certain about one thing: that I'm doubting, that I'm having thoughts. *That* at least cannot be doubted. I *think*. I *exist*.[1]

That's pretty minimal, of course. I exist. Nice, but not much of a basis for scientific knowledge. Descartes is well aware that we need something more. He goes on to argue that I can be certain not only about my own existence as a thinking being, but also about the existence of all the thoughts and sensations that I am having. When I feel pain, I can be mistaken about the cause of the pain, but I can't be mistaken about the fact that I feel pain. (It would be crazy for somebody to tell me: "It's just an illusion, you don't really feel pain!") When I see a table, I can be mistaken if I believe that there is a real table out there in the world; but I cannot be mistaken about the fact that I have a table-like sensation in my mind.

Here we encounter the Cartesian[2] conception of the subject. It consists of the idea, first, that you are essentially a *thinking being*, a *mind*. Sure, you also have a body, but for Descartes this is much less essential. If your body were to die and your mind were to survive, then *you* would survive. If your mind were to die and your body were to survive, then you would be gone and only a zombie would be left. Your mind is you.

Then, second, we have the idea that the subject is *transparent to itself*: that you have perfect knowledge of the contents of your own mind, of your own consciousness. By contrast, and this is the third important idea, your knowledge of anything else – of the world of physical objects, or of the

[1] In an earlier text Descartes wrote the more famous phrase "I think, therefore I am." But it doesn't make sense to *conclude* that you exist from first noticing that you are thinking, since you can hardly notice that you are thinking without at the same time noticing that you exist! So the later formulation "I think, I exist" is to be preferred.

[2] Descartes was known by his Latinised name 'Cartesius', and somehow the adjective 'Cartesian' has stuck. In high school math you may have learned about Cartesian coordinates, that is, standard x-y-coordinates. Those were thought up by Descartes.

thoughts and feelings of other people – is *derivative and less certain.* You can be sure that you have sensations of a table. But you can never be equally certain that there is a real table causing those sensations.

This is important enough to state again. We might think – and philosophers before Descartes usually did think – that when you look around your room, you are aware of the table in front of you. But according to Descartes, what you are really aware of is what he calls the *idea* of a table, and what we, in our current terms, would probably call a table-like *sensation* in your mind. Descartes thus posits a 'veil of ideas' between us and the world: we are never directly aware of the world, since our ideas (that is, the contents of our consciousness) are always between us and the physical objects. Think you are tasting a nice sweet piece of chocolate? What you are actually aware of is not the chocolate, but a subjective sensation in your mind. Whether there really is a piece of chocolate, and what it is like if it does exist, is something we still need to investigate.

Many philosophers today believe that Descartes's conception of the subject is a recipe for disaster: it cuts us off from the world and imprisons us in our minds. Thesw philosophers tend to argue that Descartes's arguments for the 'veil of ideas' are based on a radical mistake. I agree with this assessment. But it's not easy to point out what that mistake ise. Surely we *do* know our own minds better than we know anything in the outside world? And surely we *do* need to investigate whether that world even exists, and what it is like? Whether it is mistaken or not, Descartes's conception was extremely influential, and we will have to consider its continuing effect on our ideas about scientific objectivity.

Does the outside world even exist? This so-called 'problem of the external world' has been very important for the history of philosophy, but is less essential for the history of science. So we will set it aside.[3] Much more important for our purposes is the following question: what is the external world like? This question takes on an entirely new dimension once the Cartesian conception of the subject has been accepted. For once we make a distinction between the directly perceived idea of chocolate and the only indirectly perceived real chocolate, we must begin to wonder which of the properties of the idea are shared by the real object. Take the sweetness. Does that exist only in my mind, or is it the chocolate itself that is sweet? And what about the colour, or the shape, or the weight?

[3] Descartes himself argued in the following way. First, he proved the existence of a good and powerful God. He then argued that such a God would not deceive us; and in particular that God would not give us such an overwhelming inclination to believe in external objects if they did not exist. Therefore, he concludes, we can be sure that external objects do exist. Of course this argument depends completely on Descartes's proofs of the existence of God, which are almost universally seen as a failure. (I should stress that despite the negative things I'm saying about it here, Descartes's *Meditations* is a wonderful and rich text that I love to read and teach.)

Suppose that you put your hand in a fire. That hurts. But is the pain in the fire itself, is fire 'full of pain', or is the pain only in my consciousness? Clearly it is only in my consciousness. The fire acts on me in such a way that I feel pain, but that doesn't mean there's any pain in the fire itself. Descartes argues that the same is true for the sweet taste of the chocolate: there's no sweetness in the chocolate, that is just how we human beings experience it when it touches our tongue. Colour is equally subjective: our eyes and minds work in such a way that we perceive chocolate as brown, but strictly speaking the real chocolate does not have a colour.

The only properties that Descartes ascribes to the objects themselves are purely geometrical ones: size, shape, velocity. The objective world consists of nothing but bodies of particular sizes and shapes that fly through space, bump into each other, and push each other around. This world reminds one of a mechanism, a machine, and is therefore called a *mechanistic* world.

The distinction that Descartes makes between on the one hand size, shape and velocity, and on the other hand taste, colour, and so on, has become known as the distinction between *primary qualities* and *secondary qualities*. The primary qualities are the objective qualities, that is, the qualities that the object itself has. The secondary qualities are the effects that the objects have on our consciousness, in virtue of the primary qualities. Perhaps, Descartes might say, our tongues and minds work in such a way that small round particles on our tongue generate the idea of sweetness in our mind; then the objective, primary quality of chocolate is that it contains small round particles; and the subjective effect of the chocolate, that is, the effect of the chocolate on the subject, is the sweet taste, which is therefore one of its secondary qualities.

Our current scientific theories are of course very different from those of Descartes. Phycisists in the 21^{st} century talk about more than just size, shape, and velocity, having introduced terms like mass, electrical charge, magnetic field, and so on. But it's still as true for them as it was for Descartes that the objects of their science are very different from those that we perceive. A sweet taste is actually – a contemporary scientist might say – a particular kind of molecular structure. A brown colour is actually a complex combination of light waves of different wave lenghts. And while that bar of chocolate may be perceived as a solid object that fills up space, it is in reality composed of atoms that consist of tiny elementary particles and a lot of empty space.[4] Some theories even claim that space-time doesn't have four dimensions (three of space and one of time), but ten, or eleven, or twenty-six dimenions! When we do physics, then, the objective world turns out to be *radically different* from our subjective experience. It's not just that the senses sometimes deceive us, in certain extraordinary circumstances. It

[4] At least that is how it is often portrayed in popular science books and articles. The real story is more subtle, but we won't go into it here.

is that there is always and everywhere an immense gap between the-world-as-it-appears-to-us and the world-as-it-is-in-itself.

For our purposes, three results of the Cartesian conception of subject and object are especially important.

1. In everyday language, we tend to use the terms 'subjective' and 'objective' as meaning 'things that are different for different people' and 'things we can all agree on'. But the distinction between the subjective and the objective that follows from the Cartesian picture of the world is very different. In everyday language we say that while it's subjective that chocolate tastes good, it is objective that it tastes sweet. But both facts are equally subjective for Descartes. In this way of thinking, to be subjective is to have to do with *us*, with subjects; and so even something that everyone agrees on can be subjective. To be objective is to be independent of us humans, to have to do only with the objects.

2. This means that objective, scientific knowledge is not simply based on perception; for we now know that there is much in perception that is subjective. Objectivity requires stripping away everything human. It is precisely where we observe nature using tools, and where we describe it using the most abstract mathematics, that we rise to the level of objectivity. The sensation of heat is subjective; the number shown by the thermometer is objective. The beauty of a poem is subjective; the frequency with which a particular word occurs in a digitally searchable corpus of texts is objective.

3. If we achieve objective scientific insight by stripping away everything that is human, then it will be hard to find room for those sciences that we call *the humanities*. At best, they would consist in purely subjective acts of describing what happens in our own minds.

We can demonstrate this last point with an example. Take a text, any text. As an object, it is just a collection of physical particles, about which we cannot possibly ask what they *mean*. There is no meaning in the objective world. So if there is any meaning at all, it must reside in the subject, in us; to talk about the meaning of the text must be to talk about the ideas that the text causes the subject to have. But this makes any question of interpretation impossible. The meaning of the text *just is* whatever the reader thinks it means. You can't even talk about being right or wrong about the meaning of the text, since the text itself has no meaning. But this makes a mockery of the very ideas of interpretation, communication, and meaning. (And what is true about texts is equally true about visual works of art, music, human actions, cultural practices, and so on.)

Once we accept a Cartesian view of the world, the status of the humanities becomes deeply dubious. It becomes unclear whether they are possible at all. And if they are, there is a strong suspicion that they cannot lead to objective knowledge, that they are inescapably subjective. If an objectively true theory is a theory that is right about the world; and if there is nothing in the world for the humanities to be about; then there can be no objective truth in the humanities.

We now understand why we, who are still very much part of a world in which the Cartesian way of thinking is alive, run into trouble when we try to understand the role of objectivity in the humanities. If the Cartesian conception of objectivity is correct – if being objective is the opposite of being subjective, if we arrive at objectivity by removing everything that has to do with us human beings – then there simply doesn't seem to be anything that the humanities could be objective about. After all, what is more subjective than the meaning of things? If you take away all human ways of experiencing the world, what is left of a painting or a novel or a historical document except for a bunch of molecules without meaning?

There may seem to be only two ways of handling this troubling situation. The first is the scientistic attempt to turn the humanities into natural sciences, perhaps by focusing exclusively on measuring things using quantitative methods. The second, opposite strategy is the quasi-postmodern[5] attempt to claim that the humanities are radically subjective and not a place where we should expect argumentation or the search for truth. The choice is simple, though stark: if only the natural sciences can achieve objectivity, then we in the humanities must either turn ourselves into natural scientists or reject the ideal of objectivity altogether.

But of course there is a third strategy open to us: we can reject the Cartesian notion of objectivity. For could it not be the case that there is something wrong with the idea that maximal objectivity, and therefore the best kind of truth, is gained only when we take the subject out of the picture? Couldn't it be the case that objectivity and subjectivity are connected to each other rather than opposed to each other? And couldn't it be precisely in the humanities that this becomes evident? If so, then the third strategy is to rethink objectivity in such a way that it no longer undermines the humanities. It is this project that we will pursue in the rest of the chapter. In section 6.3 we will consider an influential way of undermining the idea of objective truth through the notion of perspectives. Not only will it turn out to not work, but it will actually give us the key to a better understanding of objectivity. This better understanding will the pursued in section 6.4. But first, in section 6.2, we will use the work of art as a case study for finding

[5] I'm calling this strategy 'quasi-postmodern' because it fits a certain caricature of postmodernism that one finds only too often in popular discussions. It isn't truly postmodern, since any postmodern thinking worthy of the name is, among other things, trying to get beyond Descartes's theory of the subjective and the objective.

out what is so special about the object of study of the humanities and why we may suspect that here objectivity and subjectivity have to work together rather than contradict each other.

6.2 Objectivity and the work of art

In the previous section, we described three strategies for dealing with the apparent lack of objectivity in the humanities. We could (1) transform the humanities in such a way that they only look for objective facts in the Cartesian sense; (2) drop the idea of objectivity and pursue the humanities as a purely subjective enterprise; or (3) rethink the notions of objectivity and subjectivity in such a way that they are no longer opposed. Described in this way, these strategies are very abstract. But we can connect them to our own experiences by considering our relationship to works of art.

Let's imagine something very concrete: you visit a museum. You walk into a room and see a painting by Frans Hals, the *Laughing Cavalier*. Next to this painting you find a little placard giving some basic facts: the year in which the painting was made, the materials used, the age of the sitter, a historical story of how the painting came from Haarlem to London. All of this is very *objective* and perfectly good knowledge. But it is knowledge that one could have learned without going to the trouble of visiting the museum. True, being in the museum we can learn more objective facts ourselves by looking at the painting. We see that the man has a large, upturned moustache; that he is looking straight at us; that six buttons are visible on his clothing. Fine. But again, we *could* have learned all of that from a good description written up by someone else. Learning these facts can't really be the reason for our visit. We came here for something that we could only get by looking at the painting *ourselves*; no list of objective facts can ever replace the need to see the work of art first-hand.

So far then for objectivity. What about subjectivity? It may seem plausible, even obvious, that we went to the museum for something subjective, namely, for an *experience*. We want to *feel* something. We want to stand in front of the painting and feel pleasure, perhaps, or admiration, or sadness, or joy, or perhaps even horror. We're here for the *feeling*. And such a feeling would seem to be wholly subjective. If I accompany you on your visit, and watching the *Laughing Cavalier* you feel pleasure while I experience only profound boredom, there's nothing we disagree about. We can't wonder who of us is *right*. The only thing we can do is tell each other how we felt. It's all just a matter of taste.

The subjective story I've just told is very powerful in our culture's thinking about art. But is it correct? Here's a first indication that something might be wrong with it. Suppose that we both stand in front of Picasso's *Guernica*, a powerful depiction of the suffering brought about by war. And suppose I say to you: "Wow, this is so beautiful. It fills me with joy, happiness and a profound sense of peace." That might indeed be my subjective response to the work. But you would probably feel that my subjective response is somehow *wrong*, somehow *inappropriate*. It doesn't do justice to the work. Someone who looks at Picasso's depiction of a town bombed by the fascists and feels happiness and peace would seem to be *misunderstanding* the work. All these terms – wrong, inappropriate, misunderstanding – suggest that we are using an objective standard to assess and even criticise my subjective response to the work of art.

Another way of seeing that the purely subjective story might be missing something is by considering judgements about the quality of works of art. There are certainly works of music, literature, painting, and so on, that I enjoy a lot, but about which I'm also perfectly willing to say that this enjoyment is no more than subjective. Listening to Bryan Adams reminds me of some good friends and the times, now irrevocably lost, that we spent together. This brings me pleasure, which is why I sometimes listen to *Cuts Like a Knife*. But if someone else thinks that his music is corny and melodramatic, well, I'm not going to tell them they're wrong.

However, there are other works of art where I might be surprised, aghast, *offended* even, when someone dismisses them. If you tell me that Leonard Cohen's late albums are uninteresting, that Emily Dickinson's poems bore you, or that *2001: A Space Odyssey* is only good for putting insomniacs to sleep, then I'll be livid. If I'm asked to explain the difference between my positive judgement about those works and your negative judgement about them, I'll say that you don't *understand* them, that you clearly haven't yet learnt to appreciate them, or perhaps, if I'm especially upset, I'll even go as far as saying that you are a shallow and foolish person whose opinion about art isn't worth listening to! And while that would no doubt be an overreaction, it does reveal that sometimes, we believe there to be an objective component to our aesthetic judgements.

What all of this suggests is that there is something in the encounter with a work of art – perhaps not with every work of art, but with some of them[6] – that is more meaningful, personal and emotional than the objective collecting of facts, but also more objective than my mere subjective emotional response. It suggests that art is something we can argue about and that we can attempt to understand.

It wouldn't make any sense to try to argue someone out of their hatred of rooibos tea. But it does make sense to try to argue someone out of their

---

[6] I'm making a distinction here between works of art that call for serious engagement and interpretation, and works of art that are not meant to challenge us at all. This is certainly not a black-and-white distinction. It is also not to be confused with any traditional distinction between 'high' and 'low' forms of art. An opera can be formulaic; a pop hit can be multi-layered and subtle; and it is almost certain that some of greatest works of the coming century will be computer games.

hatred of twelve-tone music or of Nigerian Afropop. Or perhaps we shouldn't say 'argue', because argument alone would never be enough. Rather, if we want to teach someone to hear what is good about Arnold Schönberg and Yemi Alade, we should listen to these pieces of music with them, discuss their inner structure, talk about their cultural context, perhaps also listen to some of the earlier music that led up to Schönberg and Alade, in order to slowly get them to the point where they can, if we are successful, come to appreciate these works. We need to *guide* their *encounters* with these works of art; something that can of course only succeed if they are willing to actually *engage* with them, to not rush to judgement, to really *immerse* themselves in the music, to be open to *being changed by the experience*.

(Speaking for myself, I find this among the most worthwhile things that criticism and scholarship can do: opening up new works for us, or allowing us to appreciate them in deeper ways than before. For example, I will always be grateful to Harold Bloom and Helen Vendler for helping me grow to love Emily Dickinson, whose poetry I at first found hard to appreciate.)

The experiences I have been describing can be clarified by looking at certain ideas of Hans-Georg Gadamer (1900–2002), a philosopher who is usually seen as one of the central thinkers in 20^{th} century hermeneutics.[7]

Gadamer points out that the interpretation of a work of art is curious in the following sense: no description of the meaning of the work can ever exhaust the work. We might say, as I did above, that Picasso's *Guernica* is about the suffering brought about by war. And clearly it doesn't celebrate that suffering, but condemns it. So we could say that the meaning of the painting is "suffering is bad, war is bad because it creates suffering, and you definitely shouldn't bomb villages." But clearly that sentence, which is rather banal, doesn't capture *everything* that the painting means to us. And even if someone were to write a 200-page essay on the meaning of the painting, it still couldn't capture the full meaning of the work. There's always more meaning there, more ways to experience and interpret it; whatever the critics write about *Guernica*, or about any other rich work of art in whatever medium, they'll never say everything that can be said. The interpretation of art can never be exhaustive.

Why is this curious? Because this is a phenomenon that we do not find in natural science. The physical system we are investigating might be enormously complicated and we might never be able to measure it with such accuracy that we can write down the physical state in perfect detail. But this lack of total accuracy is not what is happening with a Picasso painting or a Shakespeare play. It's not that we have captured all of the meaning of the painting or the play, albeit with imperfect precision; it's that there are always new approaches we can take to these works, new meanings that can

[7]I was helped immensely in writing the rest of this section by my student assistant Friso Timmenga.

be revealed, new forms that our engagement with the art can take. It's a commonplace of criticism that 'every generation has to discover [fill in name of favourite artist] anew'. Nothing like that is true in the physical sciences.

The difference, Gadamer might say, is that unlike a purely physical system, a work of art calls on us to engage with it. In order to do justice to it, in order to understand what it means, we cannot remain passive observers. First of all, we must allow the work to speak to us. Standing in front of the *Guernica*, I must open myself to what the work has to say to me; I must allow myself to be overwhelmed by it while remaining sensitive to its message. (Clearly, someone who looks at the *Guernica* and feels peace and happiness hasn't done that.) But opening ourselves to the work is not enough. I must also bring myself into the encounter. I must actively combine what the *Guernica* has to say to me with what I, so to speak, have to say to the *Guernica*. I must link Picasso's work to my own feelings, thoughts, experiences, circumstances, hopes and fears – for instance, with my own fantasies of violence and my own fears about a resurgence of fascism. It is only in the interplay between the work and me that meaning can come into existence; that truth, Gadamer might say, can *happen*.

This is not a process that reveals, once and for all, the truth, as a scientific experiment sometimes can. Rather, it is a process that will always remain ongoing. Every time we return to a work of art with the right openness, with the willingness to bring ourselves into play and be changed by the encounter, new and unexpected truth may reveal itself. That truth will be a truth about the work, but it will just as much be a truth about ourselves, as individuals and as humans. And so we can see why the humanities cannot embrace the Cartesian notion of objectivity as that which is left when everything subjective has been taken away. The objects of the humanities can only be understood when subject and object are not separated; when instead they come together in a dynamic, meaning-creating interplay. Art *cannot* be understood on the model of Cartesian science. To conclude that therefore there is nothing to understand about art is to fall into the trap of a superficial scientism.

We've taken the work of art as our central example because it most clearly exhibits the special nature of the object of the humanities. But we can say much the same thing about history. History too is something that we cannot approach 'from the outside'. We are *in* history; and the more we learn about history, the more we realise how profoundly it has shaped us. Thinking about history is always also thinking about ourselves, our past, our present, and our future. While there is much about history that can be studied in a 'scientific', 'objective' way, in the end, our understanding of history and our understanding of ourselves are linked in the same dynamic way that we see in the experience of art.

Finally, Gadamer might say, the fact that we ourselves are always part of what we study helps to explain why the humanities seem to generate knowledge that is less ‘certain’, less ‘final’ than that of the physical sciences. We will never have an external perspective on what we study. What’s more, we can never see the entirety of what we study, for that will continue to unfold its meaning and truth after we will no longer be there to experience it – after our deaths, that is. Life will be finished with us long before we are finished with life. And that means that while we may perhaps hope to reach absolute truths in physics, interpretations in the humanities are necessarily provisional. But that’s no reason to stop interpreting. As the ancient Jewish sage Rabbi Tarfon said (in a different context, no doubt):

> You are not obligated to complete the work, but neither are you free to desist from it. (*Pirkei Avot*, 2:21)

And if that’s true about anything, then surely about the exhilarating and terrifying work of self-knowledge.

6.3 The dialectic of perspectivism

All that sounds very nice, but a critic could object that in the end, whatever gets ‘revealed’ when we engage with a work of art, with history, or with another culture, it is always merely there from some subjective *perspective*. We have one perspective on the *Guernica*, but no doubt the fascists have a very different perspective on it (“we really did some good bombing that day, taught those anarchists a lesson, nice how Picasso has immortalised our victory”). Since there is no non-perspectival access to reality, all talk about truth and knowledge is only an illusion. There is only one’s subjective experience.

I believe that this approach to the question of objectivity in the humanities is misguided. But thinking in terms of perspectives is so common, and the tendency to slide from there into subjectivism is so great, that it will be useful to analyse how this happens. To do that, I will draw heavily on the work of James Conant.[8] The lessons we learn in doing so will allow us to develop a workable theory of objectivity in section 6.4.

Let us start our analysis with the basic idea of a perspective as we use it in everyday language. In everyday language, a perspective simply is the position and angle from which we look at something. In other words, a perspective is a particular spatial relation that a perceiving subject can have to a perceived object. And of course every time someone is looking at something, they do so from a perspective.

[8] Conant develops these ideas, primarily in relation with the thought of Friedrich Nietzsche, in a very long article that was published in two parts: Conant (2005), “The Dialectic of Perspectivism, I”, *Sats – Nordic Journal of Philosophy*, Vol. 6, No. 2; Conant (2006), “The Dialectic of Perspectivism, II”, *Sats – Nordic Journal of Philosophy*, Vol. 7, No. 1.

We can now ask a theoretical question: what does it mean for the possibility of knowledge that we always see things from a perspective? Is that perhaps a problem, something that suggests that we cannot have objective knowledge?

Actually, there seem to be both negative and positive aspects to perspectives. Let's take a look at the negatives first. Perspectives, we might notice, are always *partial*: we can never see the whole object from any perspective, but always only certain sides of it. Perspectives are also *plural*: the object usually looks different from different perspectives, so what we see is subjective at least to the extent that it *depends* on our perspective. Finally, perspectives *can distort.* Seen from straight on, a round coin looks round, so it looks the way it really is. But seen from an angle, the round coin looks like an ellipse, so it looks unlike the way it really is.[9]

For all that, the idea of a perspective doesn't immediately lead to skepticism. On the positive side, a perspective does afford us *a view of the object.* Furthermore, perspectives can be *altered* – we can take up different perspectives relating to the same object – and the observations we make from different perspectives can be *combined.* This combination allows us to *correct* for the partiality and distortion. By turning the coin, or moving our head, we can ascertain quite easily that the coin is, in fact, round. We also understand perfectly why it would look like an ellipse from certain angles. The perspectival nature of our perception requires us to do some work: we cannot simply believe what we see at first glance. But *if* we do the work, *then* it seems we can attain perfectly good objective knowledge.

This story about perspectives may sound surprisingly optimistic if we have been exposed to a particular type of argument. Here's an example of the type of argument I'm thinking of:

> We can only see the world from our own perspective. Therefore, any knowledge that we have about the world is necessarily subjective. Therefore, objective knowledge is impossible.

What I want to stress at this point is that the original everyday idea of perspectives doesn't justify this type of argument at all. We need a radically different view of perspectives if they are to lead to the idea that objective knowledge is impossible. But the original idea of perspectives lends itself easily to a development in precisely that direction. This natural development

[9] In the main text I'm going along with the dialectic of perspectivism, but this statement about distortion is already controversial. There is something to be said for the alternative claim that a round coin seen from an angle does not look elliptical, but simply looks like *a round coin seen from an angle.* It may take a conscious mental effort to see it instead as an ellipse. If this is true, then the everyday notion of a perspective perhaps doesn't involve the notion of distortion at all, and this idea is only introduced once we start theorising about perspectives – it is already the first step on a mistaken path.

of the idea is what Conant calls the 'dialectic of perspectivism', and we will now trace out its major stages.

One natural development is that the concept of perspective comes to be extended beyond merely viewing things from a specific angle. Suppose that we see a white shirt under a red light, so that it appears red to us. This has nothing to do with perspective in the original sense. But it easy to metaphorically extend the notion and say: "From my perspective, the shirt looks red." Indeed we usually push the concept much farther and say things like: "From the perspective of a historian, the political changes taking place in France right now look rather minor." "From the perspective of the Chinese government, the political actions of the United States government look morally bad."

This metaphorical extension of the notion of a perspective is, however, still compatible with the idea that perspectives can be combined in order to correct for distortions and achieve objective knowledge. The shirt really is white, and we have simple methods of arriving at this conclusion. The political actions of the United States government might really be morally bad (in which case the Chinese government has a non-distorting perspective on them), or morally good (in which case their perspective is distorting), or perhaps a bit of both (something that we will perhaps only realise once we have combined the perspective of the Chinese government with some other perspective, like that of the United States government).

The previous step has widened the notion of *perspective* by expanding it to non-spatial properties like colour, and even to non-perceptual properties like moral worth. But it didn't really change the way that perspectives relate to knowledge or objectivity. This will now change as we come to what Conant calls *stage-one perspectivism.* The crucial step is that at this point we start saying not merely that we have a perspective on things, but that some aspects of objects are *merely perspectival* – that is, that there is strictly speaking nothing in the object that corresponds to them. One example of that would be if we say that while things look morally good or morally bad from certain human perspectives, in fact there is no such thing as morality in the objective world. On this theory, morality would be *merely* perspectival. It's something introduced by a perspective, something to which nothing in the world corresponds.

We've already met a radical version of stage-one perspectivism in section 6.1. There we saw Descartes arguing that colour, taste, heat, and so on – all of the secondary qualities – are not objectively real, that they do not reside in the object itself. Only size, shape and velocity – the primary qualities – are really there in the external world.

There is thus a duality in first-stage perspectivism. *Some* aspects of our experience are *merely perspectival*: they are an effect of our perspective, and have no objective validity. *Other* aspects of our experience are perspectival in the original, basic sense: they may give us a distorted view of the world, but we can correct for that by combining different perspectives. The coin is not really coloured in any sense; but it is really round, and we can get objective knowledge about this.

But how can we defend this duality? How do we know that colour is merely perspectival, but that shape is part of the objective world? Once we have entered stage-one perspectivism, it seems natural to start doubting the veracity of *all* aspects of our experience. Perhaps the world is *totally* different from how it appears to us. Perhaps a true theory of the objective world and our perspectives on it would have to be formulated in terms that are completely unfamiliar to our normal ways of thinking. This is what Conant calls *stage-two perspectivism.*

We've briefly looked at one version of stage-two perspecitivism: the view that modern physical theories describe the world as it really is, and that it turns out to be wildly different from anything we encounter in experience. If the world doesn't consist of three-dimensional objects, but of twelve-dimensional collections of vibrating strings, then it's a very different world from what we naively thought![10] But the point of these scientific theories is still to explain our experience; to explain, starting from what is objectively real, why the world looks to us the way that it does.

But once we start taking seriously the idea that our perspectives on the world are radically different from the way the world is in itself, how can we then hold on to the idea that objective knowledge about the world is possible at all? Why should those scientific theories, for instance, be thought of as objectively true, rather than as merely true *from our perspective*? How can we ever hope to break free from our perspective to see the world as it really is? If we accept that we cannot, that the real, objective world is forever hidden from us and that we are as it were trapped in our own perspective, then we have arrived at *stage-three perspectivism.* It is only here that the argument given on page 98 starts making sense. If we are trapped in our perspectives; if we can never see or understand the world as it really is; then, surely, objective knowledge is impossible.

Arguments like these are heard regularly in debates about objectivity. But we should notice how far we have come from the original notion of a perspective. The simple fact that you and I may have a different perspective on some aspect of reality doesn't prove that there is no objective knowledge. We also have to think of those perspectives as *deeply distorting*, and of ourselves as *completely unable to compensate for those distortions.*

What's more, there is something unstable even about stage-three perspectivism. If we are trapped in our perspectives and cannot gain any know-

[10] I want to stress that this is not a claim made by modern science, but a claim about how to understand and interpret modern science. One could, for instance, also believe that scientific theories are merely tools for prediction; or that they are true, but not more true than our common experience. Debates about these issues are known as debates about *scientific realism.* We will not go into them here.

ledge about an objective world, why would we even believe in an objective world? What reason do we have for thinking of our perspectives as perspectives *on* something that exists independently of these perspectives? Can we even continue to make sense of the idea of objectivity? Possibly not. And so we may come to say – and this is what Conant calls stage-four perspectivism – that there are only perspectives, and *nothing else*.[11]

But what is this supposed to even mean? What is left of the notion of a perspective if we come to use it in such a way that it is not even a perspective *on* something, that there is nothing we actually see or understand through our perspectives? Certainly all the connections with the original, everyday notion of a perspective have been severed. It is clearly true that you and I can have different perspectives on a coin or a building. It seems relatively unproblematic to extend the notion of perspective and say that a 21^{st} century historian researching early Christianity and a 14^{th} century monk have a different 'perspective' on the Bible. But something radical has happened by the time we conclude from our thinking about perspectives that there can be no such thing as objective knowledge, or that the very idea of objective knowledge makes no sense. Something so radical that it would seem to jeopardise the very possibility of science and scholarship.

I'm not claiming that I've just given a knockdown argument against any strong form of perspectivalism. But I hope to have demonstrated that it is completely possible to be impressed by the fact that we all experience the world from a particular perspective, and to at the same time believe that some kind of objectivity is possible. Here is how Conant (2006, p. 47) describes the mature position of Friedrich Nietzsche:

> (1) if something is to count as a *perspective* (as opposed to a mere illusion), it must afford a *glimpse* of the object which it is a perspective on, (2) we must *alternate* perspectives if we wish to gauge the object accurately, (3) the glimpses afforded by one perspective must be *compared* with those afforded by another, (4) only through such *triangulation*[12] can we attain knowledge, (5) the more such glimpses we can compare the more *objective* our knowledge may become.

---

[11] This kind of point may be more familiar when couched in terms of truth. Suppose that somebody says that "there are no objective truths, there's just people's beliefs." Can we make sense of the idea of believing something without believing it to be an objective truth? When I *know* that something is merely subjective – e.g., that chocolate ice cream is the best ice cream – I don't *believe* it. I don't believe chocolate ice cream to be the best, I just prefer chocolate ice cream to all other flavours. When we discard the notion of objective truth, we seem to discard the notion of belief too, as well as the notions of argument, investigation, conclusion, theory, and so on. One may wonder whether, if we go that far, there would be anything left to talk about.

[12] *Triangulation* is a name for the process of combining different perspectives in order to attain a fuller, more objective view of the object.

This actually fits quite nicely with many of the things that we regularly do in the humanities. We try to become aware of our own perspective on things and how it might distort our view of the object; and we then try to combine different perspectives in order to get a better picture of what we want to know. Very often this is a *social* process: different people have different perspectives, and as a community of researchers we try to bring those perspectives together, critically compare them, and improve our theories and interpretations.

Perhaps it is precisely in this that objectivity exists? As we have seen, the Cartesian tradition thinks of objectivity as what remains once you strip away all traces of the subject. But it might be much more fruitful to think of objectivity as what results when subjects work together to transcend their particular perspectives, to combine all their individual views into a better view. This would be to connect objectivity to *intersubjectivity*. It is a major theme of social and feminist epistemology, and we will delve into it in the next section.

6.4 Objectivity and intersubjectivity

NOT YET WRITTEN. (Probably talk about Helen Longino and Sandra Harding.)