Philosophy of the Cultural Sciences

Preliminary draft

Staffan Carlshamre

January 14, 2016
# Contents

1 Science, knowledge and truth ................................. 10
  1.1 Words and what they mean ............................... 10
  1.2 Kinds of sciences ....................................... 11
  1.3 What science is and what it ought to be ............... 12
  1.4 Science and knowledge .................................. 13
    1.4.1 Knowing how and knowing that ....................... 13
    1.4.2 A sociological concept of knowledge ............... 14
    1.4.3 The classical concept of knowledge ............... 14
    1.4.4 A basic concept? .................................. 15
    1.4.5 The case of the flat earth ......................... 16
    1.4.6 The pragmatics of “know” ......................... 17
    1.4.7 How good must good reasons be? .................. 18
    1.4.8 Whose beliefs? .................................. 19
    1.4.9 Epistemic injustice ................................ 20
  1.5 What is truth? ........................................... 21
    1.5.1 Truth as correspondence ............................ 22
    1.5.2 A minimal theory of truth .......................... 23
    1.5.3 Truth is not the problem ........................... 23
    1.5.4 Truth by convention? ............................... 24
    1.5.5 Useful belief ..................................... 25
    1.5.6 Unknown truth ..................................... 26
  1.6 The problem of demarcation .............................. 28
    1.6.1 Science and its twins ............................... 29
    1.6.2 Science and its ancestors ........................... 31
    1.6.3 Aspects and primacy ................................ 34
    1.6.4 Science and common sense ........................... 36
    1.6.5 Knowledge and human interest ....................... 36
    1.6.6 Science and its neighbors ........................... 38
    1.6.7 Does it matter? .................................... 40

2 Questions, answers, evidence ................................ 42
  2.1 Questions ............................................... 43
  2.2 Deductive inference ..................................... 44
  2.3 Do I have a hypothesis? ................................ 45
  2.4 Testing a hypothesis .................................... 47
    2.4.1 A real case ....................................... 48
    2.4.2 What about the hat? ............................... 50
    2.4.3 Radiocarbon dating ................................ 51
  2.5 What is evidence? ....................................... 54
## CONTENTS

2.6 Inference to the best explanation ........................................ 56
   2.6.1 Testimony ...................................................... 57
   2.6.2 Interpretation and explanation ................................. 58
   2.6.3 What explanation is the best? ................................. 59

2.7 Probability – Thomas Bayes’ theory of evidence ....................... 61
   2.7.1 Prediction ...................................................... 61
   2.7.2 Likelihoods and evidentiary value* ............................ 63
   2.7.3 Prior probabilities ............................................ 63
   2.7.4 Testing for a disease .......................................... 65
   2.7.5 Bayes’ theorem* .............................................. 66
   2.7.6 What is probability? .......................................... 68
   2.7.7 Conclusion ..................................................... 70

3 Meaning ............................................................................. 73
   3.1 Concepts of meaning .................................................. 74
   3.2 Facts and interpretations ............................................ 74
   3.3 Natural meaning ....................................................... 75
   3.4 Meaning and subjectivity ............................................ 77
   3.5 Husserlian meaning: intentionality ................................ 78
      3.5.1 Words, things, concepts ...................................... 79
      3.5.2 Concepts in perception and thinking ....................... 80
      3.5.3 The heaven of the Greeks .................................... 81
      3.5.4 Nested intentionality .......................................... 83
   3.6 Weberian meaning: action .......................................... 83
      3.6.1 Thick and thin ................................................ 85
   3.7 Barthesian meaning: manipulation ................................ 87
      3.7.1 The Applicant .................................................. 88
      3.7.2 The Photo ..................................................... 89
      3.7.3 Film Music ...................................................... 90
      3.7.4 Meaning in nature ............................................ 91
   3.8 Gricean meaning: communication ................................ 92
      3.8.1 Relevant features and modes of correlation ............... 93
   3.9 Summary ............................................................. 94

4 Interpretation ..................................................................... 97
   4.1 The intentional fallacy ................................................. 98
   4.2 The hidden and the manifest ...................................... 98
   4.3 Leaving language behind .......................................... 101
   4.4 Interpretation and description .................................... 102
   4.5 Reading and claim ................................................... 105
   4.6 Interpretation and rationality .................................... 108
      4.6.1 Meaning holism and the hermeneutic circle ............. 109
      4.6.2 Pre-understanding ............................................ 111
      4.6.3 Relevant totalities ............................................. 113
      4.6.4 Relevance relations .......................................... 118
      4.6.5 Perfection and charity ....................................... 118
      4.6.6 Radical interpretation and language learning ........... 118
      4.6.7 Dialogue and triangulation .................................. 118
      4.6.8 Imperfect rationality ......................................... 118
      4.6.9 Horizons and relativism ...................................... 118
CONTENTS

4.7 Interpretation and translation . . . . . . . . . . . . . . . . . . . . . . . 119
4.8 The death of the author? . . . . . . . . . . . . . . . . . . . . . . . . . 119

5 Frameworks and relativism 120
5.1 Relativism and constructivism . . . . . . . . . . . . . . . . . . . . . . 121
5.2 Reality as a construction . . . . . . . . . . . . . . . . . . . . . . . 121
5.3 What is reality? . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 123
5.4 Two criteria . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 124
5.5 Intentional mediation . . . . . . . . . . . . . . . . . . . . . . . . . 125
5.6 Reality as a social construction . . . . . . . . . . . . . . . . . . . . 126
5.7 What is relativism? . . . . . . . . . . . . . . . . . . . . . . . . . . . 127
5.8 What is a concept? . . . . . . . . . . . . . . . . . . . . . . . . . . . 128
5.8.1 Words and concepts . . . . . . . . . . . . . . . . . . . . . . . 129
5.8.2 Concepts: classification . . . . . . . . . . . . . . . . . . . . . 129

6 Social construction 131
6.1 Social facts . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 132
6.2 The power of shared belief . . . . . . . . . . . . . . . . . . . . . . . 133
6.3 Is gender a social fact? . . . . . . . . . . . . . . . . . . . . . . . . . . 135
6.4 Kinds of rules . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 137
6.5 Fetishism, reification and alibi . . . . . . . . . . . . . . . . . . . . . 139
6.6 Identification . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 141
6.7 The body as artifact . . . . . . . . . . . . . . . . . . . . . . . . . . . . 143
6.8 Performatives and the division of linguistic labor . . . . . . . . . . . 145
6.9 Knowledge and power . . . . . . . . . . . . . . . . . . . . . . . . . . 148
6.10 Knowledge as a social status . . . . . . . . . . . . . . . . . . . . . 150
6.11 Two fallacies . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 151

Copyright © Staffan Carlshamre 2014, all rights reserved.
Introduction

The occasion for this book was an urgent need connected to a course. The philosophy department had agreed to give a course on the philosophy of science for master students in the humanities, and we needed a book to go with it. We took it for granted that there already was a suitable book, and went looking for it, but we didn't find it. To be sure, there were many introductions to the philosophy of science, but they all seemed to be cut from the same cloth and not fit for our purposes – although often excellent in themselves, they were ruled out by possessing certain features that from our perspective were bugs. One obvious problem was that there were no books at all with the right scope – there are many introductions to the philosophy of the social sciences, there are introductions to special parts of the humanities, like the philosophy of history or literary theory, but none that target the philosophy of the humanities as such.

One route, of course, would have been to use a general introduction to the philosophy of science and supplement it with humanities material as the need arose, but two considerations told strongly against that solution. First, all such purportedly general texts take the natural sciences as paradigmatic for science generally, drawing their examples from physics and chemistry and often raising doubt whether the humanistic disciplines belong to the realm of science at all – a clearly off-putting attitude in relation to our intended audience. To be sure, we wanted to treat the humanities within the perspective of a general philosophy of science, but we wanted to do it in a much more egalitarian spirit. Second, and perhaps paradoxically, such introductory texts almost invariably take a historical approach, seemingly presupposing that the first thing everybody needs in the philosophy of science is a crash course in its 20th century history: start from logical positivism, go on to Popperian rationalism, challenge it all with a little bit of Kuhn and perhaps a glimpse beyond, and you're done. That was not what we wanted our students to learn.

So we decided that we needed our own book, and I proceeded to write it. I will use the rest of this introduction to indicate and motivate the plan for the book, the sort of issues it covers, and the level it is aiming for. Many of the topics touched upon will, of course, be developed much further later in the book.

Science, the cultural sciences and the humanities

At Stockholm university, the faculty of the humanities comprises the historical disciplines, including archaeology but somewhat randomly excluding
economic history (which is counted among the social sciences); the aesthetic disciplines; and the linguistic disciplines, including not only general linguistics but also the different language departments, with each of these in turn containing linguistic and literary subdivisions. But this is not all: apart from the large groupings of the historical, the aesthetic and the linguistic, there are also outliers, like philosophy, ethnology, religious studies, gender studies and media studies – not to mention that the faculty has recently become host to an array of didactic studies, relating to teacher education.

One may doubt, of course, that such a seemingly haphazard collection of different subjects has enough real unity to merit treating them together from a philosophical and methodological point of view. In particular, one is tempted to point out that some of these disciplines, for example linguistics or ethnology, would be better placed among the social sciences, and that there is ample scope within some of them, for example phonetics or archaeology, for methods and theories from the natural sciences as well. But in voicing these concerns, one seems to reveal some deeper intuitions about what belongs where – indicating that there may be, after all, some real differences underlying the tripartition into the humanities, the social and the natural sciences, distinctions that may be interesting and useful even if they do not exactly fit the administrative classification of academic disciplines.

Most attempts to find such deeper characterizations of the main areas of science go back to the 19th century German discussion about the Naturwissenschaften and the Geisteswissenschaften – the latter category including not only the humanities but also the social sciences, a scope that I have tried to capture in the expression “cultural science”. Two suggestions have stuck in the philosophical memory, and I will start from them.

The first suggestion is that the sciences of culture differ from the sciences of nature by being about meaningful phenomena, and therefore essentially relying on interpretation and understanding. There is obviously something right in this idea. People and their actions, speech, artifacts, documents of different kinds, works of art and so on, are at the center of attention, albeit in different ways, for the historian, the literary scholar, the ethnologist, the economist, and all the other students of human culture. This is not to deny that there is room for meaning, understanding and interpretation, in some quite respectable senses of these words, in the natural sciences as well, but it would be just as odd for a book on the philosophy of the natural sciences to include special chapters on meaning and interpretation, as it would be for a book on the humanities to exclude them – here they are chapters 3 and 4.

I will argue that understanding, in the main relevant sense, involves being able to see things not only through one’s own eyes, so to speak, but through the eyes of others: understanding others involves understanding how others understand. This implies that possible conflicts between different ways of understanding is built into the cultural sciences from the start, and it is an important part of the philosophy of the cultural sciences to supply the tools for thinking about such conflicts. It is not a quirk or a passing fad that makes questions about relativism and historicism abound in methodological discussions of the humanities. To the chemist such con-
cerns are philosophical marginalia – interesting, perhaps, but not essential to her main preoccupation – to the anthropologist or historian they touch the core of what their work is about. Needless to say, this makes it all the more important to tread carefully over ground where any attempted short cut is likely to lead one astray, and I do my best to chart the territory in chapters 5 and 6.

The second old German suggestion, about what sets the Geisteswissenschaften apart from the Naturwissenschaften, is that they are idiographic, describing the individual case, rather than nomothetic, positing general laws. As a demarcation between the cultural and the natural sciences this is quite obviously hopeless. What individual event has received more scientific attention than the Big Bang? And who can deny the interest of economists and narratologists in general principles? But taken in a less absolute way, as describing competing tendencies or main attitudes, I think the distinction between the idiographic and the nomothetic still has something to say about differences between the main areas of science.

One way to get a glimpse of this is to look at the role of theories in, say, physics and in literary studies. In physics a good theory is the pinnacle of scientific achievement, that towards which the whole enterprise is directed. Individual cases are important as well, of course, but only as concrete instances of general laws, mostly in the form of controlled experiments. Their role is to serve theories, experiments are conducted to falsify or confirm theories, and have no independent interest. In literary studies, the roles are reversed. A theory is supposed to prove its mettle by what it brings to the understanding of individual works. This is why, in the humanities, the word “theory” occurs mostly in the compound “theory and method”, which is uttered as a tautology, although to a philosopher it has the air of an oxymoron. What has method, a way of achieving something, to do with theory, a general assertion about reality? The answer is that theories are primarily conceived as tools for extracting new and exciting content from old works – as a consequence, theories are consumed rather than confirmed by successful applications. Now that we have psychoanalytic and deconstructive readings of Hamlet – what’s next? The aesthetic disciplines are perhaps extreme in this regard, even among the humanities, but the same tendency is obvious for example in in history: the focus is on deep and multidimensional understanding of individual cases rather than on generalities.

That there are examples from the humanities that fit the idiographic mould is not surprising, of course, and it is also to be expected that the historic and the aesthetic disciplines should provide the best fit – after all those were the paradigmatic original Geisteswissenschaften, while the social sciences were still in their infancy. Indeed, social sciences like sociology and economy seem to provide obvious counterexamples to the idiographic model, using individual cases mainly as material for statistical summary and computation, in search of general trends and principles. What should we make of this? Here is a suggestion. The distinction between the (mainly) idiographic and the (mainly) nomothetic sciences does not split the cultural from the natural sciences. Instead it separates the humanities, on one side of the divide, from the social and the natural sciences, on the other side.

If this is right, we can work the two German suggestions in tandem to effect the classical tripartition of the empirical sciences. The humanities are
meaning-based and idiographic; the social sciences are meaning-based and nomothetic; the natural sciences are nomothetic and not meaning-based.

Armed with this scheme one might proceed to evaluate existing administrative groupings, as they have panned out over the course of time, and confirm, for example, that Stockholm has presumably misplaced linguistics – it belongs with the social sciences and not with the humanities. But this would be a rather pointless exercise, and most cultural sciences probably contain all the relevant ingredients, just in different proportions: it is the dimensions of the scheme that are important and not the taxonomic result. I will just make a couple of remarks, which are relevant for the content of this book.

As sketched, this take on the demarcation between the humanities and the social sciences obviously has something to do with another dichotomy that often pops up in this connection: that between qualitative and quantitative data. On the face of it, this distinction seems concerned with the form in which observations are recorded: quantitative data are in numerical format. But it also has to do with numbers in a more basic regard: if data shall be useful for statistical purposes, there must be enough of it. And this in itself runs counter to what I pointed to as the hallmark of the idiographic attitude: deep and multidimensional understanding of the individual case. Evidence in the humanities is typically gathered by procedures such as close reading or not too closely structured interviews; the paradigmatic form for data collection in the social sciences is the questionnaire, asking for the rating of set alternatives on a numerical scale. The latter type of data are just as meaning based and interpretation dependent as the former – the researcher expects the subject to interpret the questions in a certain way, and hopes for cooperative and informative answers. But nuances and individual variations will inevitably be discarded in the interest of having large enough quantities of comparable data.

In this dimension this is a humanities book: there is (almost) nothing here about statistical methods, sampling procedures or extensive data collection. The reason to keep the more inclusive “cultural sciences” in the title is that “the humanities” is in all other respects unduly restrictive in relation to the actual content. The chapters on meaning and interpretation are relevant for all meaning based sciences, the same thing goes for the chapters on relativism and social construction, and the first two chapters paint a picture of science that is meant to be relevant for the natural sciences as well.

So far, I have concentrated on describing and motivating the inclusion of themes specific to the cultural sciences, but maybe I need to defend the opposite front as well? Many students of the humanities are used to methodology courses giving ample space to meaning and interpretation, but at least some of them are less accustomed to speaking of their research in terms like hypothesis and evidence or even knowledge and truth. Again, this goes back to the old disputes over the natural versus the cultural sciences – returning in different terminological dressings, for example as hermeneutic versus positivistic, or, in the currently most popular guise, as qualitative versus quantitative research. It is natural in such disputes to concentrate on the differences, to forget about the common ground and to slide from the observation that X is not a science quite like Y to the conclusion that
it is not a science at all. Here I unabashedly take the standpoint that the relevant terms are properly used to mark species within a common genus – divisions within the realm of science, not frontiers between hostile empires – and that the similarities are at least as important as the differences.

Who is it for and how is it done?

As I said in the beginning, this book is primarily intended for advanced students in the humanities, where that label is taken in a liberal and inclusive sense. Most such students will have taken methodology courses before, perhaps containing deep and difficult material relating to the methodological traditions of different disciplines and university departments, and I owe them an explanation of how this book is likely to be different from what they have already encountered.

First, it is not historical, or organized by different standpoints, perspectives or schools of thought. There are chapters on meaning and interpretation, but no chapters on hermeneutics or deconstruction; there are chapters on relativism and social construction, but no chapters on Foucault, Latour or Butler. This does not mean, of course, that sources are not referenced or that the discussion is carried on without regard to the relevant literature, it just means that sources are invoked when the discussion demands it, rather than being used for setting the agenda.

Instead, it is problem oriented and systematic, discussing questions as they arise without undue regard to where they come from or in what tradition they originated. Indeed, if there is an overall credo in the book it is perhaps the possibility of a reasoned and systematic approach to questions of methodology, and against the voluntaristic relativism sometimes voiced in methodological discussions in the humanities, that seems to suggest that inquiry begins with a more or less arbitrary choice of methodological perspective.

Second, it is not a book of methods or recipes for doing research. Such things are much too specific to different subjects and topics to be treated on this level of generality. Instead, the overall aim is to supply some of the tools that an advanced student, perhaps aiming for a PhD, needs in order to reflect independently and critically on the aims and procedures of her own discipline, within the context of a reasonable image of science as whole.

What are the overall aims and value of science, in general and in relation to my own discipline? What is evidence? What counts as evidence for my type of research and how is it related to the sorts of result that I am aiming for? What am I looking for when I try to find the meaning of this or that?

So this is not a handbook for doing research, but I hasten to add that I have still tried to make it useful and practical. In suggesting and developing philosophical arguments, I have tried not to lose sight of the ways they may be relevant for the actual practice of research, and how they relate to ongoing controversies within the relevant disciplines.

Third, it is normatively oriented, without preaching any particular faith. It is impossible to do scientific work without opinions, however implicit or inarticulate, of how research should be done. The purpose of this book, just like the the course it is written for, is to help the reader to develop
a reasoned and articulate view of science and scientific research, with reference to her own subject but within the context of a general understanding of what science and research is and what it is for.
Chapter 1

Science, knowledge and truth

This book is about science, about scientific method and about the norms and values of science. So it looks as if we ought to start by getting clear about what science is, perhaps by defining the word "science". And as we do that, other words are bound to pop up and ask to be explained and defined, too — words like "knowledge", "reason" and "truth". But are words really so interesting — wouldn’t it be better to go to the things themselves?

1.1 Words and what they mean

Let us begin with a simple, yet philosophical, question. How many letters are there on the next line?

A B A

Two answers spring to mind: there are two or three letters. But which answer is the right one? It depends, of course, but on what does it depend? It depends on what we mean — in this case on what we mean by the word "letter". If we are talking about letter-types there are two letters on the line, if we are talking about letter-tokens there are three.

Trivial? Of course, but the important insight is that the answer to almost every philosophical question begins the same way: “it depends on what you mean”. Is science objective? It depends on what you mean by "science" and "objective". Is meaning subjective? Is truth relative? Is Russia a democracy? Is sex a social construction? It depends on what you mean!

Does this imply that truth is conventional and depends only on what you mean? Of course not — when we have decided what we mean we still have to count the letters. But there is no point in looking for the answer before we know what the question is.

A famous, and less trivial, example of the same kind of difficulty is told by the American philosopher and psychologist William James. In his own words:

Some years ago, being with a camping party in the mountains, I returned from a solitary ramble to find every one engaged in a ferocious metaphysical dispute. The corpus of the dispute was a
squirrel – a live squirrel supposed to be clinging to one side of a
tree-trunk; while over against the tree’s opposite side a human
being was imagined to stand. This human witness tries to get
sight of the squirrel by moving rapidly round the tree, but no
matter how fast he goes, the squirrel moves as fast in the op-
posite direction, and always keeps the tree between himself and
the man, so that never a glimpse of him is caught. The resultant
metaphysical problem now is this: Does the man go round the
squirrel or not? (James; 1907, p 43).

Experience teaches that this dispute will be repeated in almost any group
of people where the problem is presented. It shares a feature with the A B A
question: the root of the problem is not in the facts. Everybody agrees about
what the situation is and what happens there – more factual information
would not make the dispute go away. The dispute is verbal, not real, and
the solution involves making a distinction to clarify how we use words, in
this case the word “round”. William James again:

If you mean passing from the north of him to the east, then to
the south, then to the west, and then to the north of him again,
obviously the man does go round him, for he occupies these suc-
cessive positions. But if on the contrary you mean being first in
front of him, then on the right of him, then behind him, then on
his left, and finally in front again, it is quite as obvious that the
man fails to go round him, for by the compensating movements
the squirrel makes, he keeps his belly turned towards the man
all the time, and his back turned away. (James; 1907, p 44).

Such disputes, where the problem is not the facts of the matter but what to
say about them, arise again and again in philosophy. To recognize them and
to be prepared to solve them by making distinctions and clarify what one
means, is an important philosophical skill, that we will have the occasion
to exercise many times in the following.

1.2 Kinds of sciences

A first observation about the word "science" is that it is used in broad and
narrow ways. Biology and physics are sciences, by any reckoning, but what
about history and literary studies? Here, I will use "science" broadly, to
cover not only the natural sciences, but also the whole field of the social
sciences and the humanities. This does not mean that we will be blind to
differences among them – to the contrary, such differences will be among
our most important topics.

Within such a broad conception of science there are some traditional di-
visions that may be worthy of mention at the outset. First, we have the
distinction between empirical sciences and sciences that are apriori – the
latter category, as traditionally conceived, comprises only mathematics and
logic. The tripartite division between the natural sciences, the social sci-
ences and the humanities falls within the empirical sciences. Here, our
main subject is empirical science and in particular the social sciences and
the humanities – I will sometimes speak of them together as the cultural sciences.

(One should not expect every scientific discipline to fall squarely within exactly one such category. Generally speaking, one would presumably say, for example, that linguistics and archaeology are social sciences, but that does not exclude that there is large place in them for methods and theories emanating from physics, chemistry and biology – and perhaps from literary theory and the humanities as well.)

1.3 What science is and what it ought to be

One way to take the question about what science is, is as a descriptive or factual question. As such it can itself be studied by science, and perhaps most importantly by some of the social sciences. Scientific institutions are extremely important in modern societies, and by that token it is also important to know a lot about them, about how they work internally, and how they interact externally with the rest of society. Interest in the history and sociology of science has been intense during the last fifty years or so, often inspired by combined philosophers-historians like Thomas Kuhn and Michel Foucault. Philosophical conclusions inspired by such work will play a large role on our agenda, but our primary focus will not be on factual questions about science as it is actually practiced.

Another way to take the question about the nature of science is as a normative question, about what science ought to be, about the proper aims and methods of research. Such questions are at the heart of the traditional philosophy of science, that has often been harshly critical of the actual practice of working scientists. Most often, the normative essence of science has been taken to reside in a certain method, the scientific method, to be used by scientists, and perhaps in a certain attitude, the scientific attitude, that they should employ in their research. The scientific method is usually described in logical terms – as concerning the specific types of reasons or arguments that are relevant for scientific assertions – while the scientific attitude is usually taken to comprise such virtues as impartiality, open-mindedness and critical distance to the opinions of oneself as well as of others. A typical modern exponent of this type of approach to science is Karl Popper, and we will to some extent use his views as points of departure for our own discussions.

The descriptive and the normative perspectives are not independent of each other, of course. I have already stressed the importance of the normative perspective for a critical and evaluative discussion of actual science. To be able to judge some examples of science as better than others, to distinguish between good and bad scientific practice, and perhaps to judge some purported examples of research to be not really scientific at all, one must have an opinion about what science ought to be.

On the other hand, scientific norms can themselves be evaluated and criticized, and critics will often point out that a purported normative account of science does not fit what even the best scientists actually do. This is, for example, the form of Kuhn’s polemic against Popper. Kuhn argues that not even the examples of scientific research that are universally taken
to be the very best actually conform to Popper’s methodology – and, moreover, that adherence to these methods would actually have hampered or even halted scientific progress. Of course, to get that argument under way Kuhn must rely on his own normative opinions of what science aims for and what constitutes scientific progress.

1.4 Science and knowledge

A very natural idea is that science has something to do with knowledge. Indeed, the very word “science” comes from the Latin “scientia” which simply means knowledge. But science and knowledge are not the same. On the one hand, there is a lot of knowledge that is not scientific knowledge – we all know things that we have not learned from science and that have never been the subject of scientific research. On the other hand, science comprises much that is not knowledge. Open questions, unconfirmed hypotheses, bold conjectures and wild guesswork are important parts of the scientific process. And many theories that once were taken to be known, and in some cases to be almost beyond doubt, have subsequently been shown to be false without thereby ceasing to be scientific – with Newton’s mechanics as the most celebrated example.

But even if science is not the same thing as knowledge, or even as a special kind of knowledge, it may be true that science aims for knowledge – that the most important purpose of scientific practice is to produce knowledge, or perhaps a special kind of knowledge: scientific knowledge.

1.4.1 Knowing how and knowing that

It will come as no surprise that the word “knowledge” can also be used in many ways. Here are two concepts of knowledge that would figure in almost any comprehensive account.

First, there is knowledge as knowing-how, as the possession of a skill. I know how to swim and how to ride a bike. I speak my mother tongue and perhaps, to a lesser extent, a couple of other languages as well – these are skills that I possess, things that I can do, that I know-how to do.

The philosopher Michael Polanyi (1967) spoke of such skills as “tacit knowledge”, to emphasize the fact that I may possess them without being able to articulate or describe them. Some of us may know a bit about the theory of keeping afloat in water, or staying up on a bike, or the grammar of our native language – but the theory has come after the skill, and the know-how can live perfectly well without the theory.

Tacit know-how of different kinds are essential to any human practice, including science. Advanced education in any particular science is largely concerned with training the special skills one has to possess in order to do the relevant research, but, nevertheless, such skills are the means rather than the aims of research. This is not the kind of knowledge that it is the purpose of science to produce.

Second, there is knowledge as knowing-that – “propositional knowledge”, as philosophers sometimes call it. I know that Buenos Aires is the capital of Argentina, and I know that the moon is much smaller than the
earth. What comes after “that” in such a knowledge claim is a sentence that expresses a proposition, something that I take to be a fact. Such knowledge is typically expressed, or at least expressible, in language, and if science aims for knowledge it is primarily for propositional knowledge. (From now on, I will only use the word “knowledge” for propositional knowledge, if I do not expressly say otherwise.)

What about non-verbal representations? Traditional accounts of science has largely ignored the use of pictures, diagrams and maps in science – most of the philosophical discussion of other forms of representation than language has been confined to aesthetics.¹ That is unfortunate and about to change, I think, but for the moment it is enough to note that image-like representations may also be taken to code propositional knowledge: a roadmap of Sweden may inform you that Stockholm is due south of Uppsala, at a distance of about 70 km, for example.

1.4.2 A sociological concept of knowledge

But even with knowing-how out of the way, there are several different uses of “knowledge” to take into account. In the social sciences the word knowledge is often used for something like “socially entrenched belief”. When sociologists speak about how “knowledge” is produced or distributed in a certain society or social group, there is usually no implication that the relevant opinions or beliefs are true, or well supported by reasons in a normative sense – that something is “known” in this sense just means that it is taken to be known by the relevant group of people, in another sense of “known” that we will soon come back to.

This “sociological” concept of knowledge is important for us, because it is often used in discussions about science. Empirical studies of science often want to stay neutral with regard to the knowledge claims made within the studied scientific practice, and if the word knowledge is used, it is, so to speak, surrounded by invisible quotes.²

The important thing, of course, is not to confuse this “quoted” conception of knowledge, with the stronger (and, as I will argue, more basic) concept of knowledge that is the subject of the next section. It is, for example, trivial that knowledge in the sociological sense is relativistic: as long as socially entrenched beliefs vary between different societies, “knowledge” in the sociological sense will vary as well.

1.4.3 The classical concept of knowledge

There is a conception of knowledge that has been around so long in philosophy that is is often referred to as the classical conception – it is usually

¹The most sustained discussion by a major philosopher of non-verbal systems of representation is by Nelson Goodman, particularly in Languages of Art (1976). Daston and Galison (2007) gives a fascinating story of scientific atlases, i.e., books of images, and the different epistemological ideals that they embody.

²This way of describing the sense of “knowledge” relevant to the sociology of knowledge is quite explicit for example in Berger and Luckmann (1967, pp. 13–14) and in David Bloor (1991, p 5). For an argument that knowledge “really” is a social status, in a stronger sense, cf. Martin Kusch (2002).
taken to be introduced by Plato. It can be briefly expressed as “knowledge is warranted, true belief”, but it is worth the effort to be a little more explicit.

The sort of knowledge to be captured by the classical conception is propositional knowledge, knowledge that can be expressed by the formula:

\[ A \text{ knows that } p \]

where \( A \) is a person and \( p \) is a proposition. According to the classical analysis, there are three conditions that must be fulfilled if an assertion of this form is to be true. The first one is

(1) \( A \) believes that \( p \)

Knowledge implies belief – I cannot be said to know that Buenos Aires is the capital of Argentina if I don’t even believe it. But belief is not enough for knowledge, it must also be the case that Buenos Aires actually is the capital of Argentina. No matter how strong my belief, I can never be said to know that the capital of Argentina is Rio de Janeiro. So the second condition is

(2) \( p \) is true

Knowledge implies truth. But, as Plato convincingly argues, even true belief may fall short of being knowledge. I may, for example, buy a lottery ticket because I strongly believe that it will win – maybe because it hangs over the counter in the shop on a clip with my lucky number on it. And perhaps I do win, so that my belief is true. But do I know that the ticket is a winner, just because I believe it and it’s true? Hardly, there is still something missing. According to the classical analysis, the third condition is:

(3) \( A \) has good reasons for \( p \)

My belief that I will win at the lottery is pure guesswork, and it’s truth is just a lucky coincidence, and so it is not knowledge.

The classical conception is not uncontroversial, but has, in fact, been the focus of a lot of debate during recent decades. The controversial point is, almost always, the third condition. It is generally granted that (1) and (2) must hold – you cannot be said to know what is false or what you don’t even believe. And it is granted that there must be a third condition: true belief is not enough. But is it always the possession of good reasons that finally tips the scales? It has been argued both that (3) is not sufficient – that I may have good reasons for my true belief and still not know – and that it is not necessary – that I may sometimes know without reasons (for example in perception). But even so, the classical analysis seems to cover many common cases of knowledge, and, in particular, it is a natural candidate for the sort of knowledge sought by science – the aim, or at least one important aim, of science is to discover what is true and give us good reasons to believe it.

1.4.4 A basic concept?

Though there is room for disagreement over details, I think that it is fairly clear that the classical analysis comes close to capture a concept that we all
have, and which is in some ways the basic concept of knowledge. It is, for example this concept that we teach our children, when we teach them to distinguish between knowledge, belief and wide-spread opinion. Imagine little Ethan coming home from school in a rage, saying that his ball is gone and that Jonathan stole it.

– Do you know that he took it?
– Everybody says so!
– Yes, but do you know it? How do you know it?

Clearly, some sort of reason or observation is expected here, not another appeal to public opinion. And should it turn out that Jonathan is innocent it will be clear to everyone that Ethan did not know that he was guilty.

We do not want to say, of course, that the opinions of others are never relevant for a knowledge claim. In fact, most of what we know outside the circle of daily life is to a large extent based on the opinions of others. My reasons to believe that Buenos Aires is the capital of Argentina, that the moon is smaller than the earth, or that Gustavus Adolphus died in 1632 are more or less hearsay, all of them. So how do they give me knowledge? The reason is, presumably, that I have good reasons to trust my sources, and that the chain of trustworthy sources leads back to persons that have more direct reasons to believe the same things. Look at Ethan's accusation again. If it is actually true that everybody at school, including for example the teachers, believe that Jonathan took the ball, this usually gives me a reason to believe it too. But not because general consent in itself transforms belief into knowledge, but because I believe that at least some of those persons have good direct reasons to think it was Jonathan – for example that they saw him take it. If everybody involved goes only by the opinions of others, we don't have knowledge but mass psychosis and a lynch mob.

There are other cases as well, where the connection is even more intimate between what people think about something and the way it is. I know that I have 10€ in my wallet, but that piece of paper would not be 10€ unless a lot of people were prepared to accept it as 10€, on suitable occasions. That something is money is a social fact of a special kind, and social facts are partly constituted by what people believe about them. But this is a special feature of certain facts, not a general feature of everything we know.³

As I have already hinted, it may also be argued that the sociological sense of knowledge is in some way parasitical upon the basic sense. Whether something is “known” in the sociological sense, in a certain society, is established by a survey of what people in that society take themselves to know – but not of what they take themselves to know in the sociological sense! The sociological sense, where knowledge does not imply truth, is an outsider's or observer's sense of knowledge: the insider, whose knowledge-claims are investigated, will take if for granted that what he knows is true.

1.4.5 The case of the flat earth

Presumably, there was a period in human history when everyone believed the earth to be flat – perhaps in the Stone Age. Shall we also say that

³We will come back to social facts in Chapter 6.
people in that epoch knew that the earth is flat? Again, experience teaches that people will disagree over this question, some will say yes and some will say no. By now, we should immediately recognize this as a squirrel-type conflict. Everyone agrees to the facts: Stone Age people believed that the earth is flat but it is actually round. So the dispute is about how to use the word “know”, and we have already said enough to resolve it. According to the sociological sense of “know”, Stone Age people did know that the earth is flat, i.e., there was at that time a socially entrenched belief to that effect. And according to the classical sense of “know” they did not know it, because the earth neither is nor was flat, and knowledge (in this sense) implies truth.

As long as we keep to our distinctions and remember what we mean, there is no harm in any of these ways of speaking. The problem arises when we start to mix them up in an argument. Perhaps like this: In the Stone Age people knew that the earth was flat, and knowledge implies truth – so in the Stone Age it was true that the earth was flat. Simsalabim! We have what looks like an argument from the trivial fact that public belief is different at different times and places, to an extremely controversial philosophical position: relativism about truth!

Would anyone really argue this way? Well, it certainly looks like that sometimes, but it is hard to be sure. Even those who are willing to say that it was true in the Stone Age that the earth was flat, usually hesitate to draw the further conclusion that the earth really was flat at that time – just think about the earthquake when it turned into the ball that we now live on! Instead, they will perhaps retire to the position that it was true-for-them, the Stone Agers, that the earth was flat. And maybe that only means that they believed that it was flat – and who would not agree to that?

1.4.6 The pragmatics of “know”

Suppose that the classical conception of knowledge is roughly correct, that this is what “know” normally means. There still remains an interesting question about what we do with the word when we use it in ordinary life, what we have the concept for. It is useful to distinguish two cases, the third person case and the first person case.

In the third person case the emphasis is on conditions (1) and (2). When I say of somebody that she knows that \( p \), it is simply a short and convenient way to say two different things at the same time; that she believes that \( p \) and that \( p \) is true. I tell you what she believes, and I endorse the same belief myself – maybe I also imply that she has good enough reasons for her belief but usually that is not the main point.

In the first person case things must be different. If I just say \( p \), I have already said that \( p \) is true and expressed my belief that \( p \). When I say “The bottle is on the table”, I already assert that it is true that the bottle is on the table, and at the same time I indicate (though I do not assert) that I believe that it is. So what is the additional point of saying “I know that \( p \)” here? Here the emphasis is on condition (3). The point is to stress that I have (good enough) reasons to believe that \( p \), that I am warranted to assert that \( p \). And the point of that is to assure you that you may safely believe \( p \) yourself – it is a way to transfer warrant from me to you. To say “I know
that \( p \) is a bit like making a promise that \( p \) – you can “take my word” for it.\(^4\)

1.4.7 How good must good reasons be?

The third condition of the classical conception says that to know that \( p \) I must have “good reasons” to believe that \( p \). But exactly how good must my reasons be? What reasons are “good enough” for knowledge?

The philosophical tradition, including Plato himself, has been very strict about good enough reasons for knowledge: my reasons must guarantee that \( p \) is true, they must amount to a full proof that \( p \). While perhaps natural in itself, this suggestion is also very problematic. In ordinary life, and in science as well, we usually take ourselves to know a lot of things that we cannot prove to be true. Will the sun rise tomorrow? Is the coffee shop in building A open now? Did Gustavus Adolphus die in 1632? These are clearly questions to which we may claim to know the answers, but can we prove them to be correct?

Plato was prepared to bite the bullet, and admit that there is very little that we know – perhaps there is no knowledge outside mathematics and logic. If you go far enough in this direction you may even become an all out skeptic, and deny that we know anything at all. But most modern philosophers are fallibilists when it comes to knowledge. You may rightly claim to know that \( p \) even though it is possible, given your reasons, that \( p \) is false – but if \( p \) should turn out to be actually false you must, of course, admit that you did not know it in the first place.

So how good must good reasons be, if the demand for proof is too strong? If we go back to how we actually speak, and to what we usually ask of each other, it seems that the demands vary with the situation, with the contexts in which knowledge claims are made and evaluated. Suppose I tell you that there is a bus to the airport at 10.35, and you ask me if I know this. In ordinary circumstances it is presumably enough that I remember myself taking the bus at that time on other occasions, or that I have a reasonably clear memory of consulting the time-table. But suppose it is a question of life and death that you really get to the airport on time today – then I would probably be more cautious, saying that I think so, but maybe it would be best to check it yourself. Again, this points to an analogy between knowledge claims and promises – the more important the promise is, the better the excuse must be for not fulfilling it.

But what about scientific knowledge? Are there specially severe strictures on knowledge in science? Perhaps, but one may also suggest that science makes no claims to knowledge at all. In science we are concerned with reasons for a hypothesis or a theory, but whether these reasons amount to knowledge depends on the context in which the theory is to be used. So maybe the aim of science is not to produce knowledge, after all, but rather to put us, as persons, in a position to know things. Science aims to supply reasons for and against different opinions, but whether these reasons are “good enough” depends on the situation.

---

\(^4\)This aspect of the use of “know” has been stressed by the philosopher J.L. Austin, for example in “Other Minds”, in (Austin; 1961).
CHAPTER 1. SCIENCE, KNOWLEDGE AND TRUTH

This way of looking at things gets some support from some of the formal contexts in which science is appealed to. Think, for example, about the licensing of a new drug for medical use. The actual scientific testing goes through different stages, from the first rough idea through rising levels of certainty that the drug is effective against a certain disease, and that it does not have harmful side-effects that outweigh its benefit. From a purely scientific point of view there is no way to say what degree of certainty that would be good enough to actually start using the treatment on real patients – the acceptable odds on that gamble, so to speak, are decided by law (and sometimes the decision may even be left to the patient). Analogously, scientific experts giving evidence in law courts are not only expected to deliver relevant scientific opinions – they are also supposed to supply the reasons and evidence upon which those opinions are based, and it is for the court to decide whether these reasons are good enough in the case at hand. Or think about the expert panels that are sometimes assembled to decide what the evidence concerning some hypothesis really gives us reason to believe – concerning climate change and the greenhouse effect, for example. Such collective taking stock is not part of the normal scientific process, but is forced upon us by the necessity of taking a practical decision.\(^5\)

1.4.8 Whose beliefs?

There is another, related, problem with applying the classical definition of knowledge directly to science, namely that knowledge implies belief, but whose belief would that be? Science is a social endeavor but belief seems to be a personal thing.

What is it to believe something? A nominal definition of belief is that you believe something when you hold it to be true, but what is it to hold something true? There is a difference between saying that you believe something, and even thinking that you believe something and really believing it. Many of our beliefs are so basic to us that we never think about them. Do you believe that the floor will carry your weight? Do you believe that the water in the tap is drinkable? Do you believe that the driver has seen you and will stop the car? You may not have formulated these convictions even to yourself, but you show that you have them by walking on the floor, drinking from the tap, or stepping out into the street. And, conversely, we sometimes shrink from acting on what we say or think that we believe, showing to ourselves and others that we are really not so sure. Of course I believe that my team will win, but will I bet a month’s pay on it? Is there a better life after death? Yes, sure, but please call the doctor now!\(^6\)

So, there seems to be an intimate connection between belief and action: it is only when I am forced to make a practical decision that the question

---

5Oreskes and Conway (2010) gives a fascinating, and sometimes scary, overview of such work “from tobacco smoke to global warming”. The focus of the book is on how powerful interests have tried, often successfully, to derail the impact of science on public decision making, partly by exploiting the fact that scientific results (like the rest of our knowledge) are never certain and that there is always room for more research.

6Appeals to betting behavior is the idea behind modern theories about how to measure strength of belief. You show how strongly you believe something by what odds you would be willing to accept on a bet that it is true – indeed, “I would bet my life on it!” is a standard way to express subjective certainty.
of what I believe is posed in earnest. But if this is right, only an agent can have beliefs and science is not an agent. Scientists are agents, of course, and they show what they believe by what they do, just like everybody else, but their beliefs do not add up to one big collective set of convictions, which would be what science thinks, the scientific opinion.

Again, it seems that science is more concerned with reasons than with belief. Many of the rules of science have to do with transparency, in a certain sense. To report a piece of research is not only to publish a result, but to give access to the evidence and the reasoning that sustains it. Sometimes this is a handicap, that puts scientists at a disadvantage in relation to other seekers of knowledge. Journalists are required by law to “protect” their sources, i.e., to keep them secret if they do not want to come out in the open, but historians cannot use such information, they must make sure that other historians can follow their trail and check their evidence.

As we will see, an influential modern view about what sets science apart from other human activities is the specific nature of “scientific communities”, the more or less formalized groups of people, at different levels, within which research is typically carried out. We will come back several times, and in much more detail, to the social aspects of science, but as a first stab we may note that a scientific community is much more similar to a debating society than to a political party or a church. There must be a lot of background consensus, of course, as a basis for discussion, and in particular concerning the rules of the game, but there are no procedures for reaching agreement or establishing a community dogma on the issues at hand – just the ongoing debate and the gradual sedimentation of some things into the background when there seems to be no room for reasonable discussion anymore.

1.4.9 Epistemic injustice

The British philosopher Miranda Fricker (2007) has drawn attention to a phenomenon that she calls “epistemic injustice” and which is, in one of its aspects, connected to the belief clause in the classical definition of knowledge. It is a commonplace that knowledge is unevenly distributed in society. Access to education and other sources of information is tied to factors like social class, economic prosperity, ethnic background and gender. This sort of inequality has to do with access to information, but the injustice that Fricker describes belongs to another psycho-social dimension.

To possess knowledge, in the classical sense, it is not enough that you have sufficient abilities of observation and reasoning, you must also have the power to believe what your experience tells you – you must have the right degree of intellectual self-confidence. Clearly, one can be deficit in self-confidence in two directions: by having too much or too little of it. How does that come about?

One pertinent mechanism to skew one’s intellectual self-confidence is what Fricker calls “testimonial injustice”. Knowledge is a social endeavour, where we are constantly engaged in a give and take of information, and my

\(^7\)Thomas Kuhn (1970, first edition 1960) was perhaps the first to explicitly appeal to the social structure of scientific groups as a criterion of demarcation for science. Another shading of basically the same idea has been argued by Pierre Bourdieu, for example in (2001).
CHAPTER 1. SCIENCE, KNOWLEDGE AND TRUTH

21

testimony, in the broad sense envisaged by Fricker, is my contribution to
the epistemic interchange. Negative testimonial injustice is to not receive
one's due of trust and respect for one's opinions, to have one's say doubted
or neglected for reasons unconnected to one's epistemic competence – for
example because of one's gender, social class, ethnic origin or age. This is
clearly detrimental to the community, which is deprived of relevant sources
of potentially important knowledge, but it is also harmful to the individual
affected by it: if you are used to not being trusted by others, it is only too
natural that you will not trust yourself, and thus become ignorant of things
that you are really in a position to know.

Epistemically speaking, the reverse condition – to have one's opinions
respected for more than they are worth – is also harmful, both for the com-
munity and the individual, as it will lead us to believe that we know things
that we do not. But the overall effect for the "victim" of positive testimo-
nial injustice is more likely to be beneficial than harmful – the status as a
knower brings many other social benefits in its wake – while the victim of
negative testimonial injustice is likely to suffer also in other respects.8

1.5 What is truth?

Thinking about one concept always leads to others, that are often prob-
lematic in their own ways. We have analyzed the concept of knowledge in
terms of belief, reasons, and truth, and each of these concepts have occu-
pied philosophers ever since the Greeks. I will not say more about what it
is to believe something, and scientific reasoning will be the topic of the next
chapter, but here I will add a few things about truth.

One problem with truth is that the notion is so fundamental that it is
hard to see how to define it in simpler terms. We all use it as a matter of
course in daily life. We try to speak the truth most of the time; when we
sometimes lie we believe that what we say is false; we check on doubtful
assertions to see if they are true, and when someone says something that
we think is false we immediately protest. But when asked to explain in
general terms what truth is, everything we say seems to be either trivial or
wrong. The most famous definition of truth is probably Aristotle's:

To say of what is that it is, or of what is not that it is not, is true.
To say of what is not that it is, or of what is that it is not, is false.

I speak the truth when I tell it like it is. What I say is false when I tell
it like it is not. It is hard to disagree, and the standard objection against
Aristotle is not that his definition is wrong, but that it is uninformative.
To this allegation one may react in three different ways. The first way is
to deny the allegation, and try to show that there is in fact a substantial
theory of truth hidden in formulations like Aristotle's. This theory has a
name, the Correspondence theory of truth and I will say a little about it in
the next section. The second way is to admit the allegation, but to argue
that the triviality of the formulation is all right, because the concept itself is

8In chapter 5, we will come back to another form of epistemic injustice that Fricker de-
scribes: “hermeneutic injustice".
trivial. This leads to the Minimalist or Redundancy theory of truth, which was first suggested by the British philosopher Frank Ramsey in 1927 (Ramsey; 1990, pp. 38–39), and which has become increasingly popular among philosophers in recent years. I shall argue that regardless of its ultimate merit as a philosophical theory, this is in fact a very useful way to look at truth. The third way, of course, is to admit the allegation and try to supply some different and more informative theory of truth – the most well known variants of this strategy are Coherence theories, Pragmatic theories and Epistemological theories of truth.¹

1.5.1 Truth as correspondence

The basic idea behind the correspondence theory is that truth is a relation – of similarity or correspondence – between two things. “Veritas est adaequatio intellectus et rei”, says Thomas Aquinas in a famous phrase: “truth is the equation of intellect and thing”. Although this is a very natural idea, it has proved difficult to pin it down more exactly. What are, for example, the terms of this relation? What are the truth-bearers, the things that are true or false, and what are the truth-makers, the things that make the truth-bearers true and false?

In Aquinas’ formulation, the truth-bearers seem to be mental – thoughts or judgements or something like that. In other versions of the correspondence theory the truth-bearers are linguistic – sentences or assertions – and there are still other possibilities. For simplicity, I will talk as if the truth-bearers are sentences, though this cannot be exactly correct – to be true and false a sentence must at least have a certain meaning. And what about the truth-makers? Aquinas talks about “things”, and sometimes of “reality”. But it seems that things correspond to nouns rather than to sentences, and the more usual terminology is that true sentences correspond to “facts”. For simple examples, the notion of a fact seems rather tangible. The sentence “There is water in the bottle” is true, if and only if there is a corresponding fact that there is water in the bottle. And even if this fact is neither identical to the bottle nor to the water in it, it still seems to be there in some sense, at roughly the same place as the bottle. But what about “The bottle is not empty” – does it correspond to a special negative fact, or to the same positive fact that there is water in the bottle? What about sentences like “Some bottles are empty”, or “All ravens are black” – are there existential and general facts that correspond to them?

There are philosophers who have gone deeply into such issues, and tried to provide informative versions of the correspondence theory – most famously the young Ludwig Wittgenstein (1921), who developed a theory according to which true sentences are “pictures” of the corresponding facts. But another reaction is perhaps more common: the notion of a fact does not tell us anything about truth, because the only thing we know about facts is that they correspond to true sentences.

¹ Schmitt (1995) gives an accessible and useful overview of the most discussed theories of truth, and their problems.
1.5.2 A minimal theory of truth

A minimal requirement of a theory of truth seems to be that equations of the following sort should always be true:

“Snow is white” is true if and only if snow is white.

This seems to be in line with Aristotle’s definition. Now, suppose that this is all there is to the notion of truth? To say that it is true that snow is white is just a more roundabout way to say that snow is white. To say that $p$ is true neither adds nor detracts anything from simply saying that $p$.

This is what would be called a “minimal” or “redundancy” or “deflationary” theory of truth, and, as I said, it has attracted a lot of interest from philosophers in recent years.\(^\text{10}\)

But if this is correct, why do we have the concept of truth? What purpose does it serve? Well, sometimes we need to endorse (or deny) sentences in clusters, so to speak, and the notion of truth provides us with an convenient way to do so without going into all the cumbersome details of actually enumerating the relevant sentences. Suppose, for example, that I take Al Gore to be a reliable, but not infallible, source with regard to the issue of global warming. I might express that agreement as

Most of what Al Gore says about the climate is true.

In principle, I could say the same thing by (1) finding and enumerating everything that Al Gore has said and will say about the climate, and (2) forming all the possible conjunctions of these statements that include more than half of them, and (3) asserting the disjunction of all of these conjunctions. If the word true allows me to express my general agreement with Al Gore without actually going through steps (1)–(3), it seems to be a very useful word, indeed, even if it lacks all metaphysical import!

We can also note, in passing, that the notion of truth is not really necessary to state the classical conception of knowledge, either. We gave the second condition of the classical definition as “$p$ is true”, but this definition would have been equally good:

A knows that $p$, if and only if

1. A believes that $p$
2. $p$
3. A has good reasons for $p$.

1.5.3 Truth is not the problem

For us, the most important feature of the minimal theory of truth is that it helps against a natural tendency to over-generalize problems of meaning. Most arguments that there is something problematic about truth, or that truth is “relative” to this or that, start from specific problem cases, and then proceed to question the general notion of truth. The minimal theory blocks

\(^{10}\)When it comes to the details, there are several different versions of “deflationism”, but the differences between them need not concern us here.
that second step and helps us to keep the problems local, and to solve them one by one where they belong. Start with an easy case:

Fredrik Reinfeldt is bald.

Is that true or false? Not easy to say, of course, but the problematic word is not “true” but “bald”. Bald is a vague notion, and the defining feature of vague notions is that between the clearly positive and the clearly negative instances there is a fuzzy zone of problem cases. But it is still the case that

“Fredrik Reinfeldt is bald” is true if and only if Fredrik Reinfeldt is bald

and this is all there is to the notion of truth in this as in other cases. Is it true that Russia is a democracy? Hard to say, but the problem is not with “true” but with “democracy”. Is it true that it is morally wrong to eat meat? Difficult question, but the difficulty is not with “true” but with “morally wrong”.

1.5.4 Truth by convention?

Perhaps the most common form of relativism about truth is conventionalism, roughly, the idea that we decide which sentences are true by deciding what our words mean. Suppose that the bottle is on the table. Is this not just another way to say that the sentence “The bottle is on the table” is true? And is it not obvious that if the expressions “the bottle” and “is on” and “the table” had meant something else the same sentence might have been false? And as the meaning of words is clearly a matter of convention – as is proved by the existence of different languages – is it not obvious that the truth of sentences is conventional in the same way?

As a philosophical argument this belongs with the flat-earth argument that we discussed earlier: we start with something obvious and seem to arrive at an interesting, even sensational, conclusion, but only by ignoring the crucial step. When we learn a language, we master a grammar and lexicon and through them we understand what the sentences of the language say, but the rules of language do not tell us which sentences are true. Knowing English we understand “The bottle is on the table” as well as “The bottle is not on the table”, but to know where the bottle is we still have to look at it.

Suppose that we decide on a precise criterion for being bald, a certain number of hairs per cm$^2$ perhaps, and that Fredrik Reinfeldt is bald according to that convention. Does this imply that he is now bald by convention? Certainly not. The criterion fixes the meaning of “bald”, but to know if Fredrik Reinfeldt actually is bald, according to the criterion, we still have to count his hair.

There are some sentences, however, that really are true by convention, in the sense that if you know what they mean, you also know that they are true. The standard example is

All bachelors are unmarried.
To be a bachelor simply is to be an unmarried man, that is how we use the words and that is what the dictionary tells us. In the same way we need not perform elaborate investigations to find out that triangles have three angles, that socks are worn on the feet, or that liquids flow. But the price that such sentences pay for being so easily verified is that they carry no information about the world – the reason we don’t have to look at the world to check if they are true, is that they are true regardless of how the world is. Immanuel Kant called such sentences “analytical” because in uttering them we don’t add anything to what we believe, we just bring out, or “analyze”, what is already implicit in the concepts we use.

As before, we have to be a little careful with the word “sentence”, here. A string of words can be used to mean anything at all, and what is analytical or not, as the case may be, is not the string of words by themselves but the meaningful sentence, the words as meant in a certain way. The word bachelor can be used to mean other things than “unmarried man” and for some of these uses it will not be true that all bachelors are unmarried. And if you define “bald” as “having as much hair as Fredrik Reinfeldt or less” it will be true by convention that Fredrik Reinfeldt is bald – but then the statement that he is tells you absolutely nothing about his hair.

1.5.5 Useful belief

Why do we want our beliefs to be true? One reason, surely, is that true beliefs are often useful, and false beliefs are often harmful or dangerous. If I believe that this mushroom is poisonous I will refrain from eating it – which is a good thing if it actually is poisonous, but a bad thing if it is not, and I am in the mood for a mushroom sandwich. Some philosophers have suggested that the connection between truth and what it is useful to believe is so strong that it can be used to define what it is to be true, giving us the pragmatic theory of truth, which goes something like this:

“p” is true if and only if it is useful to believe that p.\textsuperscript{11}

A little reflection, however, shows that this is a rather problematic idea – useful perhaps, in some contexts, but hardly true. To begin with, there are counterexamples of all kinds: not only are there useless truths and innocent falsities, there are also useful falsities and harmful truths. It might, for example, still prove to be the case that the discoveries of nuclear physics will lead to the destruction of mankind, but that would hardly prove that the theory is false. On another note, there is a lot of research indicating that it is beneficial to have a slightly false self-image – believing oneself to be more competent and generally better than one actually is leads to greater success and a happier life, while an accurate estimation of one’s own abilities correlates with being slightly depressed. Historical knowledge

\textsuperscript{11}The pragmatic theory of truth is associated with (some versions of) American pragmatism, and in particular with William James. Another philosopher who sometimes wavers in this direction is Nietzsche, although his basic position is rather some form of correspondence theory, which prompts him towards epistemological nihilism or skepticism. One may venture the hypothesis that pragmatic theories of truth is an offshoot of Darwinism – the evolutionary process should be able to explain the survival of beliefs just like it explains other features of living things, by reference to their survival value.
is also a rich source of problem cases for the pragmatic theory. Even given that it is generally useful to have some accurate beliefs about the past, it is still difficult to point to the specific usefulness of many particular historical beliefs, but there is no corresponding problem of understanding what it would mean for them to be true.

Usefulness also seems to be relative to persons in a way that truth is not – to be useful is always to be useful for someone, but what is useful for me to believe may not be useful for you. It is probably useful, in most historical circumstances, to share the religious and moral beliefs of the society one lives in, but that is hardly an argument in favor of any particular religion.

Such examples point to an important ambiguity in saying that X is a reason to believe that p. Usually this just means that X is pointing to the truth of p, X is a good reason for p, but it may also mean that X is a good ground for wishing to believe that p. It is not unusual, for example, to argue for or against religious beliefs by pointing to the advantageous or harmful effects that would come from people having those beliefs, and one may even make comparative arguments that the consequences of believing in one religion is better than believing in another. Whatever importance one may attach to such considerations, it is clear that they are completely irrelevant to the truth of the relevant beliefs. Here, we will only be concerned with reasons to believe in the first sense, i.e., reasons for the truth of what is believed.

1.5.6 Unknown truth

We have already seen that there is an important connection between truth and knowledge: if something is known it is also true. In the other direction, however, there is no obvious connection between the two notions: something may well be true without anyone knowing about it. Indeed, the overwhelming majority of all truths are fortunately unknown – the tedium of someone having to know each trivial fact about everything hardly bears thinking about.

It is not uncommon, however, to conflate truth and knowledge. Consider, for example, the often heard declaration that “there is no absolute truth”. Usually, this simply expresses the opinion that there is no absolutely certain knowledge – whatever we take ourselves to know, there is always some possibility, however remote, that we may be wrong. But the fact that something is not known does not make it any less true, if it is true. There is, for example, a long standing controversy about who actually wrote the plays that we traditionally attribute to William Shakespeare. The majority opinion, among those who know such things best, is that they were actually written by a historical figure of that very name, but there are also dissenting voices – some who put their money on other candidates, like Francis Bacon, and some who would contend that we simply do not know. But this does not preclude that there is a fact of the matter – either William Shakespeare wrote the plays or he did not, and if it is true that he did, it is also “absolutely” true. If there were no fact of the matter, what would the fuss be about?\(^\text{12}\)

\(^{12}\)This is not to deny, of course, that the facts may be complicated in different ways. Perhaps Shakespeare wrote some of the plays and somebody else wrote the rest, or perhaps they were
CHAPTER 1. SCIENCE, KNOWLEDGE AND TRUTH

But even if truth is independent of actual knowledge, perhaps there is a more subtle connection between truth and possible knowledge? Can there be truths that are not only unknown, but also unknowable, i.e., impossible to know? Or is it is, at least in principle, possible to know anything that happens to be true?

The answer to that question hangs largely on what “in principle” means, of course. There are clearly a lot of things that we would like to know that are beyond our present means to knowledge. Think about the past, for example. In 1697 the Royal Palace in Stockholm was burned to the ground. Among the things destroyed were most of the Swedish national archives – had they been preserved we would undoubtedly have been in a position to know a lot of things about medieval Swedish history that we are now doomed to be ignorant about. Clearly, though, it was only our knowledge of Swedish history that was touched by the flames, and there was no corresponding destruction of past events.

But maybe we should say that the loss of the documents only makes the relevant historical facts unknowable in practice, and not in principle? After all, it would have been possible for the fire to have been prevented and, even as things stand, everything chronicled in the documents was presumably once known to somebody, and what someone has already known can hardly be unknowable in principle, can it?

And even if one may, perhaps, dispute that irretrievably lost knowledge should still count as knowable, it seems unreasonable to deny that things are knowable that are just beyond our present capacity to know but will be knowable in the future. The last few decades have seen enormous strides in the kinds of facts that we take as knowable about the past, thanks to progress in genetics and chemical analysis. Today, we take it as a known fact that there was interbreeding between Neanderthals and modern humans, and we have a good idea about the scale on which such events took place. We expect science to tell us, through isotopic analysis, where in the world the tin was mined that was used in a specific bronze dagger, excavated thousands of mile from its point of origin, and thereby to inform us about prehistoric patterns of trade and migration. A hundred years ago such facts were not merely unknown – one also had no idea about how they could ever become known and what kinds of evidence would be relevant to test speculations about them. There is no reason to think that the future will not bring many more happy surprises of the same kind, advancing not only our knowledge but also our capacity to know.

Despite the difficulties of saying once and for all what is or will be knowable, even in principle, some scientifically minded philosophers have thought it feasible and informative to define truth in terms of possible knowledge. Perhaps they have been inspired by an argument along the following lines. We all know that we sometimes make mistakes, that some things that we think we know are not really true, and so are not really knowledge either. How do we find that out? By the advancement of knowledge, presumably, and in particular by the advancement of scientific knowledge. But doesn’t this mean that the tangible core of the difference between the result of some sort of collaboration, so that there is no single author. But if such be the case, it still is the case – as “absolutely true” as anything, regardless of the state of our knowledge.
true knowledge and mistaken belief is just the experience of learning to know better? But if this is so, maybe we should define truth as what would be believed “at the end of inquiry”, when the process of revision and further investigation is completed?

We shall not go into the deeper philosophical motivations that might prompt someone in this direction – most of the time connected to a natural suspicion of the notion of an unknowable Ding an sich. But it is worth noting that such an attempt to substitute the hard realities of scientific progress for the lofty notion of truth involves some rather optimistic and intangible assumptions of its own. What is “the end of inquiry” and when is it supposed to take place? Clearly, it cannot be a real future moment. First, the whole idea of a completed knowledge is a bit absurd – sure, inquiry will presumably come to an end at some time, but more probably because of what befalls mankind than because of a lack of open questions to pursue. And along the way there is surely no guarantee that knowledge will not be lost as well as gained. The answer to such objections would presumably be that it is not the actual course of future science and human knowledge that is relevant, but the ideal direction that science would take if it were correctly pursued far enough into the future, i.e., forever. But then we face the difficulty of indicating that ideal direction without falling back on the notion of truth, or of a fixed “scientific method” that we would be at a loss to specify.

1.6 The problem of demarcation

So far we have been talking around the notion of science, discussing concepts that are important for describing the processes and products of research, but which themselves have much wider fields of application. Whatever the precise relation may be between science, knowledge and truth, it is obvious that science does not have a monopoly on truth and knowledge. So what is specific about science? What is it that demarcates science from other fields of human endeavor?

The problem of demarcation looms large in the philosophy of Karl Popper, who has pointed to it as the most important problem in the philosophy of science (1968, p 34), and we will soon come back to his solution to it. But first we need to say a few words about a preliminary question: what is it that we want to demarcate science from?

Popper wanted to characterize empirical science by distinguishing it from certain other forms of thinking that might easily be confused with it. Among these were the a priori sciences, mathematics and logic, but above all he was interested in the demarcation of science from what he saw as its evil twins: unscientific systems of thinking that disguise themselves as science, in order to avail themselves of its authority and prestige. Recurrent examples of such evil twins, in Popper's writings, are astrology, psychoanalysis and Marxism – no doubt he would have been ready to add, for example, creationism and scientology to the list. Sometimes he lumps these other ways of thinking together under the label of “metaphysics” –

13The whole idea originates in post-kantian German philosophy, with Hegel in particular, but most modern versions of it go back to C. S. Peirce.
he shared the habit of dividing human thinking into the two grand classes of science and metaphysics with his contemporaries and compatriots in the Vienna Circle, who presumably got it from Wittgenstein’s *Tractatus*.

It is rather obvious that any attempt to define science through such a simple contrast would yield a very revisionary conception of science – my knowledge that my children are asleep, while I’m writing this in the early morning, is hardly scientific in any normal sense but neither is it a piece of metaphysics. There are in fact many different problems of demarcation, depending on different contrasts that we may be interested in, and we will discuss some of them under different headings in this section. We will start with Popper’s contrast between science and pseudo-science – the impostors that try to pass themselves off as science. Next, we will look at the relation between a full-blown science and what might be called proto-science, the various forms of pre-scientific thinking from which a scientific discipline develops. Third, we will discuss the relation between science and common sense, the everyday forms of knowledge and investigation that science builds upon and refines. Fourth, we will touch upon the interesting fact that many scientific disciplines have specific non-science neighbors, with which they interact and have particularly intimate relations, on different levels, but which are themselves not part of science. The natural sciences often have technological neighbors, the aesthetic disciplines have a not very carefully patrolled frontier towards aesthetic criticism, history and journalism often share subject matter and treat it in related ways, some social sciences cultivate close relations with specific fields of politics and administration, etc.

It is important to keep the multiplicity of different demarcations in mind, in order to avoid a common fallacy. Concentrating on just one division, it is all too easy to take what is only a distinguishing mark of science in relation to something X as a full definition of science, and even as an essential definition, stating what science “really is”. The reality is, of course, that science is a complex phenomenon and that any simple definition of it only works in a specific context, where a host of other important features are already taken for granted.

1.6.1 Science and its twins

The Danish astronomer Tycho Brahe (1546–1601) is celebrated for his contributions to the history of astronomy, but in his own days he was just as renowned as an astrologer, deployed by kings to give advice on good or bad days for important undertakings. At the time, one saw no conflict between these different functions – astrology was seen as no less scientific than astronomy, but rather as a useful application of it. Nowadays, astronomy is still a prestigious science while astrology is relegated to the waste bin of superstition and pseudo-science. But what is the essential difference between them?

It is certainly not in the degree of theoretical elaboration or intrinsic difficulty – learning to cast a horoscope in a proper way is not an easy thing. Nor is the problem for astrology that it is false; most of Tycho Brahe’s properly astronomical ideas are today taken to be false as well, and future astronomy will presumably pass the same judgement on many of our own
opinions, but that does not make them unscientific. Popper’s basic idea is that the essential difference between astronomy and astrology lies in the different attitudes taken to them by their practitioners. When an astronomical prediction is shown to be false, astronomers take this as a reason to modify or even to abandon the theories and models on which the prediction is based. When astrological predictions are not fulfilled, the business of astrology is carried on as before. This is why astronomy has changed beyond the wildest measure since the 16th century, while astrology is just the same as it ever was.

The criterion of demarcation for empirical science, according to Popper is that a scientific hypothesis is falsifiable by empirical evidence. It does not have to be actually false, of course, to be scientific – but if it should be false there must be some possible observation that would count as showing that it is.

Popper often comes back to a specific example that impressed him as a young man, when the theory of relativity was new and still controversial. One prediction of that theory, that sets it apart from classical (Newtonian) physics, is that light that passes near a great mass, like that of the sun, will be measurably bent by gravitation. This implies that the apparent position of distant stars, on the opposite side of the sun from us, will come out differently as calculated by the two theories. A solar eclipse in 1919 provided an opportunity to test this prediction, and Einstein’s theory was vindicated in spectacular fashion.

What impressed Popper was not that the theory was shown to be true by the evidence – after all it was just one prediction that came true and it might still be proved wrong on countless other points – but the thought that if the result had gone the other way, the theory would have been regarded as falsified, and consequently abandoned. The demand for falsifiability has two aspects, one of which is logical, in a broad sense, while the other is psychological. The logical demand is that the theory and the prediction shall be formulated in such a way that they can be contradicted by an observation report – we will come back to the details of this in the next chapter. The psychological demand is that researchers shall take a critical attitude towards their own ideas, and be prepared to abandon them in the face of negative evidence. Astrology fails on both counts – its predictions are usually vague and imprecise, so as to make it difficult to tell whether they are fulfilled or not, and its proponents take a dogmatic attitude towards it, being strongly inclined to explain away or ignore anything that tells against their theory, and instead clutching eagerly to what might count in its favor.

As we will see in the next chapter, the logical demand for falsifiability is itself difficult to formulate in a precise way. There are always several options to consider in the face of recalcitrant evidence, and to abandon the hypothesis under test is not always the most reasonable one – it has even been argued that, from a purely logical point of view, it is always possible to cling to any hypothesis at all, if one is prepared to adjust one’s other opinions accordingly (Quine; 1953, p 43). Popper is fully aware of the difficulty of making falsifiability into a purely formal requirement, and it is precisely this that makes the psychological requirement of a critical attitude

important. Indeed, his criticism of psychoanalysis and marxist historiography is not concerned with the content of the respective theories, but with the dogmatic attitude of their followers – he thinks that they have devised ways to stave off all criticism beforehand, by explaining critical comments from within the theory, as symptoms of repressed desires or false class-consciousness, rather than to take them seriously on their own terms.

The focus on falsification may seem odd at first. After all, science aims for truth, not falsity. Popper agrees to that, of course, but he thinks that looking for falsifications is the best way to find interesting truth. There are two reasons for this. The first is that the easiest way to be right is to say as little as possible, beyond what one already knows – vagueness and imprecision are good ways to protect oneself from being wrong. To say with mathematical precision where one expects a specific star to be seen at a given time and date, is to take a huge risk of being proved wrong, but the reward is that one's theory, if correct, will provide a lot of information about the world.

The second reason is that finding confirming evidence, for any given theory, is always too easy. Think about the latest cure X for the common cold, that you find among the herbal drugs in your local store. You first learn about it by being told by friends and relatives how it cured them. They had a cold, it did not go away, they took X and got well again in no time! But such anecdotes give little reason to believe in the effectiveness of X – getting well again from a cold is just what you would expect in any case. Real confirmation must come from a controlled setup where X is correlated with surprising outbursts of good health, beyond what one would otherwise expect. A telling test is done from the standpoint of a skeptic, not from that of a believer. “If it's so good, can it do this?”

Popper’s philosophy is very popular among working scientists, a fact which in itself seems to speak in its favor. After all, a scientist should know a thing or two about what science is! But part of its appeal within the scientific community surely depends on the flattering picture it gives of science. Who does not want to be a critical spirit, ever prepared to set aside one’s own prejudices in relentless pursuit of the truth, always on the guard against wishful thinking and undue bias in favor of one’s own ideas? Before we bask in the pleasant glow of that reflection, however, we should remember the distinction between the descriptive and the normative. Popper is well aware that science does not always work in the way he says it should work, and that scientists are only human, like the rest of us. He tells us what we ought to do, in research, not what we always actually do.

1.6.2 Science and its ancestors

In some ways, Popper’s view of science, with his emphasis on the logical, the psychological and the normative, is typical of a classical approach to the subject, which we may classify as individualistic, in a broad sense. There is not much in Popper’s definition of science that would distinguish it from a description of knowledge in general – science as a collective endeavor proceeds in just the same way as the problem-solving activity of a single person, there are just more people involved. Starting with Thomas Kuhn’s book The Structure of Scientific Revolutions (1970, first edition 1960) this
classical approach has been challenged by accounts of science that are much more descriptive, starting from detailed historical or sociological studies of how science is actually done. This shift from the normative towards the descriptive has been accompanied by a relocation of interest from psychological to sociological factors in the process of research, and from logical towards rhetorical descriptions of what science produces – from structures of argument to strategies of persuasion. Different aspects of this general contrast will occupy us throughout the book, and in particular in chapters 5 and 6, but for the moment we will just concentrate on its possible implications for the problem of demarcation.

Kuhn has contrasted his own solution to the problem of demarcation with Popper’s, but it is worthwhile to note that it is not the very same demarcation that they are primarily interested in. When Kuhn compares an un-science, like astrology, with science, he is often more concerned with the similarities than with the differences, and the focus of his own interest is what happens when a field of learning first becomes a “mature science”, as happens, for example, with physics after Newton. His main contrast is between science and proto-science, rather than between science and pseudo-science.

The important step on the road to maturity, according to Kuhn, is the formation of a “scientific community”, held together by common allegiance to what he calls a “paradigm”. The notion of a paradigm is complex – much more complex in Kuhn’s own work than it has since become in ordinary parlance. A paradigm contains elements both of a theory and of a methodology. The theory delineates the subject matter that the community investigates, saying what sort of phenomenon it is and expressing the most important principles concerning it. The methodology specifies what sorts of problem the community takes it as their task to solve, as well as the methods that they take as permitted and relevant to those solution – ranging from the most general scientific values to community-specific tricks of the trade. An important part of Kuhn’s conception is that many components of the paradigm are not given as explicit rules or principles, but are implicit in concrete “exemplars” of scientific work, which play an important part in a scientific education. Working with such examples instills the paradigm as tacit knowledge in the trainee – as a set of skills, habits and thought patterns that will inform her own work, in ways largely beyond her conscious choice and control.

Kuhn has contrasted his own solution to the problem of demarcation with Popper’s, but it is worthwhile to note that it is not the very same demarcation that they are primarily interested in. When Kuhn compares an un-science, like astrology, with science, he is often more concerned with the similarities than with the differences, and the focus of his own interest is what happens when a field of learning first becomes a “mature science”, as happens, for example, with physics after Newton. His main contrast is between science and proto-science, rather than between science and pseudo-science.

The important step on the road to maturity, according to Kuhn, is the formation of a “scientific community”, held together by common allegiance to what he calls a “paradigm”. The notion of a paradigm is complex – much more complex in Kuhn’s own work than it has since become in ordinary parlance. A paradigm contains elements both of a theory and of a methodology. The theory delineates the subject matter that the community investigates, saying what sort of phenomenon it is and expressing the most important principles concerning it. The methodology specifies what sorts of problem the community takes it as their task to solve, as well as the methods that they take as permitted and relevant to those solution – ranging from the most general scientific values to community-specific tricks of the trade. An important part of Kuhn’s conception is that many components of the paradigm are not given as explicit rules or principles, but are implicit in concrete “exemplars” of scientific work, which play an important part in a scientific education.

Working with such examples instills the paradigm as tacit knowledge in the trainee – as a set of skills, habits and thought patterns that will inform her own work, in ways largely beyond her conscious choice and control.

We will come back to the notion of a paradigm in much greater depth.
in chapter 5, but for now we are just interested in its connection with the
notion of a scientific community, and in Kuhn's use of it for distinguishing a
“mature” science. While Popper thinks that a critical attitude to even one's
most basic convictions is what demarcates science from other human prac-
tices, Kuhn points to an element of dogmatism which he thinks is a sine qua non for science. A paradigm is not a theory or a hypothesis in the usual
sense, provisionally adopted for the purpose of testing and critical evalua-
tion, but the very framework for work in a particular discipline. It is only
within a paradigm that the specific problems are defined which scientists
try to solve, and it is only the paradigm that provides the tools for solving
them and the standards to evaluate the solutions.

According to Kuhn, it is the acquisition of a paradigm that first trans-
forms a discipline into a mature science. At the pre-paradigmatic stage,
time and energy is consumed on the discussion of fundamental issues,
partly of a philosophical nature, but once the paradigm is secure, all efforts
are directed towards concrete “puzzles” that the paradigm indicates to be
solvable and rewarding. This is what Kuhn calls “normal science”, a stage
where steady development is expected, and usually delivered, along the
lines of research outlined by the paradigm. Think, for example, of the cur-
cent stage of genetics, with its large scale projects of mapping the genome of
different species – there would be no question of undertaking such projects
unless one was reasonably sure of success, and confident of the correctness
of the underlying theory.

As I said, a scientific paradigm is carried by a community with a specific
structure, where the dominating feature is a certain sort of self-sufficiency.
There are specific criteria for entrance into the community, and novices go
through a prolonged period of difficult training, aimed at instilling the spe-
cific habitus associated with the paradigm. The community regards itself
as the sole authority on questions covered by the paradigm, and only inter-
ally defined values are allowed to influence the choice of problems and the
evaluation of solutions. Evaluations within the group, both of work done
and of the personal merit of practitioners, are done by “peer review”, and
towards the outside world the scientific community cultivates a resolute
immunity to non-expert opinion.

Normal science is, precisely, the normal state of a mature science, but
sometimes, according to Kuhn, a scientific group will be shaken by a crisis,
where confidence in the paradigm wavers, usually on account of the suc-
cessive accumulation of apparently unsolvable problems – “anomalies” that
should be explainable within the paradigm, but do not yield to even the best
efforts by the most elite practitioners. In such circumstances, questions re-
garding basic assumptions may be opened again, and discussion reverts to a

19The term 'habitus' is from Pierre Bourdieu, cf. for example (Bourdieu; 2003). Bourdieu's
work on science is similar to Kuhn's in many respects, but also supplements it in important
ways, that we will come back to below.

20I noted above that Popper gives a flattering view of science, that has made it popular
among scientists. Kuhn's views have certainly had less such appeal, at least within the sci-
cences that he himself classifies as mature. But the notion of the scientific community as
essentially self-regulatory and not to be meddled with from the outside also has a pro-science
ideological potential, in relation to attempts at political regulation of science. Steve Fuller
(2003) has forcefully argued that this ideological function of Kuhn's theory was a desideratum
that it was actually designed to fulfill.
more philosophical form, reminiscent of the pre-paradigmatic stage. Sometimes the threat to the paradigm will be fended off, but sometimes the crisis will issue in a scientific revolution, where a new paradigm replaces the old one before everything reverts to normal. Again, sociological factors play a large part in ensuring the successful resolution of a crisis. Normal science is the foundation of a scientific profession, with career opportunities and a way of life put in peril by a crisis, and the pressure to revert to normality is huge.

It is somewhat paradoxical that Kuhn’s ideas have found little resonance within the natural sciences, which are his own main examples of mature sciences, but have instead been a major influence on the self-image of the cultural sciences, to which Kuhn himself does not think they are applicable. Indeed, he has (1993) pointed to the notion of normal science not only as demarcating a mature science from its pre-history, but also as distinguishing the natural sciences from the cultural sciences. The latter are, according to Kuhn, still in the pre-paradigmatic state, where fundamentally different conceptions of the subject matter and how it should be studied are competing with each other, so that no single approach can confidently get going with the real work of normal science.

Obviously, Kuhn’s work tells us a lot of important things about science, and it also raises a lot of interesting questions that we will come back to in later chapters. But does it work as a criterion of demarcation for science? Again, we must remember which demarcation we are interested in. If we focus on the difference between science and other knowledge seeking enterprises, the sociological factors that Kuhn points to are undeniably important. But just as clearly, it does not solve Popper’s problem about the demarcation of science from pseudo-science. After all, the formation of closed groups rallied around sets of shared commitments, with almost complete disregard for the opinions of outsiders, is not a unique phenomenon that defines science. Seen in that perspective, the challenge is rather to find the peculiar features of scientific communities that explains their specific forms of development and success – why they do not more often degenerate into petrified dogmatism, and why they yield such useful products.

1.6.3 Aspects and primacy

I pointed above to the existence of different general approaches to science, distinguishing between logical, rhetorical, psychological and sociological approaches. Much of the contemporary polemics regarding the nature of science is cast in terms of such broad contrasts, and this is in itself a puzzling fact that may be worthy of a small detour. Clearly, there is room and need for all types of factors in a theory of science: there must be both a psychology, a sociology, a logic and a rhetoric of research. So what is the quarrel about?

It is about primacy, of course, about what factor is the decisive one, that explains or motivates the others. Take the conflict between logic and rhetoric as an example. Classical approaches will put logic in the driver’s seat. The proper rhetoric of scientific writing would simply be dictated by the need to write and speak so as to make the logical structure of arguments clear. Post-modern approaches, to the contrary, sees the effect of
CHAPTER 1. SCIENCE, KNOWLEDGE AND TRUTH

persuasion as the central thing, and views the emphasis on argument as just one rhetorical device among others.

How would one argue about such things? That depends on the context, and in particular on whether one’s own focus is primarily normative or descriptive. From a normative perspective, it is natural to judge scientific practices on the basis of their function, what we have science for. And the default idea, surely, is that the main function of science is to aid us in finding knowledge, reducing ignorance and avoiding error – it is in the business of finding things out, so to speak. By itself, this seems to point to a certain primacy for the logical factor, as being directly concerned with the evaluation of opinions in terms of their truth or falsity, and for a derived status for the other factors – being subject to evaluation in terms of their contribution towards the goal of increasing knowledge and reducing error and ignorance.\(^{21}\)

There are many questions, of course, that may be raised in relation to this rather commonsensical idea. On the deepest level, perhaps, we have philosophical questions concerning truth, reason and interest. Radical versions of sociologism would question whether we have strong enough conceptions of such things for them to serve as yardsticks, independent of the practices they are supposed to evaluate. If all the criteria that we can appeal to vary between different paradigms or the like – how can they be used to adjudicate between them? In practice, I think, these worries are often very exaggerated, but on the theoretical level we will come back to them in chapter 5.

Another set of questions would concentrate on the idea that there is a specific function of science, and that this function is essentially connected with finding things out. On the one hand, one may point to the personal purposes of scientists, policy-makers and other relevant agents. Is it not true that the actions of scientists, like other agents, are often explained by personal considerations – career-goals, power struggles, envy, vanity, and so on – without much concern for truth? On the other hand, one may point to other social functions that science may fulfill – to legitimize social or economic policies, for example, or simply to channel resources to the scientific community. Why should the goal of finding things out take precedence over such other functions and ambitions, in a balanced account of science?

There is no denying that such factors are important for understanding science, both as a macro-level social phenomenon and at the micro-level of explaining how scientific results and reputations are made and established. But it is also worth noting that such purposes and functions are in a way parasitical on the ambition to find things out. As an analogy we may use the playing of a game, like chess. One can play chess for a variety of reasons: for pleasure, to kill time, to please or humiliate somebody else, for money, to show off one’s intellectual skills, and so on. But all of these goals are in a

\(^{21}\)This general idea gives us a sort of criterion for judging scientific practices, in terms of their “veritism”, roughly, their tendency to increase our store of true beliefs and decrease the amount of false beliefs. Cf. Goldman (1999, ch 3) for an elaboration of the criterion of veritism, and a defense of it against rival approaches. As we noted in connection with Popper, above, truth is not enough – we want interesting truth, and are quite content with being ignorant of the vast majority of facts. We will come back to the notion of interest in the next section, and in later chapters as well.
sense external to the game itself: they cannot be appealed to in motivating or evaluating a move in the course of play. Whatever your ulterior motive for playing, you cannot play chess without acknowledging the internal goal of checkmating your opponent and winning the game. Perhaps you are not really trying to win, if you are playing with a child, for example, but if you are not even paying lip-service to the goal of winning you are not playing the game. In much the same way, at least official acknowledgement of the goal of truth-seeking, of respect for evidence, etc. is a presupposition for pursuing other goals through science – “hypocrisy is a tribute that vice pays to virtue”.

1.6.4 Science and common sense

One interesting problem of demarcation concerns the relation between science and common-sense. To what extent are scientific practices sui generis, specific rules of specific games, and to what extent are they simply refined and systematized versions of common-sense practices? Again, this is partly a question of the perspective one takes, different contexts may lead to emphasizing the similarities or the differences, respectively. When viewed from the outside, many disciplines seem to rely heavily on technical tricks of trade that appear to the outsider as hermetic secrets, perhaps more or less designed to keep non-initiates at bay. But they are still learned through intuitive exemplification and common-sense explanation, and the novice is not expected to just accept them as givens, as one just accepts the rules of a game, but is supposed to understand and appreciate the point and utility of what one does, in terms of discovery and understanding.

1.6.5 Knowledge and human interest

Jürgen Habermas has suggested an interesting take on the relation of science to everyday concerns, in terms of “knowledge constituting interests”.22 According to Habermas, each of the main divisions of empirical science is rooted in a specific type of cognitive interest, that leads back to “the natural history of the human species”. The natural sciences (“empirical-analytical”, in Habermas terminology) are rooted in a technical interest, geared to manipulation and control. The everyday context for such knowledge is manual work, where given circumstances and raw materials are treated so as to yield foreseeable results – a paradigmatic example would be the sort of recipes we find in cookbooks. Scientific laws are, so to speak, extreme refinements of cookbook recipes. Habermas thesis is not that the natural sciences are of interest to us only in so far as they are practically useful, but he holds that the forms of theories and the specific procedures of verification and control in these disciplines are rooted in the everyday procedures of work.

The “historical-hermeneutic sciences” (i.e., the humanities), on the other hand, are rooted in a different type of interest, namely a hermeneutic interest in communication and understanding. The criterion for successful

---

22The basic text is the lecture “Erkenntnis und Interesse” (Habermas; 1968), which is not to be confused with the book of the same title (Habermas; 1973) – an English translation of the lecture is included as an appendix to the translation of the book (Habermas; 1972).
communication is not effective manipulation, but unhampered consensus-
formation and cooperation, and the everyday roots of the hermeneutic dis-
ciplines are the activities of translators and cultural mediators of different
kinds.

So far, there is nothing very original in the actual descriptions that
Habermas gives of the different scientific domains – he more or less takes
the views supplied by classical positivism and hermeneutics, respectively,
and adapts them for his own purposes. When it comes to the social sciences,
his view becomes more controversial. The guiding interest of the social sci-
ences is “emancipatory”, and he introduces the basic idea in this way:

The systematic sciences of social action, that is, economics, socio-
cology, and political science, have the goal, as do the empirical-
analytic sciences, of producing nomological knowledge. A critical
social science, however, will not remain satisfied with this. It is
concerned with going beyond this goal to determine when theo-
retical statements grasp invariant regularities of social action as
such and when they express ideologically frozen relations of de-
pendence that can in principle be transformed (Habermas; 1972,
p X).

The main idea goes something like this. At different times and places, hu-
man life and cooperation are organized in social forms that appear like
invariant, natural givens to the people involved. The mode of existence, so
to speak, of these social forms is that they are internalized by the agents
themselves, in the form of a habitus, that we touched upon in connection
with Kuhn and Bourdieu, above. Although such patterns of action, feel-
ning and thought are normally below conscious control, they are to some
extent available to “self-reflection”, they may under certain circumstances
be raised to consciousness, and thereby to some extent brought under the
agent’s control.23 This sort of self-reflection has its roots, probably, in the
confrontation with other ways of life – it has been at the heart of the philo-
sophical tradition ever since its beginning, according to Habermas, and was
in due time bequeathed to the social sciences.

It is perhaps natural to be less impressed with this attempted link be-
tween science and everyday life than with the previous two. Emancipation,
in this sense, does not seem to be a daily concern in the way that work and
communication is. And what about the objectivity of science? Emancipation
appears to be a value-laden concept – is there supposed to be a common po-
litical agenda to the social sciences, or should we expect different traditions
of research, aligned along party lines?

The general question about the relation between research and ethical
and political values is best deferred to the ethics chapters, but if you just
look at the current state of many social studies, it is rather clear that Haber-
mas has a point. Among the most important theoretical and methodolog-
ical trends in the cultural sciences in recent decades are approaches like
feminism, post-colonialism and queer-theory – all inspired by emancipatory
concerns of exactly the type envisaged by Habermas. Again, we will come

23One of Habermas main examples of an emancipatory practice is psychoanalysis.
back to a more substantial discussion of the contrast between natural regularities and “ideologically frozen” social relations in Chapter 6, where we tackle the distinction between “brute facts” and “social constructions”.

1.6.6 Science and its neighbors

We have touched upon the relation between science and pseudo-science, with Popper, between science and proto-science, with Kuhn, and between science and everyday life, with Habermas. But for the working scientist another sort of demarcation will often be more practically pressing. When do I cease to be a historian and become a journalist or a novelist? When do I leave the scientific study of art or literature, and become a critic? Where is the line between political economy and economic policy? Many scientific disciplines have such neighboring fields, where the same subject matter is treated in partly similar ways, and partly by the same persons – either at different stages of their careers, or continually drifting back and forth. The demarcations in question are often vague, and it is not always important to make them more precise, but some reflection upon them may still tell us worthwhile things about what science is.

We cannot treat all such frontiers here, but must leave it to the reader to ponder what neighbors are most relevant to her own discipline, and reflect upon the pertinent similarities and differences in each case. But we will have a look at a few examples, to see what we can pick up of general interest.

The most immediately obvious difference, of course, is often the social setting – you cross the border of science as you leave the university and step into the offices of the newspaper or the government. But that is not the kind of difference we are after: we want to see a difference in the processes and products of the different settings.

Let us look first at the relation between history and journalism, as they would treat for example a contemporary international conflict. The similarities are there for everyone to see: the same sequence of events may be described in similar narrative prose, the same causes and motives may be offered as explanations, the same overall picture may emerge. But what are the differences?

One important difference concerns the use of evidence. A natural first thought is that the scientific account is expected to be more thorough, to be based on more evidence, and on evidence more carefully sifted and conscientiously employed. But this is a bit too simple. True, the historian may perhaps have had the time and the resources to dig up facts that were inaccessible to the reporter pressed by an imminent deadline. But it also works the other way around: there may be evidence that is freely available to the journalist that the historian cannot use – a bit like how a law court may be forbidden to use some available and clearly relevant evidence, because it was obtained in illegal ways. The decisive point is that the historian is restricted to traceable and intersubjectively available evidence – she must

24I choose the example, of course, because journalism is often occupied with the present or with the very recent past, and so the comparison with contemporary history springs most easily to mind. But time in itself is not important, many works of popular history are journalistic rather than historical in their approach, even if concerned with the very distant past.
leave a record of the evidence that she has used, such that it would, at least in principle, be possible for other researchers to go back and reexamine it. The journalist works under no such strictures — indeed, she is sometimes morally and even legally required to keep the sources of her information secret.

This contrast points to a general feature of science, relevant to its status as a social endeavor. Journalism, like most other forms of everyday knowledge transmission, relies heavily on trust and trustworthiness — it has an irreducible personal dimension, one might say. Science, on the other hand, seems to embody a striving for the impersonal, the intersubjective — for the objective, in one of the many senses of this notoriously ambiguous word. In practice, scientific experiments may not be repeated, and sources may not be checked, as often as one might think and perhaps hope — but it is still required that it should be possible to check them, possible for anyone prepared to take the time and the perhaps very considerable effort to do so.

Another noteworthy difference between scientific history and journalism has to do with how the agenda is set, with what questions are asked. The journalistic agenda is primarily driven from the outside: by what readers are (supposed to be) interested in, and by what happens at any given time, the news. The historical agenda is not entirely without connection to external social interests, of course, but it is nevertheless to a large extent driven by the internal concerns of the scientific community. Research focuses on aspects, themes and periods of the past that are deemed interesting from the point of view of the discipline: in order to try out interesting methods or models on new fields, for example, or to investigate questions left open by previous research.

Again, this points to a general feature of science, namely that it tends to be more systematic than other forms of investigation. This is reflected in the classical notion that scientific knowledge should ideally be presented in the form of an axiomatic theory, on the model of Euclidean geometry, but there is something to it even in fields where that ideal is clearly out of place. Researchers are generally expected to treat questions that naturally arise at the relevant stage of their speciality, and are not generally expected to motivate their choices by appeal to their external interest — a fact that may sometimes surprise and alienate the layperson.

Let us switch to another example: the relation between aesthetic research and aesthetic criticism. Many of the same things apply here as well, and, indeed, aesthetic criticism is often itself a form of journalism, albeit appearing in another section of the newspaper. But the most salient difference is surely the emphasis in criticism on evaluation and personal judgement. The critic is supposed to draw on her knowledge, of course, but even more she is supposed to apply her own taste and filter the artwork, so to speak, through her own personality. Again, the scientific side of the comparison leans more towards the impersonal and the intersubjective, while the critic is employed, precisely, to put her own subjective experience on the line.

I am not implying, of course, that there is a hard and fast line to be

---

25A famous Swedish encyclopedia, from the beginning of the last century, gives the following laconic definition of science: Science = systematic knowledge, see system and knowledge.
drawn here, between objective science and subjective criticism. We will come back in the ethics chapters to the general question of the place of values in relation to research, and in chapter 4 we will discuss the role that personal presuppositions have to play in interpretation. But a difference may be important even if it is a difference of degree, and it is rather clear that personal value judgements are much less pronounced in aesthetic research than in criticism, and that the researcher to a much larger extent is expected to supply intersubjectively acceptable reasons for interpretations that she makes.\(^{26}\)

As a final example, I want to say a few words about the relation between social science and policy making, as in the relation between political economy and economic politics. In general terms, it seems to be a rather straightforward question of application – scientific results are applied to practical decisions. Why is the application not part of science itself? I will mention two factors, that we have already touched upon.

The first factor is that to have action you need goals and values, and those are not supplied by science but by the relevant agents – for public policies, ideally, by democratically elected governments. What amount of inflation can we accept? How much unemployment? How much economic inequality? Once we have decided on such issues, and a host of others, maybe economic research can give us some hints about how to achieve them, and if they are possible – but research alone can never tell us what to strive for.

The second factor is uncertainty. In an ideal world, perhaps, science would be able to give us hard and fast advice about what goals are possible to achieve and how, but in reality there are always uncertainties of different kinds. Apart from the general uncertainty connected with empirical knowledge, which is the subject of the next chapter, there is the fact that causal reasoning in science makes use of abstract models, that are to some extent idealized in relation to real life situations. Certain factors are singled out and the relation between them are captured in the model, but are they the only relevant factors in a real situation? In some applications, the movement of a falling body is well enough approximated by purely gravitational model, in other cases additional factors, such as aerial resistance or magnetism or collisions with other bodies must be taken into account. In a similar way, economic models are based on idealized assumptions about agents and the motives and beliefs that influence their behavior. To develop such models and reduce the uncertainty is the work of science, but to take a practical decision based on an (uncertain) assessment of the uncertainties actually involved in a real life situation always involves a step beyond research.

1.6.7 Does it matter?

We have discussed the problem of drawing the border between science and some related human practices, in different dimensions. But what does it matter? Is it really so important to distinguish precisely between science

\(^{26}\)In a series of works Margaretha Rossholm has argued that the personal lived experience is in fact crucial for scholarly interpretation of the arts, and that it is largely hidden by rhetorical means, rather than absent.
and pseudo-science, between science and proto-science, between aesthetic research and criticism, and so on?

Let us look again at Popper’s concern over the false pretenses of pseudo-science. Why should we care whether the claims of astrology or creationism are scientific? Is it not much more important whether they are true and if we have good reasons to believe them? In many contexts, questions about the scientific status of this or that seems to be mainly a distraction, diverting the focus away from questions about what we have reason to believe, towards questions about the definition of science.

Popper’s worry is mainly about how the authority of science is abused. Opinions that have no other credentials are sold under the label of science, so to speak, and one might perhaps dream of some kind of certification, like the ones we have for ecological food-products or “fair-trade” commodities. But, then again, it is not a general mark of scientific opinions that they are well-founded or true – science aims at truth and evidence but most hypotheses will in time be weeded out rather than sustained, and they are no less scientific for that. In real life, the question “Why should we believe it?” is usually much more pertinent than “Is it science?”.

This is a good point, obviously, but there are still reason the think a little about the different problems of demarcation. We have not reached a definition of science, and probably for the good reason that none can be given, but thinking about the contrasts between science and related phenomena has pointed to factors that undoubtedly are important for scientific research: a critical attitude, respect for evidence, a certain form of social organization with ample space for peer review, a striving for intersubjective validity and transparency of methods and procedures, an emphasis on systematically motivated investigations, and so on.

It would be futile to draw a sharp boundary between journalism and history, or between literary theory and criticism, and devoting too much energy to classifying concrete cases one way or the other would soon degenerate into verbal quibbling. But we still draw these kinds of distinctions, and thinking a little about them can still cast important light on what science and research is.
Chapter 2

Questions, answers, evidence

One of the conclusions of the previous chapter is that the principal aim of science is to make knowledge available for us, to put us in a position to know things. More specifically, we think of science as a method for finding new knowledge, to find things out that we didn’t know before – or that something that we thought we knew was in fact not so. We seek knowledge outside of science too, of course, but usually not knowledge that is new in the absolute scientific sense. Using the Internet, or a telephone directory, or a teacher, I try to learn things that I do not know, but usually on the presumption that somebody else knows it, that the information is already there somewhere. The journalist, to take another example, tries to cure human ignorance, just like the scientist – but only the ignorance of her audience. She wants to tell her public what they did not know, but usually with the help of sources that already know it. The journalistic function is to relay knowledge and to cure local ignorance, while the scientific function is to create knowledge and cure global ignorance.

This makes it sound as if science is something very unique, and the scientific method perhaps something very exclusive. But from a methodological point of view, the most important contrast is not between new and old knowledge in the absolute sense, it is between finding something out for yourself and learning it from someone else. That is why practical problem-solving is such an important part of a scientific education: it doesn’t matter if a problem has been solved before, the important thing is that you solve the problem for yourself.

According to one traditional view, forcefully argued by for example Karl Popper, the scientific method is nothing but ordinary problem-solving, applied in a very systematic way to radically unsolved problems, in the quest for radically new knowledge. Here, at least for a while, we will follow the same approach, and come back to some critical reflections on it, later on.

So where shall we start? Introductory accounts of the scientific method usually start from the notion of a hypothesis, and go on to the procedure of testing a hypothesis by means of evidence. But where does a hypothesis come from? A natural suggestion is that a hypothesis is a suggested answer to a question, and so let us start with some observations about questions and answers.¹

¹Ref to Popper, Collingwood and Bromberger.
2.1 Questions

What is a question? A useful way to think about a question is as a set of alternatives. To answer a question is to pick one of the alternatives and give a reason to prefer it to the rest.

For yes-or-no questions this perspective is more or less self-evident. To ask if the café is open, is to ask for a choice between ‘The café is open’ and ‘The café is not open’. Most other questions cannot in this way be identified with a given set of alternatives, but, nevertheless, specifying the relevant alternatives is an important step in the process of clarifying the problem and getting ready to solve it.

Let’s look at an example from recent scientific history. What is the cause of gastric ulcers? For a long time the list of alternatives comprised mainly three types of factor, namely (a) different kinds of food stuff, like coffee, fat, cigarettes, etc. that were thought to increase acidity in the stomach, and (b) life-style factors, like stress, and (c) genetic factors. Numerous attempts were made to correlate factors of these types with gastric ulcers, and treatments were based on the results. But the real breakthrough came only with the addition of a new alternative, namely (d) bacterial infection.

Research begins with a question, and when looking for the answer it is useful to think of the process under two different aspects:

A. Find the relevant alternatives – these are the hypotheses that you want to consider.

B. Reduce the set of alternatives, by reasoning and testing.

Ultimately, of course, you want to eliminate all the wrong alternatives and have exactly one left – the complete answer to your question. But progress in research is usually partial, and the elimination of some previously open alternative is often an important result in itself.

So what are the relevant alternatives? What options should we consider when tackling a research question? There is, of course, no mechanical method for generating hypotheses, but there are, in fact, some general considerations to point to. We have, for example:

1. The suggested alternatives.

In science you are seldom the first person to work on a specific problem, and an important part of your scientific expertise is to know which solutions that have been suggested before, and the reasons that have been given for and against them. In your own research you are not only expected to argue for your own solution, but also to show in what ways it is superior to rival proposals. A second important class of alternatives is:

2. The natural alternatives.

Regardless of what has actually been proposed, most problems come with a natural set of alternatives, options that you would be expected to come up with if you only took the time to think about if for a while. A new problem may, for example, belong to a certain type of problem, where the right answer usually falls within the range of certain types of alternative.
Unfortunately, of course, there is no guarantee that the actually correct answer is neither suggested nor natural, so we must mention a third class of relevant alternatives, namely

3. The right alternative.

What makes really outstanding research special, as in the case of gastric ulcers, is often that it changes the range of alternatives, by finding one that is neither previously suggested nor natural, in the above sense. If you fail to consider alternatives from the first two classes, you will be blamed; if you find the right alternative outside of them, you are bound for glory.

2.2 Deductive inference

There will be some talk about deduction and inference in what follows. But what do these words mean? Here is an example of a deductively valid inference:

If Sven wants the job, he has a haircut today.
Sven wants the job.

Sven has a haircut today.

First some terminology. The two sentences above the divider are the premises of the inference, the sentence under the divider is the conclusion. What does it mean for the deduction to be “valid”? Precisely this: if the premises are true, the conclusion must also be true. A valid deduction “preserves truth”. We also say that the conclusion of a valid inference can be derived or deduced from the premises, or that it follows logically from them.\(^2\)

The validity of an inference has nothing to do with the actual truth or falsity of the premises or the conclusion. This inference is as valid as the first one:

If Lisa is a beetle, she is immortal.
Lisa is a beetle.

Lisa is immortal.

The inference is valid, but it is still useless as an argument for the conclusion, which goes to show that validity is not enough. A good argument must not only be valid, it must also be sound, i.e., its premises must be true. There are two ways to criticize a deductive argument: you can show that it is invalid, that the conclusion does not follow from the premises, or you can show that it is unsound, that at least one of the premises is false.

How about the following inference, is it also valid?

\(^2\)In modern logical terminology we would distinguish the concept of a derivation, which is a syntactic concept, from logical validity, which is a semantic concept. The description of deductive validity in terms of truth is a semantic description, but we will take no further notice of this distinction.
If Sven wants the job, he has a haircut today.
Sven has a haircut today.

Sven wants the job.

A little reflection shows that it is not valid. It may very well happen that the premises are true but the conclusion is false. Perhaps Sven does not want the job, so the conclusion is false. But if he wanted it he would have a haircut, so the first premise is true. He cuts his hair anyway, but for some different reason – maybe he wants to borrow money from his grandmother – so the second premise is also true.

So, this is not a valid inference but a fallacy, and even a fallacy of such importance that it has its own name: the fallacy of affirming the consequent. The type of valid inference exemplified by the first two examples also has a name: modus ponens.

Arguments are supposed to extend or increase the quality of our knowledge. How do they do that? That an assertion is the conclusion of a sound and valid inference is not in itself a reason to believe it. For a derivation to be a reason to believe the conclusion, the premises must be knowledge-wise superior to the conclusion – if I already have good reasons to believe the premises, the inference gives me a reason to believe the conclusion as well. This is the basis for a classical view of mathematical knowledge. According to this view, mathematics starts from specific sets of premises, called axioms, that are obviously true – they are self-evident, as the saying goes. The business of mathematicians is to proceed by valid inferences from the axioms to less obvious conclusions, the theorems, thereby proving these to be true as well.

Proof, in the mathematical sense, is not to be expected in the empirical sciences. But valid inferences can play a role in arguments in other ways, too. Suppose that I have reason to believe that the conclusion of a valid inference is false. Then I automatically have a reason to believe that at least one of the premises is false as well: for if they had all been true the conclusion would also have been true. The step backwards over the inferential divider preserves falsity, so to speak. If Sven does not have a haircut today, it must either be the case that he does not want the job, or there must be no connection between his desire for the job and the length of his hair, of the type expressed by the first premise – maybe he wants to be a roadie for a rock band? As we will see, this sort of backwards reasoning from the falsity of a conclusion to the falsity of at least one of the premises plays an important role in the testing of an empirical hypothesis.

2.3 Do I have a hypothesis?

The bulk of this chapter is about how to think about testing a hypothesis. Most of it will be spent on some different models of the relation between a hypothesis and the evidence that may support or weaken it, as the case may be. But before we come to that, we need to say a little bit more about what a hypothesis is and what role it has in research.
It is customary to define a hypothesis as a statement that one is not certain whether it is true and that one wants to test in one’s research. It is also often presupposed that to "have" a hypothesis is to believe in it, so that research would consist in forming beliefs and then trying to prove these beliefs to be true – as in "my hypothesis is that ...". In the present context, however, it is more useful to think of a hypothesis in a less personal way, simply as a possible answer to a question. The aim of research, with regard to a question, is to find out which answer is true, by accumulating evidence for and against the different possibilities. Which answer you happen to believe in at different stages of the research process is, by and large, unimportant – your results should be relevant for anyone who is interested in the question, regardless of their prior beliefs.

It is sometimes assumed that hypothesis testing is most at home in the natural sciences, and largely irrelevant to most of the humanities and to at least some of the social sciences – perhaps because they are "descriptive" or "inductive" or "interpretative", rather than "theoretical" or "deductive" or "factual". While there are indeed pertinent differences (which we will come back to) behind such ways of speaking, the conclusion is not true – as long as there are questions that you aim to answer in your research, you will also have hypotheses to consider and to evaluate against relevant evidence.

One reason to think that hypothesis testing is irrelevant to some research is the misleading notion that this would imply that the research process must always start from a hypothesis. It seems obvious that a lot of research begins without any clear idea of the result to be reached, and that this is often a good thing – signalling an open mind and a readiness to be informed by one’s interlocutors or the material at hand. But wherever the process starts, it would not be research unless somewhere along the way there were questions asked and answers considered. To consider an answer to a research question is the very same thing as to formulate a hypothesis, to contemplate what possible evidence would be relevant for and against it, and then try to ascertain what the actual evidence says.

Another reason why the role of hypothesis testing in research may be overlooked is precisely the fact that it is omnipresent – that we do it all the time without thinking about it. As soon as we formulate anything that might be in any way contentious, we start to ponder arguments for and against it, usually by relating it to actual or possible evidence. Once you realise that you do this all the time without even thinking about it, however, you may wonder why you should need a theory of empirical reasoning at all – would that not be as awkward as using a formal grammar to speak your native language? There is something to this objection: a formal theory of evidence and testing can never replace common sense and thinking for oneself. Properly used it is a supplement to common sense reasoning, a supplement which can help us to clarify complex issues and to avoid potential pitfalls, by being a bit more careful and systematic. It is in this spirit that one should approach the theoretical models of the evidence relation that will be sketched in the following. The point is not to slavishly accommodate one’s practice to a formal model, but to use these models as points of comparison and inspiration for a deeper reflection on what one already does.

There are two basic ideas about the evidence relation, one of them
grounded in prediction and the other in explanation. The first idea is that a hypothesis is supported by the evidence that it predicts. Very roughly: if the actual evidence is the way we would expect it to be if the hypothesis were true, this can be taken as an indication that the hypothesis actually is true. If playing a lot of computer games makes gamers more aggressive we would predict them to score higher on tests for aggressiveness than non-gamers – do they? The second idea is that a hypothesis is supported by the evidence that it explains. Again very roughly: if the actual evidence would be explained by the hypothesis if it were true, this can be taken as an indication that the hypothesis actually is true. The fact that I have a fever would be explained by a flu infection, so my fever is evidence that I have the flu.

As one can imagine, there are a lot of complications to accommodate before these rough ideas can be made precise, for example connected to the fact that there are usually rival hypotheses to consider, that would predict or explain the same given evidence. In the following sections I will, first, consider one classical way to develop the prediction idea, traditionally called “the hypothetico-deductive method”, and, second, present a more recent way to develop the explanation idea, most commonly called “inference to the best explanation”. I will not try to force a choice between the two perspectives, but instead try to show that they both give valuable and complementary insights into how we already think and how we should think about evidence in science.

Both these models are qualitative in character – they characterise what it is for a hypothesis to be supported or undermined by evidence, but they do not give a numerical measure of how strong the support is. There are also ways, however, of thinking about the evidence relation that have a more numerical look to them, usually based on the mathematical theory of probability. Here belongs various theories of statistical inference, which are of fundamental importance in large regions of the social and the natural sciences, but also the more general framework of “bayesian” reasoning. This is not the place to present such theories in any detail, but some of the basic ideas should be part of any general education in science, and can also throw some light on intuitive reasoning about empirical support that have no relation to statistics.

2.4 Testing a hypothesis

Let us suppose that we have a research question and a range of alternative answers to it, a range of hypotheses to be considered. What next? A natural impulse is to pick the hypothesis that looks most promising, perhaps one that you for some reason already believe (or hope) is correct, and try to prove it to be right. For a variety of reasons, this is not generally the best strategy. Contrary to popular imagination, research is not primarily a process of proving right ideas right, but of proving wrong ideas wrong – what you antecedently believe is, of course, often important as an incentive to try this or that line of research, but it may also act as a bias that makes you overlook other possibilities.

What alternatives to test, and in what order, is a pragmatic question.
Sometimes there is a “received wisdom” to take into account: some hypothesis is so generally accepted that you cannot get other alternatives on the table without challenging it. Sometimes the cost – in effort, time or money – of testing different options are different, and it may be wise to try the cheaper options first. But the general point always applies: you have not done enough to support “your” hypothesis until you have eliminated all the competition.

A good metaphor for scientific research is criminal investigation. You have the question – who did it? – and the first thing you need is a list of suspects. You fill this list from a variety of sources: the “usual” suspects, tips from the public, people with strong motives, etc. And as you fill the list you start trying to eliminate names from it by testing for each one, as it were, the hypothesis that this is the actual perpetrator. How do you do that? You ask questions of the form “If this person did it, what then?” He must have been at the scene of the crime – was he? No, he has an alibi, strike him from the list. Yes or possibly yes, keep him on the list. He must have had the requisite skill, strength and equipment to accomplish the deed in the way it was in fact done, did he? Etc.

A lot of the time, this work is painstaking routine, eliminating candidates that were never very hot in the first place, but sometimes someone emerges as the prime suspect, and all the resources go in that direction. Hope is surging and the end seems near, but this is also a point of high risk – concentration on one alternative may lead to an impasse, when the real culprit is someone else on the list, or, even worse, when we didn’t throw the net wide enough in the first place, so that we end up with an empty list and a killer going free.

2.4.1 A real case

In history, murder cases abound. A famous question from Swedish historiography is “Who killed Charles XII?” The list of possible suspects comprises all those close enough to the trench outside the Castle of Fredriksten, Norway, that evening on November 30, 1718, to reach the king with a bullet. But in the process of elimination we don’t go directly to the individual suspects, but try to find questions that will prune the list of as many of them as possible at once. Was the bullet fired from the Norwegian or from the Swedish side? The Swedish alternative is more interesting, in its ramifications for the further interpretation of Swedish history, but the Norwegian alternative is antecedently more likely so let us try that first. Suppose the king was shot from the Norwegian side – what then? The bullet must have entered his head from the front.3 Suppose that it did – what then? According to a well confirmed general principle, exit holes in such cases are much larger than entrance holes, so the hole in the front of the kings head should be smaller than the hole in the back. Well, is it? No, the front hole is bigger, so the hypothesis is rejected. Or is it?

3My account of this case is, of course, drastically simplified. In particular, the holes in the skull are not in the front and the back, but on the left and the right side. It is, however, generally agreed that Charles stood with his left side towards the castle, and so I refer to the left hole as the front hole.
Let us describe the relevant piece of reasoning a little more pedantically, in the technical terminology of the philosophy of science. It may be taken as a clear example of what is usually called the “hypothetico-deductive method”. The first step in applying this method is to have an hypothesis and deduce a testable consequence from it – that’s where it gets its name from. Basically, we try to find a useful answer to the question “what then?”. Let us formalize this step as a derivation, with premisses and a conclusion. The first premiss is the hypothesis itself: the king was shot from the Norwegian side, i.e., from the front. The conclusion, the testable consequence, is that the hole in the front of the skull should be smaller than the one in the back. But to derive that conclusion we need to assume the general principle mentioned earlier. So here is our reasoning:

First premiss: The King was shot from the front.
Second premiss: Entrance holes are smaller than exit holes
Conclusion: The front hole is smaller than the back hole

The conclusion is our testable consequence, something that follows from the hypothesis and that we can actually observe, if it is true. It has also been actually tested by historians. The body has been disinterred and examined several times, most recently in 1916, with the result that the hole at the front is in fact the bigger one. So we draw the further conclusion that the hypothesis is false. How do we do that?

The deduction of the testable consequence from the hypothesis may be summarized as an implication: if the King was shot from the front, then the front hole is smaller than the back hole. And we may use this implication as a premiss in a further argument, that goes like this:

Test implication: If the King was shot from the front, then the front hole is smaller than the back hole.
Actual evidence: The front hole is not smaller than the back hole.
Conclusion: The King was not shot from the front.

This is a logically valid argument, which means that if the premisses are true, the conclusion must be true. This form of inference also has a name in traditional logic: modus tollens. So does the evidence prove the hypothesis to be false? We will come back to that . . .

What about the contrary hypothesis, that Charles was shot from the Swedish side, i.e., from the back? Is that proved to be correct by the evidence? Let us skip most of the steps and go directly to the final argument, which would look like this:

Test implication: If the King was shot from the back, then the front hole is bigger than the back hole.
Actual evidence: The front hole is bigger than the back hole.
Conclusion: The King was shot from the back.

As we already know, however, this is not a logically valid argument – it is just our old friend the fallacy of affirming the consequent. This logical asymmetry, between negative and positive evidence, is the basis of Karl Popper’s famous denial that a hypothesis can ever be verified by positive
evidence. At least in some of his moods, Popper would claim that the best we can say about a hypothesis that accords with the evidence is that it is not yet falsified.

But isn’t Popper’s pessimism out of place in this case? Surely, if the King was not shot from the front he must have been shot from the back? There is something in this, of course, and the reason is the symmetry of the case: there seems to be only two alternatives, and so whatever proves one of them wrong must prove the other one right. This situation is not uncommon in the history of science: when just two alternatives face each other we try to device a “crucial experiment” to choose between them.

We will not go further into the logical formalities, but only note that this, again, shows the importance of viewing a hypothesis in relation to its alternatives. The actual evidence supports the hypothesis of a Swedish killer, because this seems to be the only alternative that is compatible with the evidence. Had there been other alternatives, the process of elimination would not have been complete.

2.4.2 What about the hat?

As anyone familiar with Swedish history knows, the excavation of the King’s body did not put an end to the debate about the killer. The case is not taken to be settled, and the hypothesis of a Norwegian bullet is, I think, the more popular one among historians today. What is wrong – can historians not recognize a valid argument when they see one?

Look again at the first argument, the alleged falsification of the Norwegian hypothesis. Sure, if the premisses are true the conclusion must be true, but are the premisses true? The test implication was, in turn, based on a deduction which relied on an auxiliary hypothesis, to the effect that exit holes are always bigger than entrance holes. One reason for this is that the bullet gets deformed, flattened, at the entrance and so is actually bigger when going out. But suppose that it hit something else before hitting the skull, and already was deformed going in? And perhaps it lost some of its momentum on the way, and actually made a smaller hole going out?

It turns out that this may actually be the case. The King was wearing a thick felt hat and there is a matching hole in the hat, as well. Experiments with similar cloth and bullets show that the hat is solid enough to cause the deformation, and the size and form of the two holes is just what should be expected, given the passage through the hat.

When relying on the principle of smaller entrance holes, we treated it as a law – as if it was true in every case. But like almost any principle that would be used in similar circumstances, it is not a strict law, about what must happen, but only a ceteris paribus principle, telling us what normally happens if there are no intervening factors. And the presence of the hat is just such a factor in this case.4

The important point, for us, is that the deduction of testable consequences will always have to rely on auxiliary suppositions that we take for granted in order to test our main hypothesis. And this means that even falsification by means of evidence is not as clear cut as Popper sometimes

4“Ceteris paribus” means, roughly, “with other conditions remaining the same”.
makes it out to be – and that even “crucial” experiments are not always so crucial. As the prediction of the evidence relies on several premisses, there is always a possibility that the main hypothesis is innocent, and that the fault lies with some of the auxiliary suppositions.

Does this render the hypothetico-deductive method worthless, as a tool for measuring a hypothesis against the evidence? To the contrary, this is one of the main benefits of thinking according to this pattern. The necessity to rely on auxiliary assumptions, when trying to fit one’s theories to the evidence, is real, but most of the time these assumptions are not made explicit but are doing their secret job in the darkness. The exercise of formulating your prediction of the evidence as something approximating an explicit derivation helps you to see, and then maybe to question, the assumptions you actually presuppose.

2.4.3 Radiocarbon dating

Here is another example, to further illustrate the notion of an auxiliary hypothesis and to prepare us for some questions about the notion of evidence, that will concern us in the nest section.

Think about radiocarbon dating in archaeology. The basis of this method is the fact that all organic material contains carbon, which the living organism takes from the environment. Most of the carbon in nature is ordinary C\textsubscript{12} but the atmosphere also contains the radioactive isotope C\textsubscript{14}, which is continuously created there by the impact of cosmic radiation. As long as it is alive, the proportion of ordinary and radioactive carbon in an organism is the same as in the environment, but once the organism dies, and the exchange with the environment stops, the amount of C\textsubscript{14} starts to go down, because of radioactive decay. The proportion of the two forms of carbon becomes an index of the time that has passed since the tissue was part of a living organism. Let us take an archeological hypothesis, like

This grave is approximately 4000 years old.

and suppose that we want to test it by radiocarbon dating. For this purpose we need a test implication of the form:

If the grave is 4000 years old, the proportion \( \frac{C^{14}}{C^{12}} \) in this sample = \( \frac{x}{y} \).\(^5\)

But how can we derive the assertion that the proportion \( \frac{C^{14}}{C^{12}} \) in this sample = \( \frac{x}{y} \) from the assumption that the grave is 4000 years old? There is obviously no rule of logic that takes us directly from the one to the other, and so the road from the assumption to the consequent must be built with large amounts of background knowledge. That background knowledge supplies the auxiliary hypotheses that we need to derive the purported evidence from the hypothesis. Let us sketch the derivation and fill in some of the relevant assumptions:

\(^5\)In real life, assertions about measurable quantities, in scientific contexts, always carry an explicit or implied margin of error, but I will ignore that complication here.
The grave is 4000 years old.
The sample comes from the grave.
The proportion $\frac{C^{14}}{C^{12}}$ in the sample when it was placed in the grave was $\frac{z}{r}$
The half-life of $C^{14}$ is 5730 years

The proportion $\frac{C^{14}}{C^{12}}$ in the sample is now $\frac{x}{y}$

The derivation is not really complete even in this form, of course, but it seems reasonably clear what would have to be added to specify it in passable detail. For example, we need to know not only that the sample is from the grave, but also that it is contemporary with it, and neither was there when it was constructed, nor has been moved there afterwards. We also need to perform some mathematical operations to transform our assumptions about the original proportion and the half-life into an assertion about the current proportion, and so on. But let us suppose that we have made the deduction and that we check the conclusion against the facts. What happens then?

The conclusion of our deduction is a “prediction” about the evidence, about what we shall observe when we make the relevant observation. As before, there are two possible outcomes: the prediction may come out as true or false. That the prediction is correct ought to speak in favor of the hypothesis; that it is incorrect ought to speak against it.

Start with the negative case. If the prediction is contradicted by the evidence, and the derivation itself is correct, we learn that at least one the premisses must be false. But which one? The focus of our interest is the assertion about the age of the grave. But how do we know that it is precisely this premiss that is the guilty one? We only know it, of course, under the assumption that all the other premisses are true. In principle, one can always evade a purported falsification by questioning some of the auxiliary assumptions.

When radiocarbon dating was first introduced, in the sixties, it led to many surprising results, that overturned existing chronologies and forced revisions of earlier views about cultural priorities and directions of influence. Such revisions did not go unopposed, of course, and the reliability of the method was widely questioned. In particular, one pointed to the auxiliary assumption about the original rate of $C^{14}$. How do we know the proportion of radioactive to ordinary carbon at various periods of the distant past? At first, one just assumed that this proportion is constant over time, at the same value as today, but the surprising results could just as well be taken to falsify that assumption, making it possible to preserve the well-entrenched framework of traditional archaeology. How can one handle such a suspicion? The natural strategy is to look for an independent way to test the auxiliary assumption, and this is also what actually happened. The main tool was dendrochronology: by correlating series of annual rings of very old trees with radiocarbon analysis one could calibrate the initial values of $C^{14}$ for long periods of time. (The outcome was that the original radiocarbon dates had to be corrected even further in the sensational direction.)
What about the positive cases, where the real evidence fits the prediction – what do we learn from that? Very little, it seems, at least from a purely logical point of view. The negative case excludes the possibility that all the premisses are true, and sends us looking for the culprit. The positive case, as we have seen, does not exclude anything – it is compatible with the truth of all the premises, to be sure, but it is also compatible with all or any of them being false. So what shall we say?

One possibility is to bite the bullet and admit that experience can never verify an empirical hypothesis. The best we can ever say about a hypothesis that has successfully passed every test is that it has not been falsified. This is Popper's standpoint, at least in some of his moods, and it still has followers. But for most of us, surely, this conclusion is much too radical. We demand that a new drug shall be submitted to wide-ranging tests before it is released for public use, to ensure that it is effective and not too harmful. If the testing goes the way we hope, we take ourselves to have reason to believe that the treatment is indeed beneficial. To just say that it has not yet been proved harmful and inefficient does not seem to be a fair summary of the procedures.

As before, the intuitive solution is to look at the alternatives. Suppose that the actual outcome is only possible under the assumption of the hypothesis, that there is no alternative way to account for the evidence. If this were really the so, the evidence would indeed show that the hypothesis must be true. In principle, this is perhaps very seldom the case – we may always dream up some wild and crazy alternative explanation of the evidence, or just hold out for the possibility that future research will provide some alternative that is beyond our present imagination. But in practice, we will take our hypothesis as confirmed by the evidence when we have excluded all realistic alternatives, like random variation or the placebo effect in the medical case, or other values for the original radiocarbon rate in the archaeological case.

Take another current and important example: global warming. A pertinent hypothesis in this connection is that higher levels of CO2 in the atmosphere, due to the burning of fossil fuels, leads to a warmer earth. A plausible place to look for empirical confirmation is to temperature statistics, to see if the planet has indeed become warmer in a way that reflects rising levels of CO2. But, as most of us have learnt over the years, things are not quite so simple. Even if we disregard the considerable problems of actually measuring global temperature in a meaningful way, we have to deal with the fact that temperatures fluctuate with a bewildering plurality of factors, working in different directions. How can we single out the contribution made by human induced CO2? The hypothesis of global warming, in the relevant sense, is even compatible with long stretches of actual cooling – it does not imply that temperatures will actually rise, but only that they will be higher than they would otherwise have been. And if we find that temperatures do rise, how shall we exclude alternative explanations, like volcanic activity or increased activity of the sun? The challenge is to find a “fingerprint” of CO2 emission, a type of influence on earth temperature that would not be expected on the basis of the alternatives. To exclude volcanic activity one may look at temporal variations, to see if they correspond with known volcanic sources, or one could try to ascertain the size of the
expected effects, to see how much of the measured temperature increase they would explain. To exclude the explanation by increased influx from the sun, one could look at the precise distribution of temperature changes: sun activity would be expected to raise temperatures in the outer layers of the atmosphere most, while greenhouse gases would primarily trap heat at lower levels, and so on. From our current point of view, the point remains the same: confirmation of one hypothesis is inseparable from the falsification of alternatives, and the evidence one needs to consider depends heavily on which alternatives have been suggested and are considered relevant.

2.5 What is evidence?

One may have all kinds of reasons for, or against, a belief. Why do I believe that Buenos Aires is the capital of Argentina, or that the moon is much smaller than the earth? Because people have told me so – I believe it on authority. Is that a reason, or even a good reason? Yes it is, if the authority is trustworthy, or, perhaps, if I have reason to trust the authority. To provide authoritative reasons that others can rely upon is one of the main functions of science.

But scientific knowledge is not itself based on authority, but on evidence. What is evidence? Roughly: a piece of evidence is an observable fact that may be used to support or refute a hypothesis which is not itself decidable through observation in the same way. An historical hypothesis is about the past, but it is tested against remnants and documents that may be examined now. A grammatical hypothesis is tested against what speakers of the relevant language are prepared to say and accept as correct. An interpretation of a literary text is tested against the wording of the text. An ethnological hypothesis is tested against what informants say in interviews, or how they behave in interactions. And in every discipline there are specific rules about evidence: how observations are to be conducted, how evidence is to be recorded or preserved.

But what is it for a fact to be “observable”? The natural suggestion is that observable facts are those that we can directly ascertain through our senses: what we can see, hear, touch, taste or smell. But I cannot see the rate of different carbon isotopes in a sample. What we do, is to measure it by means of sophisticated equipment, which is designed on the basis of scientific theories, reaching right down to fundamental quantum mechanics. If my evidence is the radiocarbon rate, it is already heavily infused with theories – it is “theory-laden”, as the saying goes. And the same thing goes for most other things that passes as evidence in science. Sure, I can test my hypothesis about Strindberg’s misogyny against what the text says – if I know the language and the literary conventions of his day. A quick glance in a book in a language we don’t understand is enough to dispel the illusion that we can actually see what is in a text.

So what shall we say – is it not the rate of the isotopes that is our evidence, but only the meter reading? To some extent, this is a terminological question. If we decide that it is the reading that is the real datum we will have a longer and much more complicated derivation of the evidence from the hypothesis, and will face an explosion of new auxiliary assumptions,
for example about the supposed working of the equipment and whether it really works as supposed. Taking this option, we go down a path that classical epistemology has investigated in great detail, looking for “pure” data that do not presuppose any theoretical assumptions at all. On the way, we will encounter problems that are philosophically important, but have little practical bearing on scientific research.

One problem is that it is hard to give examples of pure observations. Is it even OK to stop at a simple thing like the reading of a meter? Does that not presuppose that my senses work as they should, and that I am not dreaming, for example? Maybe the real datum is only that I have a certain type of visual experience, that it looks to me as if there is a meter there, showing a certain value? That my experience is really caused by such a meter – is that not a hypothesis? But how would I test that hypothesis, if I cannot presuppose anything at all about the external world? And is it really possible to build objective science on a basis of purely subjective observations, where different researchers never really observe the same thing?

Another problem, that we already touched upon, is the explosion of auxiliary assumptions. The less theoretical baggage that is already built into the evidence, the more complicated the derivation of it from our hypothesis becomes. The American philosopher W. V. O. Quine (1953) has taken this line of thought to its ultimate conclusion, claiming that each empirical test involves the totality of what we take ourselves to know, in science and in daily life.

For practical purposes, it is surely better to stick to the idea that it is the radiocarbon rate that is our evidence. Speaking like this, the distinction between evidence and hypothesis becomes functional and relative, rather than absolute. What is taken as evidence in one context, may be treated as hypothesis in another context, that may in principle always be actualized by questioning the evidence. So we accept that evidence is usually and perhaps always laden with theory – does that matter? Not in general. The only thing we have to guard against, is if the evidence depends on the very hypothesis that we want to test – then we will be faced with a piece of circular reasoning, that may play havoc with the whole idea of empirical testing. Suppose, to make up another example with archeological flavor, that I claim that a certain location that I excavate is a cult-site, and cite the occurrence there of some cult-objects as evidence for that claim. If I then cite their occurrence on a cult-site as evidence for them being cult-objects, I am clearly on my way down a slippery slope. But the radiocarbon case is not like that. Sure, the assertion that the rate of $\text{C}^{14}$ in a certain sample has a certain value is laden with vast amounts of physical and other kinds of theory, but the important thing is that it is independent of the hypothesis that we want to test, concerning the age of the grave. The interesting question is not whether our evidence is theory-dependent, just in general – in practice it always is – but on what theories it depends.

This slippery slope seems related to the “hermeneutical circle”, and we will have more to say on the vices and virtues of different kinds of circularity in chapter 4.
2.6 Inference to the best explanation

Our question, at the moment, is about the connection between hypothesis and evidence. We want to know who killed the king, and we measure the size of the holes in his cranium. But what is the connection between the one and the other? The hypothetico-deductive method supplies one answer to that question, roughly:

Evidence E supports hypothesis H, if and only if, a statement that describes E can be deduced from H (together with suitable auxiliary hypotheses).

This way of thinking about evidential support is geared to the thinking pattern we have described, when we start from a hypothesis and ask “what then?”. But there are other ways of thinking about evidence, that fits more naturally in other situations. In research, and perhaps particularly in the humanities, we often seem to start from the evidence, rather than from a distinct hypothesis – the archaeological finds, the archives, the artworks, the interaction with informants, these are what we live with on a daily basis from the beginning of our projects, and it is from them that our theories and conjectures arise. The guiding question seems to be, not about what evidence we can find to support or refute this or that hypothesis, but about what hypothesis we can find to explain this or that evidence. And this question points to another way of formulating the evidential relation:

Evidence E supports hypothesis H, if and only if H is part of the best explanation of E.

This pattern of thinking also has a name in the philosophy of science: “inference to the best explanation”. And it seems to fit our example perfectly: are the actual deformations of the skull best explained by a shot from the front or from the back? The best explanation is also the hypothesis that the evidence supports.

How important is the difference between the two formulations? Do they say the same thing in different words or are they rivals that will lead to ultimately different results in concrete cases? This is a difficult question, which hangs on the notion of explanation – on some theories of explanation they are more or less equivalent, on other theories they may be importantly different. We will not go into that here. Even if the hypothetico-deductive method and inference to the best explanation are ultimately equivalent, they are both useful, by giving us different perspectives on the process of research, that complement each other.

One thing that makes inference to the best explanation worthwhile, as a methodological tool, is the explicit emphasis on alternatives, that we already noted to be so important in connection with the examples above. From the hypothetico-deductive perspective this comes as an afterthought, so to speak, as an added caution, but here it is granted from the outset.

---

7The basic idea goes back to Charles Sander Peirce, but it was reintroduced into the modern debate by Gilbert Harman (1965). Peirce used the term “abduction” for this form of inference – an interesting application of it, from the point of view of the cultural sciences, is in Peirce (1958).
that the hypothesis under current scrutiny has alternatives, that it is not enough that it explains the evidence – it must be the best explanation. Particularly when evaluating a favorite hypothesis, this helps one to keep a critical perspective. Well and good, my hypothesis explains the evidence, but are there other possible explanations of the same findings? Perhaps even better ones?

2.6.1 Testimony

Evidence in the humanities is often in the form of linguistic narration: a written text or an oral utterance that purports to describe a chain of events or other circumstances. Is the theory of inference to the best explanation applicable to such evidence, too?

Let us get back to the murder case. So far, we have concentrated on the forensic evidence, but apart from that we have the witnesses and their testimonies. Mrs X tells us that Mr Y did it, and that she even saw him do it, with her own eyes. Our evidence is her utterance, what she said. Now, what is the best explanation of that utterance?

One possibility, of course, is that she said what she said because she believed it to be true, and wanted us (or whoever she said it to, her addressee) to know the truth. Our next question, then, becomes: why did she believe it? What is the best explanation of that? Again, one possibility is that she believes it to be true because it is true: her visual experience which caused her belief (which caused her utterance) was in turn caused by the actual event of Mr Y doing what she says he did. The complete explanation of any of these steps would, of course, be much more complicated, but that doesn’t matter – the important thing is that Mr Y’s actually committing the crime is part of the best explanation of Mrs X’s saying that he did.

But, of course, there are many other possible explanations of her utterance, and we have to make sure that none of them is better than the one involving the guilt of Mr Y. She may have lied to us, i.e., she did not want us to believe what she said because she believed it herself, but for some other reason – her dislike for Mr Y, perhaps, or her wish to protect someone else who she believes to be the real killer. Or she may herself be deceived, i.e., her belief that Mr Y did it may be caused by something else – perhaps she saw some other person do it, that she mistook for Mr Y.

Evidently, these questions about the possible explanations of Mrs Y’s testimony correspond exactly to the questions that an historian or a journalist have to ask about her sources. Are the sources sincere, i.e., do they say what they say because they believe it, or is there a hidden agenda somewhere, another reason for the sources to say what they say? And are the sources well informed, i.e., are their beliefs the products of reliable processes that lead back to the actual events, so that these events are part of the best explanation of those beliefs?

Is this a fruitful way to think about sources and testimony? Yes, it is. To assess the evidentiary value of a source one must look at the whole range of possible explanations of it saying what it says, and not just stare blindly at the question of whether what it says is true. Our ultimate goal, of course, is to know what is true, but our first task is to judge what this testimony gives
us reason to believe, and for this the question about the best explanation of it saying what it says is exactly the right one.  

### 2.6.2 Interpretation and explanation

With these questions about the evidentiary value of sources, we seem to be closing in on the subject of interpretation, that will occupy us in subsequent chapters. Here I will just make a few remarks on the connection between interpretation and explanation, that we will later place in a broader perspective.

When confronting “meaningful” material – like texts and other utterances, for example – there are prima facie two types of questions to be asked. We may want to know what the object, for example the text, itself means, and we may want to know what information it carries about other things, what it reveals about circumstances external to it. When using a text as a source it is the second question that is in focus, while the first one is only a step on the way. In other cases, as for example in literary interpretation, the focus may instead be on the first question.

Using utterances as sources is of course not the prerogative of detectives, journalists and historians. The ethnologist conducting an interview is doing the same sort of thing and needs to apply the same perspective: to see the utterance of the other as the product of an action which is itself the outcome of a complex set of circumstances. How does the other view the situation of the interview? What are her goals? What pressures is she under? What bias and interests does she have? In short – why does she say and do what she says and does, what is the best explanation of her actions? A large part of that explanation, of course, will concern what she wants, what she intends, how she feels and what she believes – and sometimes this is all we want to know, we are using someone’s acts and utterances as sources of knowledge about their “subjective reality”. But, as every child knows, there is no particular place where you must stop asking why, and sometimes we want to go further, and know why she feels and thinks the way she does – what social, historical and personal circumstances that explain the beliefs and intentions that in turn explains her acts and utterances.

What about the first type of question, about what the text or utterance itself means – can it be seen in the same light? What is in the text is, of course, the most important evidence for or against an interpretation of it, i.e., a hypothesis about what it means. But can the relation between evidence and hypothesis, between text and interpretation, in this case, too, be seen as an inference to the best explanation. Does the meaning, as ascribed by the interpretation, explain the text?

Well, that depends on what we take the meaning of an utterance to be, and we will not go deeply into that question until the next chapter. But this way of thinking about testimony is not uncontroversial. According to some, testimony has a sort of intrinsic credibility that allows us, and perhaps even obliges us, to take it at face value as long as we do not have specific reasons to challenge it. Such positions are defended for example in Coady (1992) and Kusch (2002) – both works also give useful overviews of the debate up to their own publication. Lipton (1998) sketches an account of testimony based on inference to the best explanation.

The notions of a text and of something being “in” the text are in fact rather complicated, and we will come back to them in the next chapter.
CHAPTER 2. QUESTIONS, ANSWERS, EVIDENCE

at least for some notions of meaning, the theory of inference to the best explanation gives a surprisingly good idea about what interpretation is. In many cases, asking about what something means is to ask what someone means by it, and the answer refers to the beliefs and intentions of that someone, the speaker or the writer, for example. Interpretation, in such cases, becomes a special case of intentional explanation – one explains the action of speaking or writing, the speech-act, just as one explains any other type of action, and the explanation is the interpretation.\(^\text{10}\)

As we know, looking for intended meanings is not à la mode in most humanistic disciplines today, and sometimes the very notion of utterances as intended products of actions is put into question. Instead one searches for hidden layers of meaning, beyond the awareness and control of the speaker – as in psychoanalytic interpretation or deconstruction. Can such interpretations also be seen as inferences to the best explanation? Yes, they can. When the psychoanalytic renounces the surface meaning and the conscious intentions of the analysand in favor of unconscious and repressed intentions, it is precisely because she thinks the conscious intentions fail to give a satisfactory explanation of what the analysand says and does, so that other factors must be taken into account. The quarrel between intentionalists and anti-intentionalists, in this arena, is not about the goal of explaining what is in the text but about the value of different kinds of explanation, i.e., about what explanations really are the best.

2.6.3 What explanation is the best?

Inference to the best explanation seems to provide a useful and illuminating way of looking at the relation between evidence and hypothesis, in the humanities as in many other contexts. But what is an explanation and how does one tell which of several competing explanations really is the best? I will not say much about the first question here, but leave it for another chapter; but we must say at least something about what makes one explanation better or worse than another.

The most natural suggestion, I suppose, is that the best explanation is the true or correct explanation, or, for practical purposes, the explanation most likely to be true. In a sense, of course, this must be so, but it does not help us much in the present context. We want to use the notion of a best explanation to evaluate competing hypotheses in the light of the available evidence, i.e., to decide which one is most likely to be true. But then we cannot presuppose that we already know this. The British philosopher Peter Lipton (2004) makes a suggestive terminological distinction between the likeliest and the loveliest explanation, and notes that for the purpose of inference to the best explanation it is the value of loveliness that is relevant. We want to find those other virtues, so to speak, of explanations that can guide us to the explanations that are likely to be true.

So far, we have spoken about the relation between hypothesis and evidence as if it was a “local” phenomenon, concerning one piece of evidence in relation to a range of competing hypotheses. This is an oversimplification,
of course. The relevant base of the evaluation is all the available evidence. And not only that: the hypothesis that we accept should be a convincing partner not only to all the other direct evidence that we have but to everything else we take ourselves to know.

Let us go back again to the death of Charles XII. On the assumption of a Swedish murderer, how does the rest of the story go? Is there a plausible candidate for the role of the actual killer? Is there a plausible motive? Was the murderer alone or was he part of a larger conspiracy? Who else was in on it? Supposing that answers to such questions are suggested, is there independent evidence to support them?

We introduced inference to the best explanation partly to remedy the tendency of the hypothetico-deductive method to put too much weight on “top-down” reasoning, and ignoring the fact that we often seem to reason “bottom-up” from some given evidence to a hypothesis that explains it. But inference to the best explanation, at least when naively interpreted, seems to err on the other side, ignoring the fact that we often look for evidence to support or refute a given hypothesis. As soon as we delve into the real world we see that we are really dealing with two aspects of the same process, where we are constantly going back and fro between evidence and hypothesis. Once we have a suggested explanation to whatever evidence we started from, we go on to investigate its consequences and further ramifications by “if so – then what?” reasoning. And as new evidence pops up we look for explanations that integrate it into the emerging total picture.

In fact, the real problem with the idea that research starts either from the bottom, with the evidence, or from the top, with a theory or hypothesis, is the notion that it “starts” at all. The search for knowledge is always an ongoing process, where evidence is sought and explained, and hypotheses are suggested and tested, on the basis of a background of given beliefs that comprises not only the received wisdom of a given discipline, but the rest of science and common sense as well.\footnote{This is the sort of point that philosophers of science are after when they heap scorn on “inductivism”, as implying the hopeless notion that the knowledge process can ever start from scratch. But, of course, “deductivism” can be given a similarly hopeless interpretation. Dagfinn Føllesdal (1994) has argued that the hypothetico-deductive method involves something like the “hermeneutic-circle” of classical hermeneutics and, given a broad enough understanding of the different kinds of presuppositions involved, that seems to be a good analogy.\footnote{We have become so accustomed to relativistic ways of speaking, that someone may perhaps be tempted to question this platitude. Isn’t it generally accepted, for example, that there may be several incompatible but equally good interpretations of a literary text? We will come back to such issues in later chapters, and treat them in more detail, but for the moment I will just note that purported examples of such incompatibilities generally fall in one of two classes. Either the incompatibility itself is spurious, and it is just a question of different answers being relevant to different questions. Or the valuation “equally good” is just epistemic, referring us to the obvious fact that sometimes we have no compelling reason to prefer one hypothesis to another which contradicts it. But the fact that two opinions are equally likely to be true, given the available evidence, is no reason why we should believe both of them.}}

With this we have introduced one of the virtues usually associated with good explanations: \textit{coherence}. It is not easy to pin down exactly what coherence is, but it has to do with how tightly a given hypothesis is integrated into a body of other opinions. A lower bound to coherence is \textit{consistency}, which means that the given hypothesis should not contradict what we otherwise believe – two contradictory opinions cannot both be true, and so a choice has to be made between them.\footnote{We have become so accustomed to relativistic ways of speaking, that someone may perhaps be tempted to question this platitude. Isn’t it generally accepted, for example, that there may be several incompatible but equally good interpretations of a literary text? We will come back to such issues in later chapters, and treat them in more detail, but for the moment I will just note that purported examples of such incompatibilities generally fall in one of two classes. Either the incompatibility itself is spurious, and it is just a question of different answers being relevant to different questions. Or the valuation “equally good” is just epistemic, referring us to the obvious fact that sometimes we have no compelling reason to prefer one hypothesis to another which contradicts it. But the fact that two opinions are equally likely to be true, given the available evidence, is no reason why we should believe both of them.}
CHAPTER 2. QUESTIONS, ANSWERS, EVIDENCE

2.7 Probability – Thomas Bayes’ theory of evidence

Empirical knowledge is never certain. No matter what amount of positive evidence we collect, we can never strictly prove that a hypothesis is correct – the possibility always remains of new data that contradict the hypothesis, or of a new hypothesis that gives a better explanation of the available evidence. Data cannot even prove a hypothesis to be false – maybe the problem is with some auxiliary hypothesis, perhaps with some hidden assumption that we haven’t even noticed. But if certainty is unattainable, maybe we should lower our standards and strive for some easier goal, and just try to make our theories as credible or probable as possible? Can we use the concept of probability to say something useful about testing a hypothesis?

The concept of probability is rooted in common sense, but it is also anchored in a mathematical theory, with wide-ranging applications across the sciences. The suggestion to use this theory to shape inductive inferences is certainly attractive: just think that one might use the calculus of probability to give an exact measure for the probability of a hypothesis given a certain body of evidence, or, at least, for how that probability is affected by “new” data.

There have been many attempts to use the calculus of probability as a basis for a theory of empirical evidence, but here we will concentrate on an approach that has become increasingly influential in recent decades, although it has its roots in the 18th century. The approach usually goes under the name “bayesianism”, after the British clergyman and mathematician Thomas Bayes (1702-61), who was the first to formulate some of its basic ideas. The main battles for and against bayesianism has been fought in the theory of statistical inference, and concrete applications of it can be mathematically heavy, but the basic ideas are intuitive and do not require mathematics beyond what every newspaper reader is supposed to know.

2.7.1 Prediction

Let us go back to the basic idea that evidence has to do with prediction: a piece of evidence supports a hypothesis if it is predicted by it. This is the idea that the hypothetico-deductive method tried to explain by means of deduction: the evidence supports the hypothesis if it follows logically from it (together with suitable auxiliary assumptions). How can we think about prediction in terms of probability instead?

The first thought that comes to mind is presumably that an hypothesis predicts some evidence if it is probable that the evidence would occur if the hypothesis were true. This sounds like a weaker version of the hypothetico-deductive method, which demands that it shall be certain that the evidence occurs if the hypothesis (and the auxiliary assumptions) is true. But as we

our evidence, does not make them equally true.

13Unlike other theories of empirical evidence Bayesianism has even become the topic of popular books aiming for the best-seller market. McGrayne (2011) gives an enthusiastic historical survey of the ups and downs of bayesianism, while Silver (2012) delivers a richly exemplified plea for bayesian thinking in all walks of life.

14Some readers may wonder what “probability” means in this context. That is indeed a good question which we will come back to in a little while, but for the moment I hope that you are willing to sail along on whatever pre-understanding you have.
saw, this idea runs into difficulties when there are rival hypotheses that predict the same evidence — rivals that make the evidence as probable or even more probable. The obvious solution to this problem seems to be that a hypothesis is supported by some evidence if it makes that evidence *more probable than it would otherwise have been*.

To explain what this means involves some complications, but let us first get the feel of it by means of an example. There is a neurological condition that used to be called *paralysie générale* — sometimes suggested as the diagnosis of Nietzsche’s madness. It is caused by having an untreated syphilis infection, and usually occurs 10-20 years after the original infection. It does not, however, affect all the relevant patients: the chance of developing it is about 25%. Is the occurrence of *paralysie* in a patient good evidence for the hypothesis that the relevant person has been infected by syphilis a decade or two earlier? The answer is surely yes, but the hypothesis does not make the evidence probable in an absolute sense — even if the infection has occurred it is more probable that the patient will escape the condition than develop it. But the relevant factor is not the absolute value of the probability, but the fact that the evidence is much *less* probable if the hypothesis is false — in this case it is in fact impossible.

So, if we want to evaluate what some evidence tells us about a certain hypothesis, we have at least two probabilities to ponder and to compare. First, we need to know how probable the evidence is if the hypothesis is true, and, second, we need to know how probable the evidence is if the hypothesis is false. In the terminology of probability theory such probabilities are called *conditional* probabilities: the conditional probability that a patient shall develop *paralysie general* given that he has an untreated syphilis infection is approximately 25% or 0.25.\(^1\)

Let us take another example: confession to a crime. Suppose that we have a murder trial and that the accused admits, during the interrogation, to having killed the victim. Is that good evidence that he or she is in fact guilty? It is certainly not conclusive evidence: we all know that people sometimes confess to crimes they have not committed. But even so, it is much more probable that a guilty person will confess to such a serious crime than that an innocent person will do it, and so the confession does speak rather strongly for guilt. Again, it is not the absolute probabilities that matter. Maybe the chance of obtaining a confession if the accused is guilty is only 10%, but it is still good evidence, if the chance that a guilty person shall confess is much smaller, say 0.1% or one in a thousand.\(^2\)

In the same way we understand why the denial of a crime is rather weak evidence for innocence. We expect almost all innocent persons to deny having committed a murder, but, alas, we know that many guilty persons would also deny it. Perhaps 99.9% of all innocents are deniers, but if 90% of the guilty would also deny, a denial does not tell us much one way or the other.

I said that we need to *compare* the conditional probability of the evidence if hypothesis is true with the probability of the evidence if the hy-

---

\(^1\)Reference for the example.

\(^2\)This is why circumstances that give innocent people strong reasons to confess, like torture or the offer to escape punishment, destroys the evidentiary value of a confession.
CHAPTER 2. QUESTIONS, ANSWERS, EVIDENCE

63

...ypothesis is false. If the hypothesis is that someone is guilty and the evidence is the confession, we compare the probability of a confession by a guilty person with the probability of a confession by someone who is innocent. But how do we compare it?

One suggestion might be that we should look at the difference between the two probabilities, but the contrasting examples of the evidentiary value of confession and denial make it clear that this cannot be right. The difference between 99.9 and 90 is the same as the difference between 10 and 0.1, namely 9.9. But it is intuitively obvious that if the figures in the examples were right the evidentiary value of a confession would be vastly greater than the value of a denial. The relevant relation is not the difference between the probabilities but the ratio between them, how many times one goes in the other. The probability of a confession from a guilty person, according to the example, is a hundred times greater than the probability of a denial from someone innocent, while the probability of a denial from someone guilty is only 1.1 times bigger than the probability of a denial from someone innocent.

2.7.2 Likelihoods and evidentiary value*

So far I have kept numbers and formulas to a minimum, and I will go on in the same way in the next section. But it is convenient, for those who dare make the attempt, to summarize the conclusions so far with the help of some standard mathematical abbreviations.

First some ways to talk about probability. We express the probability that a certain hypothesis \( h \) is true by means of the formula \( \Pr(h) \). For example, if the hypothesis is that the next throw of a certain fair coin will be heads, \( \Pr(h) = 0.5 \). The conditional probability that one proposition \( a \) is true given that something else \( b \) is true is expressed as \( \Pr(a|b) \). For example if \( a \) is the proposition that a certain throw of a six-faced die will be a six, and \( b \) is the proposition that it will be an even number, then \( \Pr(a) = 1/6 \) and \( \Pr(a|b) = 1/3 \).

In this terminology \( \Pr(e|h) \) expresses the probability of the evidence given that the hypothesis is true, and \( \Pr(e|\neg h) \) means the probability of the evidence given that the hypothesis is false. The conditional probabilities in the confession example thus become \( \Pr(\text{confession}|\text{guilt}) = 0.1 \) and \( \Pr(\text{confession}|\text{innocence}) = 0.001 \). The evidentiary value of the confession, according to the conclusion of the last section, is the ratio of these two numbers: \( 0.1 / 0.001 = 100 \).

In a standard, albeit a bit confusing, terminology \( \Pr(e|h) \) is often called the likelihood of \( h \), while \( \Pr(e|\neg h) \) is the likelihood of \( \neg h \). The ratio between them, as in the last paragraph, is the likelihood ratio, and this is often, as explained above, taken as a measure of the value of \( e \) as evidence for \( h \).

2.7.3 Prior probabilities

The likelihood ratio is a nice measure of how strongly a piece of evidence supports a certain hypothesis, but it does not in itself tell us if we should expect the hypothesis to be true, or even if we should actually expect the
evidence to occur. To see why, let us go back to the confession example once again.

It is well known that the police in many investigations, especially with high profile cases, get a lot of confessions that they routinely dismiss. How can that be right, if the evidentiary value of a confession is so high? The key is that even a low probability of a false confession will give us many “positive” cases if we put the question to enough people. Let us continue with the assumption that one in a thousand of innocent people will confess to a certain crime. Let the question be put, by way of the media, to each of a million people whether he or she is the one who did it, and you should expect about a thousand people to confess! If there is only one murderer, at least 999 of them are innocent. Based only on the evidence of the confession the probability that any specific confessor is guilty is very low, and the police are right to rather summarily dismiss most of them from the investigation.

So were we wrong to ascribe such a high evidentiary value to the confession? No, but we must think some more about what it means to have a high evidentiary value. Positive evidence raises the probability of a hypothesis from what it was before we acquired the evidence, but the result does not only depend on the raise but also on where we begin. Suppose that the murderer actually is one of the million people who reads about the murder in the papers. Based only on that information the probability that any specific person is the murderer is one in a million. If someone confesses, the probability for him or her goes up to one in ten thousand – considerably higher than before but not much to base a case for the prosecution on.

Suppose, on the other hand, that the police has already narrowed the search to one of three equally likely suspects and one of them confesses. Under these conditions the chance of obtaining a false confession is very small in relation to a true one, and the chance that the confessor is guilty rises from 33% to 98%.\(^{17}\)

So, in addition to the likelihood of the hypothesis and its alternatives, we have one more background probability to take into account: the “prior” probability of the hypothesis itself. It takes a lot more evidence to justify us to believe in something which is in itself very improbable than we need to believe in something antecedently plausible.\(^{18}\) The prior probability that someone picked at random from a million people shall have committed a particular murder is very low, while the prior probability that someone on the police short list will be the guilty one is usually quite high.

Where do prior probabilities come from, and how do we know how probable an hypothesis is before we have tested it? In most cases, we can think about prior probabilities as another way to express the demand that we talked about above, that a hypothesis must always be judged against all the available evidence. A confession – to continue with the same example – adds new evidence to the hypothesis that someone is guilty, but we always have other relevant considerations as well. Assessing the prior probability sums up these other considerations, and bayesian reasoning will help us to

\(^{17}\) We will soon learn how to make these calculations.

\(^{18}\) David Hume used this idea in a famous argument that it is never rational to believe on the basis of testimony that a miracle has occurred. By definition, a miracle is a maximally improbable event, and so it is always more probable that the testimony is wrong than that the miracle has happened.
update that probability in the light of the “new” evidence.

We will soon come back to the actual structure of bayesian reasoning, and also to the problem of assigning prior probabilities, but first we will try to make things clearer with the help of another simple example.

2.7.4 Testing for a disease

Suppose that I worry about having a certain disease, for which there is a rather simple test. The condition is not very common, but the death rate is high and it afflicts something like 5 out of 1000 men in my age group. I go to my physician and ask to be given the test, which is accepted. The test is not infallible, of course, but reasonably reliable. Let us assume that it finds all real cases: if I have the disease I will test “positive”. But in the inverse case there is a margin of error, the reliability in that direction is, say, 0.95 – there is a 5% risk that I will test positive even if I am perfectly healthy.

I take the test and nervously await the outcome, and finally learn that my fears have come true: the result is positive. The test result is my evidence, my hypothesis is that I have the disease. How probable is the hypothesis in light of the evidence? Experience tells that most people will make a guess in the vicinity of 95% – they think that the test shows the hypothesis to be overwhelmingly probable. But this is decidedly wrong: the real probability is under 10%.

This result can be calculated with a simple formula, that we shall soon mention by name, but it can also be reached by common sense reasoning. Suppose that we let a thousand randomly chosen men of my age take the test. About five of them will have the disease and will test positive on account of this. But, because of the margin of error, the same thing will happen to approximately 5% of the healthy men as well – i.e. to roughly 50 persons. There is, of course, a much larger chance that I belong to the latter group than to the former.

Let us reflect a moment on the information that we used to reach this verdict. The hypothesis, in this case, is that I have the disease, and the evidence is the positive outcome of the test. We had to know the answer to three different questions:

1. How probable is the evidence, given that the hypothesis is correct? In the example this probability is 1, or 100%.
2. How probable is the evidence, given that the hypothesis is false? In the example this probability is 0.05, or 5%.
3. How probable is the hypothesis before the evidence, i.e., before I take the test? In the example this probability is 0.005, or 0.5%.

Hopefully, we recognize these probabilities from the previous section. The first two are the “likelihoods” of the hypothesis and its negation in relation to the evidence, while the third one is the prior probability of the hypothesis.

Why does almost everybody give the wrong answer to the question in the example? The mistake is that one forgets about the prior probability, and only concentrates on the evidence. This is a very common and important
fallacy, sometimes called the “fallacy of ignoring the base rate”, and it does not only afflict ordinary folks but also persons who through their education should know better, like doctors, judges and scientists.

It is the role of the prior probability that motivates the hesitation to use many medical tests on a broad scale, to “screen” large sectors of the population for a specific form of cancer, for example. When the prior probability of having the disease is low, it will, so to speak, be covered by the margin of error of the test procedure, and the result will be correspondingly useless. The very same test that is of little value for the purpose of a general screening may very well be an excellent diagnostic tool, however, for specific cases where the prior probability is higher, for previously known risk groups. Suppose that the doctor that orders the test for me has other strong indications that I have the relevant disease, so that the prior probability in my case is much larger, say 50%. In that case a “positive” outcome of the same test would, indeed, indicate a 95% risk of me actually being ill – as can be verified by the same kind of common sense reasoning as above.\footnote{The example also suggests another description of the fallacy of ignoring the prior probability: namely, that we have a tendency to presuppose that the prior probability of any hypothesis is fifty-fifty.}

A similar fallacy afflicts our intuitive evaluations of the “likelihoods”, where it is easy to be blinded by a high positive likelihood, i.e., by the probability of the evidence given the hypothesis, and forget to ask how probable the evidence is if the hypothesis is false. Think about a watch that is stuck at a quarter past noon. The probability that it will show 12.15 if that is indeed the time is 100%, but that fact alone does not make the watch a reliable indicator of this point in time – the reason, of course, being that it would just as surely give the same verdict if the hypothesis were false. The same point is obvious in the case of testimony. That a certain witness claims to know that the accused is innocent, is of little value if there is strong reason to suspect that the same witness would say the same thing if the accused were actually guilty.

2.7.5 Bayes’ theorem*

It is time to give a little more structure to these reflections, in order to discuss the possibility of generalizing them to other kinds of knowledge and evidence. We want to know the posterior probability of a hypothesis, i.e., the probability of the hypothesis given the evidence. In line with the example, this probability varies positively with the prior probability – a higher probability before the evidence gives, other things being equal, a higher probability after the evidence. It also correlates positively with the positive likelihood: if the hypothesis renders the evidence more probable, the evidence gives more support to the hypothesis. And it varies negatively with the inverse likelihood: if the evidence is very probable even if the hypothesis is false, the occurrence of the evidence lends correspondingly less support to the hypothesis.

The formula of the probability calculus that I alluded to above, expresses precisely these correlations. It goes by the name of “Bayes’ theorem”, after the British clergyman and mathematician Thomas Bayes (1702-61), who was the first both to formulate it and to apply it to the problem of empirical use.
The suggestion that one can use a simple mathematical formula to calculate how strongly a certain piece of evidence supports a certain hypothesis is very appealing, of course, but it is also very problematic, and the basic problems have to do with the very notion of probability, which is the subject of the next section.

We have already introduced notation for the prior probability of a hypothesis $Pr(h)$, and for the conditional probabilities that we call the likelihoods of the hypothesis and its negation, with regard to the evidence, $Pr(e|h)$ and $Pr(e|\neg h)$. When using Bayes’s theorem we are supposed to have all of these as already given, and the problem that the formula will solve for us is how to calculate the “inverse” conditional probability of the hypothesis given the evidence: $Pr(h|e)$.

The idea that we started from in this whole discussion was that a hypothesis is supported by some evidence to the extent that it makes that evidence more probable than it would otherwise have been. In its simplest form this is almost exactly what Bayes’s theorem expresses. Here it is:

$$ Pr(h|e) = \frac{Pr(e|h)Pr(h)}{Pr(e)} $$

What it says is that the probability of the hypothesis given the evidence is a ratio between two numbers, where the denominator is just the overall probability that the evidence should occur and the numerator is the probability that the evidence should come about by way of the hypothesis being true.\(^{20}\)

In the very simplest case, like the one involving syphilis and paralysis, these two numbers are equal because the evidence can only come about if the hypothesis is true. In that case $Pr(h|e)$ is simply equal to 1, and the evidence would conclusively prove the hypothesis.

In normal cases, however, the situation becomes a little more complicated, because there is some chance that the evidence should occur even if the hypothesis is not true, so that $Pr(e)$ is in fact the sum of the probability that the evidence occurs when the hypothesis is true and the probability that it occurs when the hypothesis is false. In this case

$$ Pr(e) = Pr(e|h)Pr(h) + Pr(e|\neg h)Pr(\neg h) $$

If we plug this complication into Bayes’s formula we get another and slightly more informative version of it:

$$ Pr(h|e) = \frac{Pr(e|h)Pr(h)}{Pr(e|h)Pr(h) + Pr(e|\neg h)Pr(\neg h)} $$

Again, it looks more forbidding than it actually is. The key to understanding how it works is to see that it has a very simple overall form, namely just like this:

$$ Pr(h|e) = \frac{A}{A + B} $$

\(^{20}\)It is the probability of the evidence given the hypothesis, "weighted" by the prior probability of the hypothesis. In slightly different words, the first factor gives the chance that the evidence shall occur in case the hypothesis is true, and the second factor gives the chance that this case shall be real.
Here A is the probability that the evidence comes about by way of the hypothesis being true and B is the probability that it comes about in any other way. Obviously, the smaller B is in relation to A, the closer to 1 the whole ratio becomes, i.e., the higher the "posterior" probability of the hypothesis, its probability "after" the evidence, will be.

So, a bigger value for A and a smaller value for B makes the posterior probability of the hypothesis larger. How does A get bigger? By increasing one or both of its components: the likelihood (the conditional probability of the evidence given the hypothesis) and the prior probability of the hypothesis. How does B get smaller? By decreasing its corresponding components.

This concludes our little detour into formalities but before we go it may be useful to have a look at how the machinery works with numbers instead of symbols, and check that the result is as we would expect. So, let us stick the numbers from the medical test example into the formula. Here they are:

\[
\begin{align*}
Ph &= 0.005 \text{ (five men in a thousand have the disease)} \\
P\neg h &= 0.995 \text{ (995 men in a thousand are healthy)} \\
P(e|h) &= 1 \text{ (there is no negative error)} \\
P(e|\neg h) &= 0.05 \text{ (the positive error is 5%)}
\end{align*}
\]

And the result is just what it should be, according to our previous reasoning and counting of cases:

\[
0.09 \approx \frac{0.005}{0.005 + (0.05 \times 0.995)}
\]

Again, to repeat the lesson: even if the probability of testing positive is much higher if you have the disease than if you do not have it, it is still much more probable that the positive result is a false positive than a true positive, if the prior probability of having the disease is low enough.

### 2.7.6 What is probability?

The example of medical testing has some special features that makes it suitable to introduce Bayes’ model. First, we have, in many such cases, reasonable estimates of the probabilities that we need to get started: the prior probability of the hypothesis and the two likelihoods. We know how common different diseases are and, with time, we are often able to find out whether people who test negatively or positively on the relevant test actually have the disease in question. Second, we have in these cases a reasonably clear and useful idea of what it means for something to be more or less probable: it simply means that it is more or less common in a certain reference group or “population”. How large is the probability that I – this specific individual – have cancer of the prostate? In a statistical sense there is no such probability: I either have it or not. The probabilities that we used in the example were supposed to be frequencies of the disease in certain groups of men: of men in general, of men who show some kinds of symptom, of men who test negatively or positively on the relevant test. That Bayes’ model is applicable to such probabilities is incontestable, and is proved by the fact that one can reach the same results without it, by simply going
back and count cases as we did above. But is it applicable to hypotheses that are not themselves statistical or backed by some supporting statistical hypothesis?

Both in science and in daily life, we often speak of probabilities without any clear connection to frequencies in populations. We would probably (or probably not) have escaped the second World War if Hitler had died in the first. The bullet that killed Charles XII was probably (or probably not) fired from the Norwegian side. It is probable (or not so probable) that the US will have a female president in the present century. It is hard to see what such probabilities have to do with frequencies, as they deal with events that are evidently one-off. The relevant concept of probability is *epistemic* rather than statistical — it is not about prevalence but about *credibility*.

As soon as this is pointed out, we realize that it is always probability in the latter sense that we are really after when we discuss the testing of a hypothesis: how credible is a certain hypothesis in the light of the available evidence? It is really the same thing in the medical case, too. I want to know how plausible it is that I have cancer, as the ground for a decision whether to submit to unpleasant and painful further investigations, and ultimately to a costly, time-consuming and even more unpleasant treatment — and, of course, I want to know how credible it is that the relevant treatment will actually help.

In many cases there seems to be a simple connection between statistic and epistemic probability, between frequency and credibility. It just seems reasonable to adjust one’s beliefs to certain given frequencies. Is the coffee shop in building A open now? Well, it is 11 am on a Thursday, it is usually open at this hour, so it is reasonable to believe that it is open now as well. But in many cases there is no such underlying frequency to lean on. Can we still use Bayes’s model? On this, opinions fall apart.

To generalize the model we need a notion of probability that is not based on frequencies, and we need to show that this notion, too, behaves in a way that fits the mathematical probability calculus. On top of that we need to make it plausible that in practical situations we can estimate such probabilities with sufficient precision to make the model useful. The most important suggestion for such a concept of probability goes back to the British philosopher Frank Ramsey, and is often discussed under the heading of “subjective probability”. Ramsey starts from the observation that we seem to have different strengths of belief in different things: of some things we are completely certain, of other things we are less certain but still rather sure, further down the road there are things we are just leaning towards or weakly presume, and at the other end of the scale there are things that we doubt, or absolutely do not believe. Such degrees of belief is the basis for the concept of subjective probability.

Ramsey’s idea is that we can measure degrees of belief on the basis of

---

21This does not mean that such inferences are unproblematic. Apart from the fundamental question of what motivates the step from a frequency in a set of cases to a probability in the single case, there is always the case of what reference group is relevant in different situations. The frequency of cancer of the prostate is not the same for men in general, as for men over fifty, for men who have the disease in the family, and so on. But which group is relevant to the probability in my case? Ultimately we are all unique, but in a group of one there are no frequencies.
betting behavior. How strongly you believe something is shown by the odds you are willing to accept on a bet that your opinion is correct. By this criterion we can order degrees of belief on a scale from 0 to 1. Ramsey then supplies an ingenious argument to the effect that a rational person must strive to arrange the totality of her subjective probabilities so that they obey the mathematical theory of probability: if you do not it is always possible for a smart enough bookmaker to set up a bet that you would accept but will always lose, no matter the outcome.²²

Are subjective probabilities a reasonable basis for a theory of scientific evidence? On this question, opinions wildly diverge. On the one hand, one may think that such a basis is much too personal and “subjective” in a pejorative sense – should science not be built on something more settled and secure than the degree of belief of individual scientists, measured by what they are willing to bet on this or that? On the other hand, one can argue that, whether we like it or not, this is just the situation that we are in and have to manage: we start from a given set of opinions of different strengths, opinions whose origins we don’t know very much about, and we must decide how to modify our faith in those beliefs in the light of the new information that is always coming in, as a result of our investigations and continuing exposure to the world.

Here, as always, we will dodge the deeper philosophical problems, and only take it for granted that there is enough substance in our intuitive judgements of probability to let the bayesian perspective throw some light on how to reason about evidence in science.

2.7.7 Conclusion

So what do we learn from the bayesian model, about how to evaluate our theories in relation to our evidence? The mathematical apparatus seems to point us in the direction of a way to calculate a numerical value for the probability of a hypothesis in relation to a given body of evidence, but as soon as we leave the domain of statistical inference, in the proper sense, this prospect is rather obviously vain. The probability assessments that are feasible in the humanities are, at best, estimations in the style of something being very probable, rather probable or hardly probable at all, or that one possibility is more probable than another one, and so on. But this does not exclude that there are practical lessons to be drawn from bayesian thinking, even in the humanities.

The most important lesson is that we learn what factors to concentrate upon: the prior probability of the hypothesis and its alternatives, and the two likelihoods. How plausible is the hypothesis in itself (compared to the alternatives), apart from the current evidence? How probable is the evidence given that the hypothesis is true? How probable is the same evidence given that that hypothesis is false, i.e., given that some alternative hypothesis is correct?

As an illustration, let us cast a quick glance back to the death of Charles XII. As possible evidence for a Swedish killer, one might refer to a persistent rumor, in the Swedish army and later in Stockholm, that the king was

²²In the literature, such a bet goes by the name of a “Dutch book”, and so Ramsey’s argument is known as “the Dutch book argument”.
murdered. This rumor is usually taken to lack evidential value, however, on the ground that similar rumors went around at the death of Gustavus Adolphus, and will generally occur every time a famous war leader is killed in battle. Why is that relevant? The implication obviously is that even if the rumor might be true on this occasion, it has scant evidential value because it would probably arise even if it were false – a high likelihood of the evidence on the assumption that the hypothesis is false nullifies its value as evidence for the hypothesis. On similar grounds one gives scarce importance to negative testimony from persons that may be implicated in a possible murder – if the hypothesis were true they would presumably have denied it anyway.

An important part of the argument would concern the prior probability of the hypothesis of a murder: to make it large enough to compensate for the rather weak direct evidence, one would have to point to a convincing motive and a plausible killer – issues that have drawn considerable attention from pro murder advocates. Correspondingly, someone who wants to deny the murder hypothesis must argue that there is a high probability for alternative causes of death. The only plausible alternative is that the king was hit from the Norwegian fortifications, and so one must show that there was a significant risk to be hit by “a honest enemy bullet” – which explains the relevance of the circumstantial information that there were seven soldiers killed and fifteen wounded during four hours that same night, while digging the trench from which the king was watching.

Finally, how does the bayesian model relate to our earlier two approaches: does it give other advice, or does it say essentially the same things?

Let us begin with the hypothetico-deductive method. In the present terminology, it seems to give the more limited advice to concentrate entirely on the positive likelihood, and even on the special case where the probability of the evidence on the assumption of the hypothesis is 100% – the model demands that the evidence shall be a deductive consequence of the hypothesis. The first limitation is real. As we pointed out in the discussion of the hypothetico-deductive model, it ignores the importance of alternative hypotheses, i.e., of alternative explanations of the evidence, and we may now add that it ignores the prior plausibility of the hypothesis, as well. The other limitation is more verbal. The demand of the HD-method is not that the evidence shall follow deductively from the hypothesis alone, but only in conjunction with suitable auxiliary assumptions. To the extent that these assumptions have varying degrees of probability, the relevant uncertainty will be inherited by the conclusion.

In fact, the deductive apparatus that we introduced in connection with the hypothetico-deductive method is just as useful in connection with bayesianism. To ask for the positive likelihood is to ask what would be the case with the evidence if the hypothesis were true, and a good way to

23 Perhaps this may be argued to be implicit in the demand to take all the relevant evidence into account, which would include all of the past experience of mankind – summing up this background knowledge in a prior probability seems to be a reasonable way to think about it.
start on this question is to ask what auxiliary assumptions one would need in order to deduce the evidence from the hypothesis, and what degrees of credibility these assumptions have.

In a similar way, the emphasis on the competition between different possible explanations, which is a key strength of the inference to the best explanation perspective, is built into the bayesian approach: when the prior probability of the alternatives shrinks or grows, the probability (both prior and posterior) of the main hypothesis grows or shrinks, correspondingly.

So, what shall we finally say about our three theories? Which one shall we choose? For our purposes, and at the level where we are approaching the discussion, the best option is not to choose at all – even the question if they really contradict each other, and if so on what points, would burst the limits of this presentation. Instead, I will close this chapter in the same ecumenical spirit in which I have conducted it throughout, and point to one question from each perspective that any researcher should ask about the hypothesis he or she finally decides to embrace:

1. What, in precise detail, is the connection between my hypothesis and the evidence that I present?
2. Are there other possible explanations of this evidence, and, in that case, what is it that makes my explanation the best?
3. How likely is it that the evidence would occur even if my hypothesis were false?
Chapter 3

Meaning

According to a traditional and still popular view about the distinction between the cultural sciences and the natural sciences, the boundary line should be drawn by means of the concepts of meaning, understanding and interpretation. The cultural sciences stand apart with regard to their objects – meaningful phenomena – and with regard to the cognitive abilities and procedures that are required to access them: understanding and interpretation. Other words may be used to draw the contrast – opposing hermeneutics to positivism, or qualitative to quantitative methods – but the basic point remains the same.

Although popular, this view has never been unopposed. The separatists has always been confronted by unionists – people who either deny that there is any essential distinction between the natural and the cultural disciplines, or at least that the distinction can be drawn in terms of meaning and understanding. Traditional unionists come from the positivist camp and are nature imperialists: they deny that there is a specific domain of meaning which the natural sciences cannot grasp. Meaning categories should either be reduced to natural categories or abandoned as unscientific. Nature devours culture. Recent unionists, on the other hand, have tended to be culture imperialists instead: meaning is ubiquitous, and natural science relies just as heavily on understanding and interpretation as do the social sciences and the humanities. Culture swallows nature.

Whatever the case may be, with regard to the problem of demarcation between the natural and the cultural sciences, it seems prima facie obvious that meaning and interpretation play a dominant role in the research activities of cultural scientists, so that it would be a good thing to have a clear idea of what meaning is and what interpretation does. Poring over books, transcriptions, or other remnants of human signifying activities, while trying to figure out what they mean, is more or less what scholars do with their time. In some subjects, most obviously in the aesthetic disciplines, such remnants are the principal objects of study, and the ultimate result of research is often, precisely, an interpretation. In other disciplines, perhaps with history as a paradigm, they are used as sources for the knowledge of other things – the historian may well have to interpret a letter or an agreement found in an archive, but her ultimate interest is not in what the document itself means, but in what it reveals about historical circumstances external to it.
CHAPTER 3. MEANING

3.1 Concepts of meaning

As usual we have to be wary of the fact that words are used in different ways, with different meanings – the words “meaning” and “interpretation” are no exceptions. To the contrary: they seem to have got much more than their fair share of ambiguity and conceptual confusion. There are concepts of meaning and interpretation that are obviously useless to demarcate the cultural sciences, either because they are clearly too wide and applicable to the study of purely natural phenomena as well, or because they are clearly too narrow, and only applicable to some cultural phenomena, for example those directly involving language. The challenge for the Verstehen enthusiast is to find a reasonable sense of meaning and interpretation that has just the right field of application, that covers the domain of the cultural sciences and stretches no further. And if we live, as Leibniz thought, in the best of all possible worlds, that concept should then be capable of further differentiation, to give rise to the narrower concepts of meaning and interpretation relevant to different subfields of the cultural sciences, allowing for example linguistic meaning to be seen as a special case of a more general concept of meaning.

It is not obvious, of course, that there is such a unified concept of meaning, that keeps the cultural sciences together and apart from the rest. It might, for example, be the case that each part of the social sciences and the humanities deals with “meaningful” phenomena, but that the relevant concepts of meaning are so wildly different that the impression of a similarity is spurious – more or less as if one expected a deep connection between riversides and certain financial institutions just because they are both called “banks”. Nevertheless, I will argue that there is a unified notion of meaning that roughly demarcates the cultural sciences, and that it is illuminating to see the domain of the cultural sciences in the light of this notion, so that the Verstehen tradition really has a point.

Before we come to that, however, I will devote a couple of sections to some other notions of meaning and interpretation that do not match the prescription, that are not unique to the cultural sciences. In the first case the main reason is to get it out of the way, so that it won’t be insinuating itself into the discussion later on. The second case, however, has independent interest for us, and we need to have it along for the ride.

3.2 Facts and interpretations

The uses of the word interpretation that interest us are the ones that correlate with a notion of meaning, where to interpret something is to try to find out what it means, and where an interpretation is the result of such an undertaking. But there is at least one common usage where there seems to be no such correlation, namely the one where an interpretation is contrasted with a fact. “Is that a fact, or is it just your interpretation?” Here, the relevant contrast is not between something that has meaning and the meaning that it has, but between something that is taken to be known, and something that is less well known, that is still up for grabs, as it where. The dividing line between facts and interpretations, in this sense, in not
stable and rooted in the nature of things, but moves with our knowledge: what was once an interpretation becomes a fact as more evidence comes in.

The important thing to notice about this sense of interpretation is that it is more or less built into it that the competition between rival interpretations is not decided by the evidence at hand, that opting for one interpretation over another involves a degree of choice, that it has an element of the arbitrary. Once the evidence is clear, the need for interpretation is over, and we are left with the facts.

The reason that it is important to note this feature of this notion of interpretation is that it is otherwise easy to subconsciously bring it along to interpretation in other meanings of the word, and mistakenly take it for granted that interpretation always involves an element of arbitrary or subjective choice between epistemically equivalent alternatives. It may, of course, be true that there are examples of undecidable competition between rival interpretations in some cases where interpretation has to do with finding out what something means, but this must then be argued for in each case, not just be taken for granted because we use the word interpretation about them.

3.3 Natural meaning

When talking about hypotheses and testing we used the detective on a murder case as a metaphor. Clearly, the detective does a lot of interpreting and trying to find out what things mean, in a certain sense. Lipstick on the rim of a glass indicates that a woman has used it, big footprints under the window mean that a big man has stood there, and so on – the standard fare of crime stories. The same sort of reasoning may be applied where there are no human agents present at all. The dark cloud means that it is going to rain, a certain pattern in the spectrum of the light emitted by a distant star means that it contains a lot of helium, and so on.

This sort of meaning is often called natural meaning, and the things that get interpreted in this way are called natural signs. In this sense, we can say that part of the debate about the death of Charles XII, that we discussed earlier, is about the interpretation of the holes in his skull: do they mean that he was shot from the back or from the front? The basic relation that we use to travel from natural signs to their meaning is causality. We reason backwards from the lipstick on the glass to the woman who put it there, we reason forwards from the present clouds to the future rain, and sometimes we do both as when we take the fall of a barometer to indicate a future storm, by reasoning backwards to the decrease in atmospheric pressure, and forwards from that decrease to the storm.

It is obvious that this is not the sort of meaning that anyone would think distinguishes the cultural sciences. One might even think that it is specific to the natural sciences, which are sometimes thought to have a sort

\footnote{Perhaps apologies are in order for the somewhat dated character of the lipstick example – the improved quality of lipstick and the changing distribution of its employment surely means that its function as a clue in whodunits is on the wane. That traces of lipstick in some historical contexts can function as natural signs of the presence of women does not imply, of course, neither that it is natural for women to use make-up, nor that someone who reasons in the indicated manner cannot be mistaken.}
of monopoly on the notion of causality. That would be a big mistake, however. No domain of empirical knowledge can do without causal reasoning, and the interpretation of natural signs, in the indicated sense, permeates every aspect of human cognition.

One interesting case is where natural meanings attach to signs that also have meanings of other kinds, as for example linguistic utterances. When I speak English it is immediately obvious to most listeners that I am not a native English speaker, and many would also guess that my native language is Swedish. My speech indicates my ethnic, geographical and linguistic background, as well as many other facts about me: that I am a man, my approximate age, whether I am nervous or angry, and so on. I do not have to say that I am Swedish for someone to know this from my speech.

Interpretation of linguistic utterances for what they indicate in this way, for the information that they carry behind the back, so to speak, of the person speaking and writing, plays a very important role in many humanistic disciplines. A good example is author attribution and dating of doubtful texts by means of stylometry — the small scale linguistic habits of different individuals give rise to a sort of linguistic fingerprint that can be used to identify the author of a work. When a historian reads a text not for its message but for what it reveals about the language or the mentality of an epoch, it is the same mechanism at work.

In classical historical methodology one used to classify evidence in two large groups: the written sources and the remnants. Written sources were treated as testimony and were evaluated with regard to their reliability. One was supposed to read a document for what it said, and then to judge whether the author was truthful, well-informed and generally trustworthy, or, conversely, biased or ignorant or otherwise suspect. Remnants, on the other hand, were taken as natural signs, and interpreted along causal lines — basically by means of inferences to the best explanation. We find roughly the same distinction in law courts, with the classification of evidence as testimony and forensic evidence, respectively.

A strong tendency in modern research is to treat documents not only as testimony but also as remnants, to interrogate them not only for what their author wanted to say but for the information they carry in other ways. It may even be argued that some celebrated modern theories of interpretation, like deconstruction, are mainly geared to treating the texts they focus on as symptoms rather than symbols, i.e., as a kind of natural signs of some of the circumstances involved in their production.2

A specific complication is that natural signs may be taken up by human agents and integrated into intentional “messages” of other kinds. A simple case is deceit, as in the planting of evidence. The lipstick may be smeared on the glass by a man wanting the police to think there was a woman present, and I may deliberately speak English with a French accent to make you think that I am French. With this, however, we have left the domain of natural signification, and we will return to this type of examples below, when we are ready for them.

2This way of thinking is very explicit in Foucault’s work, as witnessed for example by the title The Archaeology of Knowledge for his main methodological treatise (Foucault; 1969).
3.4 Meaning and subjectivity

What is the next step? What would be the contrasting kind of meaning, a kind of meaning that is not natural meaning? One sometimes sees the contrast drawn in terms of convention: natural meaning is distinguished from conventional meaning. Undoubtedly there is such a thing as conventional meaning, and we will come back to it in due time, but that notion is much too narrow for our present purposes. The natural/conventional dichotomy does not exhaust the field of meaning, and when it is said that the meaningful is the domain of the cultural sciences, conventional meanings cannot fill the bill.\(^3\)

One basic intuition concerning the notion of meaning that we are looking for is that it has to do with human cognition and agency. Roughly: if there were no people there would be no meaning either. A text in itself is nothing but black ink unevenly distributed over a white surface. A spoken sentence, in itself, is just a certain disturbance of the air produced by someone’s vocal apparatus. An oil painting is only color on canvas unless someone is there to see something in it.

In the rest of this chapter we are going to follow this lead and start from an elementary understanding of what human cognition and agency is, largely drawn from the phenomenological tradition in philosophy, and build a hierarchy of meaning phenomena on that basis, helped by insights from semioticians like Roland Barthes and speech-act theorists like Paul Grice. The result will, by and large, be an agent-action centered perspective on meaning. We will start, however, at an even more basic level, with an analysis of a basic trait of human experience: what phenomenologists call intentionality.

Before we do that, however, I want to sketch the place of the considerations of this chapter in an even larger whole. One common way to introduce meaning phenomena is through a communications model comprising at least three components: the sender, the receiver, and the language that they use. Competing theories about meaning in the humanities tend to take one of these as their base, and try to derive an understanding of the others from the chosen foundation: approaches based on speech act theory start from the sender and the activity of meaning something, hermeneutical approaches start from the receiver and the activity of interpretation, structuralist approaches start from the notion of a language or a symbolic system, conceived as a background that makes the activities of meaning and interpretation possible.

It goes without saying, that a complete theory of meaning and interpretation must have something to say about all three components, and about a lot of other things besides. So the first guess would be that such one-sided approaches will, at the end of the day, turn out to be complementary rather than competing. Basically, this is the working hypothesis that I will adopt, and after the chapter of meaning, there will be a chapter on interpretation,\(^3\)

\(^3\)Paul Grice, in a seminal study of communicative meaning that we will return to below (Grice; 1957), simply distinguishes natural meaning from non-natural meaning. In itself this would be a broad enough term, of course, but he then goes on to use it for a quite specific, albeit very important, kind of meaning, leaving vast regions of relevant phenomena outside the apparent dichotomy.
and after that a chapter on languages and the conceptual frameworks that they embody. Which is not to say that we will just ignore the usual polemics between the different approaches – even if all three aspects must be present in the final picture, there is ample room for disagreement about priorities, about what should be analyzed on the basis of what, and we will comment on such things as we go along.

3.5 Husserlian meaning: intentionality

The founder of the phenomenological tradition in philosophy is usually taken to be Edmund Husserl (1859-1938). Prominent later members of the same tradition are Martin Heidegger, Jean-Paul Sartre, Simone de Beauvoir and Maurice Merleau-Ponty – when discussing the methodology of the cultural sciences one should also mention Alfred Schutz. But if Husserl is the father of phenomenology, its grandfather is Franz Brentano, and it is from Brentano’s discussion of intentionality that we will start here.

In his book *Psychology from an Empirical Standpoint* (1879) Brentano wants to demarcate the subject matter of the emerging new science of psychology by means of a definition of what it is to be a mental phenomenon. He gives several different characterizations of the mental but the most important one is by means of the property of intentionality. Here are his own words:

Every mental phenomenon is characterized by what the medieval scholastics called the intentional (or mental) inexistence of an object, and that we, by means of not entirely unambiguous expressions, call the relation to a content, the direction towards an object (which is not understood as something real), or the immanent objectivity. Every mental phenomenon contains something as its object, albeit in different ways. In a representation something is represented, in a judgement something is affirmed or denied, in love loved, in hate hated, in desire desired, etc.

To imagine something, to see something, to believe something, to love or fear something, etc., are examples of what phenomenologists call intentional acts. That “something” towards which the act is directed is called the intentional object of the act. The basic idea is not difficult, even if the terminology may be daunting – when I see an apple tree I have a special kind of intentional act, namely a perceptual act with the tree as its intentional object.

---


5In fact, the whole terminology surrounding the notion of intentionality is bewildering. One should, for example, not confuse the intentions of intentional acts with the motives of intentional actions. An intentional act is not something that I do with the intention of achieving something, it is rather something that happens to me. To intend something in the action sense, is a specific kind of intentional act, like wishing or believing. To add insult to injury, we will soon learn that the intentional object is not an object either …
CHAPTER 3. MEANING

What does Brentano mean by the parenthetical remark that the intentional object is not to be understood as something real? He is referring to the fact that I can think of, and otherwise mentally relate to, things that do not exist. I can be afraid of ghosts, I can think about Pegasus, I can wish to meet Gandalf. How is that possible? I cannot shake the hand of Tintin or sit behind Donald Duck on the bus, but I can think about them. What odd type of relation must intentionality be, if I can have it to things that do not even exist?6

The plausible answer is that intentionality is not a relation at all, or at least not a relation between the act and what it is about. The act must so to speak be able to accomplish its directedness by itself, without any help from its object. How? The standard answer to this question is easiest introduced with the help of a related problem concerning linguistic expressions.

3.5.1 Words, things, concepts

What is the meaning of a word? Well, it depends, of course, on what kind of word we are talking about – the word “not” hardly works in the same way as the word “Paris”. But let us concentrate on names and other nouns, words that purport to “stand for” something.

The simplest theory about what the word “Socrates” means is that it just stands for a certain individual, i.e., Socrates himself – it is, one might say, a label that attaches to him. Analogously, the word “man” might be taken to stand for a certain group of individuals, namely all men. Saying that Socrates is a man, according to this theory, would be tantamount to saying that this individual belongs to that group.

But this theory runs into the same sort of problems that we just encountered regarding intentional acts. We readily understand and use names like “Donald Duck” and “Tintin” but there are no individuals for them to label. We understand the nouns “centaur” or “ghost” just like we understand “man” – but there are no groups of centaurs or ghosts for them to stand for.

The solution is to distinguish between the meaning of a word and its reference, between what we have to grasp about the word in order to understand it, and the thing or group of things that it stands for. “Socrates” and “man” have both a meaning and a reference, “Donald Duck” and “ghost” do not refer to anything but they still have meaning. And it is natural to think of the meaning as being more intimately connected than the reference to the word – it is the meaning of the word that decides what it refers to, and if it refers to anything at all.7

This way of thinking goes back at least to Plato, who argued that a word

---

6 Some may want to protest that Donald Duck does exist, namely as a figure in a cartoon. As a way of speaking this is quite OK, of course, but it should not be made to obscure the fact that Donald Duck, the creature that the cartoon purports to be about, does not exist. Donald Duck, as portrayed in the cartoon, is not supposed to be a cartoon but a duck, habitually dressed in a sailor shirt but no trousers, living in a city called Duckburg populated by other man-like animals speaking English to each other, having an enormously rich and greedy uncle named Scrooge McDuck, etc. Had Donald Duck been real there would actually have been someone of which these descriptions were true, but there is not. There are Donald Duck pictures, Donald Duck stories and Donald Duck movies – but there is no Donald Duck.

7 Everything I say here is very simplified, and there is in fact a huge amount of controversy and complication regarding every aspect of these seemingly simple notions.
like “man” primarily relates to an idea or a form that we grasp with our minds, rather than with the senses, and that it is only at second remove that the word is applicable to ordinary concrete people, because they “take part in” the idea of man. Ever since Plato it has been a subject of controversy what sort of things these ideas or forms or concepts, or whatever we like to call them, really are in themselves. Are they psychological phenomena, within the soul or mind of persons? Or do they have an objective existence independent of anybody’s mental life? Or are they perhaps social entities presupposing communication and a language that is shared among a group of people? We will come back to such questions later, but for now it is enough that we accept that we cannot explain how language works without invoking something “between” words and things in the world, a stratum of meanings or concepts, that you have to apprehend in order to use a language.

3.5.2 Concepts in perception and thinking

So concepts have a role in the explanation of language, but that is not all they do. They seem to be important to many other aspects of human cognition as well – we cannot even begin to explain things like perception and thinking without recourse to concepts.

A simple way to introduce the conceptual component in perception is by means of ambiguous pictures, like the duck-rabbit.

This is a picture of a duck – it has a sharp beak and it is looking to the left. This is a picture of a rabbit – it has long ears and it is looking to the right. Once you have seen both pictures it is not hard to go back and forth between them. In one sense of the word ‘see’ you see the same thing all the time – a certain pattern of black lines on paper – but in another sense you see two different things. A natural way to describe the situation is that you see the same lines, but you see them now as a duck and now as a rabbit. Where does the difference come from, where is the change? It is not in the lines themselves, nothing on the paper changes as I go between the duck and the rabbit. It seems that the change must be in me rather than in the picture, and it seems to involve my concept of a rabbit and my concept of a duck.

8More carefully put, you see them as a duck-picture or as a rabbit-picture. This qualification maybe important sometimes, if we are not to be bogged down in needless paradox. While it impossible to be a rabbit and a duck at the same time, the picture proves that it is quite possible to be a rabbit-picture and a duck-picture at the same time – although it may still be impossible to see both the rabbit and the duck in the picture at the same time.
see the lines and I apply different concepts to them: if I had the concept of a rabbit but not the concept of a duck, I would only be able to see the rabbit and not the duck, and vice versa. And it seems almost irresistible to invoke the notions of meaning and interpretation here: as I go between the duck and the rabbit I interpret the lines differently, the lines stay the same but their meaning changes.

According to Husserl there is a conceptual component in all perception, all perception is meaningful in roughly the indicated sense. To see (or hear, etc.) something is always to see it as something.9 And it is the same type of meaning-component – Husserl calls it a noema – that accounts for the intentional direction of intentional acts of every kind: to remember something, to hope for something, to fear something, is to have an intentional act of a certain kind which is directed in a certain way in virtue of a certain conceptual content, the noema. In many cases, of course, there will also be a real object fitting the noema.

One important consequence of this way of looking at things is that your intentional capacities – what you are able to think, perceive, and even do (as we shall see later) – depends on your conceptual resources: you can only see and think (in this sense) what you have a concept for.

3.5.3 The heaven of the Greeks

This way of thinking is what inspires present day unionists to argue that there is no pure nature for the natural sciences to describe, and that all knowledge is in a sense knowledge of culture. Step one, in this argument, is the thought that even our most direct contact with reality, in perception, is already shaped by our conceptual resources. Step two, is the notion that concepts themselves are cultural products, that our conceptual resources are given to us through language and culture and that they vary between different languages and cultures. Step three, the conclusion, is that what we experience as the world is, in fact, also a cultural product, and that people belonging to different cultures in an important sense live in different worlds.

We will come back to this argument in much more detail in a later chapter, but for now we just need to say a few things about its implications for the possibility of basing the distinction between the natural and the cultural sciences on the notion of understanding. As an example, I will use an argument by Thomas Kuhn.10

The question around which Kuhn’s discussion circles is, precisely, whether the phenomena of nature are given in the same way for all people, irrespective of their social, cultural and historical background, or whether these phenomena are, in some suitable sense, socially constituted, accessible only by way of some medium that may vary between cultures or historical epochs – the medium of a language or a conceptual system. His main

---

9 Actually, Husserl was not the first think about perception in this way. The basic idea comes from Kant and it was standard among the neo-Kantians of Husserl’s generation, but he develops it in such a rich and and fruitful way that I will continue to speak about it as his theory.

10 He has expressed similar ideas on many occasions, but I will refer mainly to the article “The Natural and the Human Sciences” from 1991, reprinted in (Kuhn; 2002).
example is what we see when we look up into the sky. Was the heaven of the Greeks different from ours?

It is clear that the Greeks classified what they saw in the sky differently from how we do it. For them, the sun and the moon belonged with Jupiter, Mars and Venus in the category of the planets; the Milky Way they took to belong in the category of “meteorological” phenomena, along with the rainbow and rings around the moon. According to Kuhn, this historical fact has deeper implications. He denies that there is something common and objective – the heavenly bodies – that we and the Greeks classify in different ways. They applied different concepts to what they saw in the sky, but that means, eo ipso, that they also saw different things, because “concepts shape the world to which they are applied”.

Using Husserl’s terminology we might rephrase Kuhn’s point as saying that the Greeks saw different things in the sky because their intentional acts were imbued with different noemas. From this he argues that even the natural sciences are in a sense “hermeneutical”, that they are sciences of meaning, just like the cultural sciences. Here is the reasoning:

The heavens of the Greeks were irreducibly different from ours. The nature of the difference is the same as that [Charles] Taylor so brilliantly describes between the social practices of different cultures. In both cases the differences is rooted in conceptual vocabulary. In neither can it be bridged by description in a brute data, behavioral vocabulary. And in the absence of a brute data vocabulary, any attempt to describe one set of practices in the conceptual vocabulary, the meaning system, used to express the other, can only do violence. That does not mean that one cannot, with sufficient patience and effort, discover the categories of another culture or of an earlier stage of one’s own. But it does indicate that discovery is required and that hermeneutic interpretation – whether by the anthropologist or the historian – is how such discovering is done (Kuhn; 2002, p 21).

As it stands this argument clearly misses its target. What Kuhn wants to argue is that the natural sciences, astronomy in the example, are hermeneutic disciplines. But the conclusion he ends up with is only that anthropology and the history of ideas are hermeneutic, and who would deny that? Astronomy is about what is in space, not about how the Greeks or anyone else (including ourselves) have thought about or perceived what is in space.

We can express it this way. Human cognition, in all its aspects, is intentional and, if the above reasoning is correct, this implies that it is also conceptual, in the sense that it must use concepts. But this does not imply that all human cognition is about concepts. The Greeks used their conceptual resources to describe and investigate what they saw in the sky, just as present day astronomers use our concepts for the same purpose, but in neither case are they investigating their own or anyone else’s concepts – they are investigating the world by means of their concepts.
3.5.4 Nested intentionality

When I see an apple tree, or wish for a bite of an apple, the respective objects (the tree or the bite) of my intentional acts (the seeing or the wishing) are not themselves intentional acts. But an intentional act can, of course, have another intentional act as its object – I can think about perceptions, beliefs and desires that I, rightly or wrongly, attribute to myself and others, I can wish that your desires were different from what they in fact are, and so on. I will say that an intentional act whose object is itself not intentional is (or exhibits) a *first-order intention* – for example, my belief that there is a cat in the basket. An intentional act that has a first-order intention as its object I will call a *second-order intention* – for example, your belief that I believe that there is a cat in the basket. Extending this usage in the obvious way, we can talk of intentions of the third, forth, fifth order, and so on, as long as we like. For example, if I *intend* to make you *believe* that I *believe* that you *believe* that there is a cat in the basket, that would be a 4th order intentional state. Because of the way it weaves different intentional acts together in sometimes very complex patterns, I will speak of higher order intentionality as *nested* intentionality.\(^{11}\)

The capacity for nested intentionality is very important to humans, and although there are traces of it in other higher animals, particularly in other primates, the sophisticated development of this capacity seems to be a special characteristic of our species.\(^ {12}\)

With this terminology we can finally express what distinguishes the cultural from the natural sciences, in terms of meaning. While Kuhn is right that all science involves the use of intentionality, only the cultural sciences involve higher order intentionality, in the sense that they have intentional phenomena as their object. In simpler terms: the natural sciences cannot do without human subjectivity, without thought and feeling, but only the cultural sciences are *about* human subjectivity – without implying, of course, that this is *all* that they are about.

As we shall see the notions of nested intentionality and orders of intention can also be used to illuminate and further differentiate among forms of meaning relevant to different fields within the cultural sciences.

3.6 Weberian meaning: action

One of the classical exponents of the view that understanding is a basic and distinguishing feature of the cultural sciences is the German sociologist Max Weber.\(^ {13}\) The primary object of sociology, according to Weber, is human action, and actions are meaningful phenomena that are only accessible to us through understanding and interpretation.

What is it to understand an action, to know what a person does? I see someone taking cover from the rain, under at tree. To apprehend that ac-

\(^{11}\)I don't know if he was the first to use the term, but I learned it from Ingvar Johansson (1989).

\(^{12}\)According to a widely accepted hypothesis, impediments to this capacity is the core phenomenon in autistic disorders.

\(^{13}\)The introductory sections of *Wirtschaft und Gesellschaft* (1922) give a good introduction to Weber's methodological views.
tion is partly to apprehend a purely “external” process: to see someone’s body being transported to a place under the branches of the tree. But it also involves something more: namely to see that piece of bodily behavior in the light of a certain desire (not to become wet) against the background of certain beliefs (it is raining, the tree will provide some protection). To apprehend an action is not only to apprehend a certain behavior, it is to understand something about how the actor herself apprehends her behavior as relating to her beliefs and desires.

In the present terminology we can say that understanding someone else’s action always involves an intentionality of (at least) the second order. To have an opinion about what someone does, ipso facto, to have an opinion about what he or she believes and desires – it presupposes that one, to some extent, views the situation through the eyes of the agent. To apprehend an action you not only has to see-it-as something, you have to see-it-as-seen-as something. This is what Weber has in mind when he says that the social sciences presupposes Verstehen, and this is what R. G. Collingwood means by saying that all history is the history of ideas. And this is what, pace Kuhn, sets the cultural sciences apart from the natural sciences – the astronomer must, of course, see the heavens through his own eyes, but he need not see them through the eyes of another. The astronomer’s description of the heavens presupposes only first-order intentionality, the historian’s description of the astronomy of the Greeks presupposes second-order intentionality.

To forestall a possible confusion, I should perhaps stress that what has to be second order with regard to an action is not the agent’s but the observer’s (the scientist’s) intentionality. Understanding an action presupposes second-order intentionality on the part of the observer; performing the action may in many cases get along with first-order intentionality – desiring and believing may be first-order intentional states.

The crucial point is that the description of an action depends upon the agent’s conceptual resources and world-view. It cannot be true of someone that he is chopping wood if he doesn’t know what chopping wood is, and it cannot be true of someone that she voted Labour in the last general election if she doesn’t know what it is to vote. And the conception of her own action that the action description implicitly ascribes to the agent is not an isolated atom, but involves large chunks of her world-view. The wood-chopper has to understand that this is an axe, that this is wood which will be fit to burn in a stove or a fireplace, if divided into suitable pieces; he must have a conception of what fire is and how it may be used for human purposes, like heating and cooking, and so on. The voter must have ideas concerning the political system of her society, concerning the connections between the act of voting, the outcome of the election and the wielding of political power.

We might say, as many have said, that the undertaking to describe and explain human actions makes the observer dependent upon the perspective of the agent. The natural sciences are, in this respect, more free in their choice of conceptual apparatus. Physics may shape its concepts so as to allow the simplest, most fruitful and elegant theoretical description of the universe – without asking the protons for their opinion. But in the cultural sciences we must, at least as a beginning, frame our descriptions in terms intelligible to the agents we study. This dependence is not total, of course.
Even if I have to *describe* the actions I want to explain in terms comprehensible, and perhaps even acceptable, to the agent – “A voted Labour,” “B consulted the poison oracle to find out if his neighbour is a witch” – my *explanation* of these actions may go far beyond, and even come into conflict with, the agent’s own views of the matter. And if I want to explain the actual *outcome or result* of the action, I must always go beyond the agents intention – obviously so when the actual result is not what was intended, but not even when the intention is realized in the intended way is the intention sufficient to explain what comes about. It is not sufficient that the wood-cutter *thinks* that the axe is sharp enough, for the wood to be cut – it must actually *be* sharp enough.

3.6.1 Thick and thin

June 28 1914, the archduke Ferdinand of Austria and his wife were killed, during a visit to Sarajevo, by shots fired by a young Bosnian Serb, Gavrilo Princip. The action performed by Princip at this occasion came to have world historic significance. But which action did he perform? Here are some suggestions:

1. He contracted the muscles of his index finger
2. He bent his finger
3. He fired the gun
4. He killed the Archduke
5. He created havoc in the Balkans
6. He started the first World War

What we have already said seems to be enough to exclude (1) and (6). One can only perform actions that one, in some suitable sense, *intends* to perform. To start the first World War seems to be beyond Princip’s possible intentions – it is not reasonable to credit him with a conception of this war that allows him to have this intention. To contract the muscles of his index finger seems, on the other hand, to come *before* his possible intentions – we normally do not have any conception of the various physiological processes that are involved in the bodily movements we perform. On other grounds, (5) seems to be less than reasonable as a specification of Princip’s action. He may, of course, have intended, by his action, to create havoc in the Balkans, but the actual coming to be of this havoc seems too indirect, and too dependent on other factors, to be included in the action itself – it would rather be counted among its intended effects.

But (2)–(4) all seem to be possible actions for Princip to perform. Which is the one that he actually performs? A plausible answer seems to be: all of them. He kills the Archduke *by* firing the gun, he fires the gun *by* bending his finger. The three actions constitute an hierarchy of means and ends, of a direct and transparent kind, that makes it reasonable to say that Princip in one blow performs them all. Or we might prefer to say that there is only one action, which can be given different descriptions, some being “thinner”
and some being “thicker” than others. In this terminology (3) gives a thicker description of Princip’s action than (2), but a thinner description than (4).

Hierarchies of means and ends is not the only dimension in which action descriptions may be variously thick. Compare the following statements, uttered in the context of a football match.

(7) L kicks the ball

(8) L takes a corner

It is just as possible, in this case also, to say that L takes the corner by kicking the ball, but the relevant relation here is not that of means to end – that the corner gets taken is not an effect of the ball being kicked. In certain circumstances kicking the ball counts as taking a corner. One relevant circumstance is that the kick shall be part of a game of football (in the European sense) – the other circumstances are specified by the rules of the game. I will call this dimension, in which action descriptions may vary in thickness, “convention thickness.”

The convention thicker description (8) of L:s action is only available to someone who knows the applicable set of rules: who knows that football is and who knows that football is being played here. Someone without this background will have to contend himself with thinner descriptions – and if he produces a report of what he has seen, it will, to aficionados, be only too obvious that he has missed the point completely. Similar things are true of other conventionalized or ritualized actions – only in the thinnest sense are they available to the one who does not know the game. Compare “having a small sip of wine” with “taking communion”, or “putting a piece of paper in a box” with “voting”, or “pressing some buttons so that bits of colored paper with pictures on them come out” with “drawing money from the cash dispenser”.

The expression “thick description” is originally from Gilbert Ryle, but it was Clifford Geertz’s use of it to explain the goals of anthropological research that gave it wider currency.

The task of the anthropologist, according to Geertz, is to enable us to produce and understand thick descriptions of behaviors that from the beginning are only available to us in the rudimentary way exemplified by the kicking of the ball, the sipping of the wine, or the pressing of the buttons.

Conventional action opens the possibility of a conflict between different ways of ascribing meaning to the same piece of behavior. On the one hand we have the meaning supplied by the agents own understanding, what the agent sees his behavior as. On the other hand we have the meaning supplied by the system of rules. Normally we take it for granted that the agent voluntarily takes part in the relevant activities, aware of what her activities “count as”. But this is not always the case – many sporadic churchgoers have only vague inklings about the point of the ritual and get along by doing what the others do. One can easily imagine someone imitating those

---

14 More than anyone else, it is John Searle who has emphasized this dimension of action thickness, most elaborately in Searle (1995). We will come back to it in ch. 5.

around him taking communion, without knowing one bit about Christian doctrine. Will he have taken communion, or just have tasted some wine and crackers in a church?\textsuperscript{16} For many ritualized actions that have important consequences conventionally attached to them – to vote, to marry – the main point of the ritual is that it shall make it unambiguous whether the action was performed or not, but it is still taken as a condition of performance that the agent was aware of his act.\textsuperscript{17}

Means-end thickness and convention-thickness are not the only dimensions in which action descriptions may be variously thick and thin, however. I will briefly indicate two more thickness dimensions – we will need them later when we come to the topic of the “chaining” of significations.

\textit{Part-Whole thickness.} I write a word as part of writing a sentence, I write a sentence as part of writing an article. One can correctly describe my writing of the word as my writing of the article, and also say that I write the article \textit{by} writing the words in it, but the relation between the two descriptions does not seem to be neither of the means-end nor of the counts-as variety. It is not a convention that my writing of a particular word should be part of my writing a particular sentence, and so forth, but neither is it a means to that end.\textsuperscript{18}

\textit{Exemplification thickness.} My action may have descriptions on varying levels of generality: I may compliment someone on his haircut in order to be nice to him. Again, this case does not seem to be covered by any of the earlier categories. The compliment is not the means that I use for being nice to him, it is my being nice to him; there is no convention that makes issuing the compliment count-as being nice; the compliment need not be part of my being nice to him but may very well be the whole of it.

### 3.7 Barthesian meaning: manipulation

I have tried to capture the difference between two concepts of meaning by way of the distinction between intentionality of the first and of the second order. The point may be expressed in terms of the notion of \textit{seeing-as}. Kuhnian meaning is a concerned with \textit{seeing-as}. Weberian meaning is concerned with \textit{seeing-as-seen-as}.

So far, neither of these notions has anything to do with someone \textit{meaning} anything – with \textit{designating}, \textit{symbolizing}, \textit{communicating}, conveying a \textit{message}, and so forth. The wood-cutter’s behavior is meaningful – to the agent and to the observer who understands is – but it is not an act of communication. Not even in a metaphorical sense is the wood-cutter \textit{saying} that he is cutting wood, he is just doing it.

So, what more has be true of an action for it to be counted as \textit{symbolic} or \textit{communicative}, even in the most elementary way? A natural suggestion is that we have to climb yet another step on the ladder of intentional orders. Meaning something, in this further sense, is to act \textit{so as} to influence the intentional or emotional state of someone – to act \textit{so as} to affect the beliefs,

\textsuperscript{16}Compare Max Weber’s discussion of “traditional” actions, in the first chapter of (1922)
\textsuperscript{17}We will come back to conventional actions and “performatives” in chapter 6.
\textsuperscript{18}Searle (2001, p xx) actually mentions this sort of example, but takes it to be a species of counts-as thickness.
CHAPTER 3. MEANING

Meaning something by something is to produce or display something with the intention that it shall be seen-as so and so, by someone. To see something as meaningful, in this third sense, is a case of seeing-as-intended-to-be-seen-as.

Meaning in this sense is central to a lot of classical semiology, for example to much of the work of Roland Barthes. Think again about the smearing of lipstick on a glass, in order to make someone believe that there has been a woman present. To be able to ascribe such an action to someone we must ascribe second order intentionality to the agent himself – the agent must see the result of his own action in the light of the impression it would make on somebody else’s mind.  

I will use the word manipulation as a general term for the intentional structure of this level, taking its place between (simple) action, on the level below, and the even more complex structure of communication, that we will find on the level above. The term is not an entirely happy one – while it fits certain central cases quite well, it may seem to misrepresent others – but I have found no better alternative. I will present the different forms of manipulation by means of three examples, making some general observations as I go along.

3.7.1 The Applicant

Wearing clothes is a type of action that can have different purposes – to cover the body or to protect against the weather, for example. Let me use the word “functions” for such non-semiotic purposes. Now, suppose that one day, when I’m going to the bank to apply for a loan, I dress in a different way from what I usually do – instead of my habitual jeans, t-shirt and sweater, I put on a suit and a tie. From a functional perspective there is nothing to motivate the change – rather the reverse, in fact, as I will feel a bit uncomfortable in my disguise. So why do I wear it? In order to make a certain impression, of course, on the bank’s representative – I want to be apprehended as a financially trustworthy person, someone who will be eager and able to pay his debts. I do what I do in order to be seen in a certain way by my vis-a-vis, with my choice of dress I try to manipulate his or her beliefs and attitudes.

In this case, the word manipulation, with its negative connotations, seems particularly apt, because of the element of deception involved in my action. While the success of a “communicative” act, as we will see, depends on my intentions being disclosed to the “audience”, the desired effect of the Applicant’s action depends on the intention being hidden – were my audience to know the motives behind my temporary adoption of a different dress code, the effect would vanish, or even be reversed.

The possibility of such deception depends on an important phenomenon, that I will dub the meaning-function ambiguity. Apart from any manipulatory purposes, my act of wearing a suit still has a purely functional moti-

\footnote{The component of deceit makes the example more vivid but is not really essential, of course. We can easily imagine a woman smearing her own lipstick on the glass, in order to indicate to some future observer that she herself was there.}

\footnote{The term is from Roland Barthes, as is the general spirit, though not the details, of the example. Cf. for example (Barthes; 1988).}
vatiation. I wear it to cover my body, and the semiotic intention, the meaning, attaches to the function as something extra. Now, as the function in itself is sufficient motivation for my act, I need not openly avow my manipulatory intention. The non-semiotic function acts, as Barthes would say, as an *alibi* for the semiotic intention.

Note also, that here we have not only something that might be called a message, something the agent wants a receiver to pick up as the result of his action, but also a *carrier* or *vehicle* of the intended meaning. The Applicant’s action involves displaying something, the dress, from which he wants his audience to infer to something else, a certain financial status. In fact, any human action may, through a process of “semantization”, take on an extra symbolic dimension. While my primary intention in chopping wood is, usually, to produce firewood, I may also do it in order to manifest that I do it – perhaps with the further intention of demonstrating that I fulfill my share of household responsibilities, or to show off my manual labor skills.

The behavior that serves the function of getting firewood is simultaneously displayed in an act of manipulation – the woodchopper intends his behavior not only to *be* a case of chopping wood, but also to be *seen-as* such. In many cases of conventional and cooperative actions this semiotic dimension is essential to the primary act being successfully performed – the consequences that should attach to it, by convention and/or by the subsequent actions of other agents, can only take place if it is made clear that the act *has* been performed.

### 3.7.2 The Photo

To clarify the contrast between “natural” and “non-natural” meaning, Paul Grice (1957) compares a photograph and a drawing, with the same pictorial content. In both cases we are to imagine the picture showing a “Mr. Y. displaying undue familiarity to Mrs. X.” The actions whose meaning we are asked to contemplate is showing the photo and the drawing, respectively, to Mr. X, with the intention of getting him to believe that “there is something between Mr. Y and Mrs. X.” Only in the case of the drawing, says Grice, can this count as non-naturally meaning something.

What is the difference between the two cases? Surely that in the case of [the photograph] Mr. X’s recognition of my intention to make him believe that there is something between Mr. Y and Mrs. X is (more or less) irrelevant to the production of this effect by the photograph. Mr. X would be led by the photograph at least to suspect Mrs. X even if, instead of showing it to him, I had left it in his room by accident; and I (the photograph shower)

---

21It may, perhaps be objected that the function motivates my wearing clothes, but not my wearing this particular set of clothes. But what I want my adversary to think is that the clothes that I wear is my unmarked choice, the “natural” thing for me to wear – I want him or her to take my dress as a natural, not as an intended, sign of my economic status and my social habits.


23As we will see, there may be manipulation without display, however.

24This is why the form of of many conventional actions is ritualized, and also why there are in many cases special agencies with the task of deciding whether the action has been performed, in cases of doubt.
would not be unaware of this. But it will make a difference to the
effect of [the drawing] on Mr. X whether or not he takes me to
be intending to inform him (make him believe something) about
Mrs. X, and not to be just doodling or trying to produce a work
of art (Grice; 1957, p 218).

I will come back to Grice’s notion of non-natural meaning, below, but for
the moment I am only interested in the photographic side of the contrast.
Clearly, we have here another case of manipulation, in the defined sense –
an act of displaying something to Mr. X with the intention of influencing
his intentional state, i.e., getting him to believe something. The main dif-
ference between the Applicant and the Photo, is that in the latter case there
is no deception involved – the intended effect does not depend on the ma-
nipulatory intention being hidden from the audience, it is just independent
of it.\footnote{Independent in principle, that is. In practice there may be some complications. If the
audience suspects the sender of a bias, consciousness of the senders intentions may weaken
the effect of the evidence presented. If the audience, on the other hand, has trust in the sender,
consciousness of the senders intentions may strengthen the effect.}

The structure of this example is not odd or peripheral – it is the typical
structure of a very important form of manipulation, namely the giving of
evidence. In contradistinction to telling someone something and expecting
him to take my word for it, I display some facts that points to what I want
him to think and expect him to draw his own conclusions.\footnote{The Applicant, in a way, works by the same mechanism – I intend my audience to take
my dress as evidence of my financial status – but the evidence is “fabricated”. One might be
tempted to say that I lie, and use the functional alibi to escape from blame. Note however that
I need not be lying in the sense that I try to make my addressee believe something that I, the
sender, believe to be false – I may very well be the person I want to be taken to be. In that case,
my dishonesty will consist only in trying to make my addressee arrive at a true conclusion by
way of false reasoning.}

Of course, one common way of “displaying” the facts, is to tell my audience what they
are, and so an act of full gricean “communication” may be the means for
accomplishing an act of “manipulation”. This is a form of “chaining” of
significations – I will discuss chaining below.)

3.7.3 Film Music

Motion films supply a multitude of examples of manipulatory techniques,
often differing in informative ways from those we have already considered.
Consider the use of music to affect the viewers mood – to frighten, to startle,
to soothe. Clearly this is a means of manipulation, in the indicated sense:
something intended to affect the intentional states of someone.\footnote{If you are inclined not to think of moods as intentional states, you may think the example
less clear. I will bypass those worries, for the moment – they can be answered either by arguing
for the intentional status of moods, or by slightly changing the example.}

How does it differ from the Applicant and the Photo?

One salient difference is that there seems to be no display involved, or
at least not in the same sense. The audience is, of course, expected to hear
the music, but they are not expected to attend to it. And not even in the
weakest sense is the intended effect intended to be reached by any form of
inference, from features of what is heard to something else. The effect is, so
CHAPTER 3. MEANING

to speak, intended to be more direct; the music is a *means* for accomplishing the effect, but it is not a *signifier*. The same thing goes for many other cinematographic effects, for example cutting techniques. “Accelerating cuts” may be used to impart a sense of acceleration and breakneck speed to the events portrayed, but, again, the transference from features of the film to features of the action is not one of inference.

What about the importance of the manipulatory intention being hidden or revealed? The tricks of movie making are no secret to many spectators, but that does not stop them from working their magic – or does it? It seems as if the achievement of the intended effect is to some extent impaired by too vivid a consciousness of the intention to produce the effect, and of the means employed. What makes a “turkey” into a turkey is not only that it is generally bad, but that it constantly reminds the spectator of what it is trying to accomplish, and thereby of its failure to do so.

3.7.4 Meaning in nature

Before we go on to even higher orders of intentionality, and correspondingly more complex ways of meaning something, I want to add a few remarks to our earlier discussion of natural meaning. Standard examples of natural meaning comprise things like spots that mean measles, smoke that means fire, or dark clouds that mean rain. Extending the use of a medical term we may call such things *symptoms*. In thinking of symptoms as meaningful, we are referring to them as possible sources of information about something other than themselves – the doctor can use the presence of a certain kind of red spots on the skin to infer that the patient has measles. But that use is not a *function* of the spots, they are not there in order to facilitate that inference, in anything like the sense in which I wear a suit and tie in order that the clerk shall infer something about my solvency. Interpreting a symptom, I exploit a certain natural regularity for my information gathering purposes, but I do not think that anyone or anything is actually informing me or telling me anything.

But there are things in nature that may be said to be meaningful in a stronger sense, similar to the meaning of manipulatory signs. In particular, there are features and behaviors of living organisms that actually have the *function* of influencing the behavior of other organisms. Take the example of pheromones – substances produced and released by many animals to notify suitable other members of its species of the fact that they are available for procreation. Or take the bright colors flaunted by many poisonous plants and animals, to protect themselves from potential predators by alerting them to the dangers of taking a bite. Such phenomena are not just symptoms, they are *signals* – they exist in order to influence the behavior of others, their function is, so to speak, to convey a message, to be interpreted in a certain way.

I take it as clear enough that there is a distinction between signals and (mere) symptoms, and that signals have an interesting similarity to Barthesian meanings of the type described above. But it still is a philosophical problem, to say what can properly be meant by the word “function” in this context, and of how functions in this sense relate to the intentions and purposes of the preceding discussion. According to one way of thinking,
intentions are conceptually prior to functions, and talk about meaningful signals in nature is basically a metaphorical extension of intentional vocabulary – it is only the human observer that can supply the valuations of different outcomes that underlies the distinction between a function and a mere effect.\footnote{A prominent exponent of this view is John Searle, for example in REF.} According to a contrary line of thought, however, it is natural function that is the primary phenomenon, and human intentionality and meaning should be understood on the basis of biological functions.\footnote{The basic idea is that functions are effects that explain the (continued) existence of the features whose effects they are, in the manner of evolutionary theory (REF to Wright’s article). The most vigorous development and defense of the program of basing human meaning in biological function is by Ruth Millikan, for example in REF.}

### 3.8 Gricean meaning: communication

There is a lot more to be said about manipulation, and I will soon come back to it, but first I want to introduce a fourth, and last, level of meaning, which I will call “communication”. Most of what is interesting about manipulation comes out best when it is considered in contrast to communication. As already hinted, I will lean heavily, for the notion of communication, on Paul Grice’s ideas about non-natural meaning. Grice has extended and modified his account on many occasions, to fend off or accommodate various criticisms. I will come back to one of his later and more elaborate versions of it in a moment, but I will first use the simpler original idea to indicate the place of Grice’s conception in the hierarchy of intentional orders. Grice means to capture a notion of what it is for someone to mean something by uttering something which is prominently involved in ordinary linguistic communication. He argues extensively that, in this context, it is not enough for the utterer to intend to influence the audience in a certain way, the utterer must also intend the audience to grasp the speaker’s intention and thereby to be influenced in the intended way. In Grice’s own words

\[
"U \text{ meant something by uttering } x" \text{ is true if and only if, for some audience } A, U \text{ uttered } x \text{ intending:}
\]

(1) $A$ to produce a particular response $r$

(2) $A$ to think (recognize) that $U$ intends (1)

(3) $A$ to fulfill (1) on the basis of his fulfillment of (2).\footnote{Grice (1989, p 92)}

Let us take a simple example. Someone asks me what time it is, and in reply I say “It is a quarter to five”, meaning by that utterance that the time is a quarter to five. For my utterance to have that meaning, according to Grice, it must be true (1) that I intend to produce a certain effect in my hearer, namely that she shall believe that the time is a quarter to five, and (2) that the hearer shall recognize that I intend her to believe this, and (3) that she shall adjust her beliefs accordingly, precisely because she recognizes that this is what I intend. My utterance has the meaning it has
in virtue of my having these intentions, and the hearer understands what
I mean if she recognizes that I have them.\footnote{It is not a part of the understanding, however, that she actually comes to believe what I say. While understanding perfectly well what I say, she may think that I am lying, or that I am simply wrong.}

What makes meaning something, in this sense, belong to a higher or-
der of intentionality than manipulation is, of course, clause (2). Using the
seeing-as terminology, we can say that to take someone as meaning some-
ing, in this sense, is a case of seeing-as-desiring-to-be-seen-as-desiring-
to-be-seen-as. If we count the intentional verbs in this formulation, we can
see that, strictly speaking, we are now at intentionality of the fifth order,
but as I will go no further, and as I can see no use for an intermediate level,
I will continue to speak of this level as number four.

\subsection{Relevant features and modes of correlation}

Grice's original analysis has been subjected to a number of counterexam-
ples and criticisms, that have caused him to modify it in various ways. I will
look at one such modification, inspired by a counterexample by John Searle
(1965). Searle wants us to imagine an American soldier being caught by
Italians during the war in North Africa. In the hope of making his captors
think that he is a German officer, the American utters the only German
sentence he remembers from school: "Kennst du das Land wo die Zitronen
blühen." Running through the clauses of Grice's definition we seem to be
led to the conclusion that the American not only means something by his
utterance, but that he actually means that he is a German officer.\footnote{As it stands, the definition only says what is is to mean something in general, but Grice's intention is that meaning different things shall be distinguished by differences in the response R. In Searle's case the intended response is that the audience shall come to believe that the utterer is a German officer, so this is what the utterance means, according to the analysis.}

That conclusion is obviously absurd, thinks Searle, and he suggest mod-
ifying the definition by a clause demanding that the utterance $x$ shall be
of a kind conventionally used to elicit responses of the type $r$. This sugges-
tion is not approved by Grice, for at least two reasons. First, he wants his
definition to cover cases where there is no conventional connection between
the utterance and the intended effect. Second, he wants to use the notion
of non-natural meaning to understand what conventional meaning is, and
so cannot allow himself just to presuppose it.

Grice's own solution to Searle's problem case involves two new ideas:
that of message-relevant features of the utterance and of different modes of
correlation between such features and the intended response of the receiver.
It sounds a bit complicated but the basic idea is simple. To mean something,
in the relevant sense, is to intend an audience to react in a certain way
to what I say or do – but how does the audience know what reaction is
expected of them? The audience cannot peep into my mind directly, and so
I have to design my utterance in such a way that it gives a good enough
cue to what my intentions are. I must, so to speak, supply a vehicle that
the audience can use to travel from the signifier to the signified.

When I produce or display something as a sign there are always some
specific features of it that I intend to carry the message, while other proper-
ties of it are irrelevant for my purpose. In the case of the Applicant, for example, it is relevant that I am wearing a suit, from within a certain normal range, but the weight and the exact type of cloth do not matter. Uttering “It’s a quarter to five” in response to a question about the time, it is relevant what English words I use, but my accent or the pitch of my voice are immaterial. I also intend the audience to think that the relevant feature is correlated with the intended response in a certain specific way, and to use their knowledge of that correlation to reach the intended response. In the case of the Applicant the mode of correlation is dress habits – I intend the bank clerk to use his beliefs about how different sorts of people habitually dress to reach a conclusion about what type of person I am. In the case of telling what time it is, the mode of correlation is linguistic convention: I intend the listener to recognize the words and to use her knowledge of what these words conventionally mean to figure out what I intend her to believe. In other cases I may rely on purely contextual clues, or on similarities that I hope the audience will perceive, and so forth. In this way, Grice can allow a role for linguistic conventions in his intention-based account of meaning, while still being able to explain the how such conventions can arise from other modes of correlation, for example by repetition and the formation of habits.

3.9 Summary

In this chapter we have tried to chart some of the uses of the word meaning that are pertinent to the sciences, and in particular to the cultural sciences. As a sort of guiding thread we used the traditional idea that there is a specific notion of meaning (and a corresponding notion of interpretation) that can be used to demarcate the cultural from the natural sciences.

The first contrast that we drew was between natural meaning, on the one hand and the sort of meaning primarily connected with human intentionality, in the phenomenological sense, on the other hand. Intentionality is metaphorically described as the capacity of mental states to be “directed” to things and states of affairs in the world; it is the property of thoughts, fears, wishes, fantasies, and so forth, to be of something. This property, again, we associated with the use of concepts to render experience “meaningful”.

Neither of these conceptions of meaning gives us a hold on the distinction between the natural and the cultural sciences. We pointed out that “natural” meanings play an important role in the social sciences and the humanities – indeed, a role that seems to be increasing with the tendency of modern theoretical approaches to emphasize the “forensic” aspect of human products, to treat them as remnants of past structures and processes rather than as documents carrying intended messages. It is just as clear that intentionality is as basic to the natural sciences as to every other field of human activity. No form of knowledge or experience gets by without concepts: to find and elaborate suitable conceptual tools is an integral part of progress in any discipline.

Nevertheless, the notion of intentionality gives us a clue to the sense in which a preoccupation with meaning really does distinguish the cultural...
CHAPTER 3. MEANING

95

The social sciences and the humanities don’t just *employ* concepts, like any human activity, they are also *about* the use of concepts; all forms of scientific inquiry *is* intentional, but only the social sciences and the humanities are *about* intentionality.33

In passing, we may note that this way of looking at things gives the cultural sciences a “subjective” subject matter – they deal with phenomena related to consciousness and mental life. But this does not imply, of course, that they only deal with subjective phenomena, in this sense, nor does it imply that they are themselves “subjective” in the pejorative sense of being dependent on the whim of the investigator – knowledge of subjectivity can be just as objective as any other kind of knowledge.

A picturesque way to sum this up, is that it is an integral part of research in the cultural sciences to see the world not only through one’s own eyes but also through the eyes of others, namely of those that one studies. I must be able not only see the world through my own concepts, as this or that, but also through the concepts of others – I must see the world as they see it.34

Next we tried to give some structure to the field of intentional phenomena, by introducing the notion of *nested* intentionality, and of different orders or levels of intentionality. On a first level we found the study of *action*, i.e., of human behavior described through the meaning it has for the agent. One important consideration in this connection is the notion of action descriptions being variously “thick”, and we indicated some different dimensions in which descriptions can differ in thickness.

All actions are meaningful in the minimal sense that the cannot be identified apart from the way the are conceptualized by the agent, but they need not have meaning in the further sense of being signals or communicating messages. To have phenomena that are meaningful in this stronger sense we must go to the next level of nested intentionality and look specifically at actions that are intended to influence the mental states and behavior of other agents. For this level of meaning we used, with some misgivings, the term “manipulation” – chosen to highlight the fact that this order of meaning need not involve the intended *cooperation* of a receiver, without implying that it must always involve a malicious intent on the part of the sender.

The notion of cooperation, in contrast, is central to meaningful actions on the next, and for our purposes the last, level of complexity, where it is natural to talk of “communication” in the fullest sense, exemplified by, but not restricted to, ordinary linguistic communication. To communicate, in this sense, is to perform an action which is not merely intended to *influence* the state of an audience in a certain way, but to do this precisely by way of an intended *recognition* of this intention on the part of the audience. We ended this discussion with a few observations on some of the different

---

33These formulations point to a certain relation between the cultural sciences and philosophy, which has also often been said to be the study of concepts or intentionality. Using clues from Max Weber and the later Wittgenstein, Peter Winch (1958) has developed this analogy very far, and argued that the relation between philosophy and the social sciences is indeed very intimate.

34Again, this is not all that I do, of course, as a cultural scientist. I must also see the world from my own perspective, for example in order to explain why they see things as they do.
means that a speaker can use to get this intention across, the intended “modes of correlation” between the signal and the meaning.

***

With all this said and done, we have by no means exhausted the bewildering variety of uses of the word “meaning” in science and in daily life. If there is one lesson to be drawn from this chapter, it is, perhaps, the need to be aware of this variety and to be prepared to analyze and explain what “meaning” is supposed to mean in each context where the word pops up. As always, when in doubt, it is a good exercise, to try to do without the word and explain oneself in different terms.
Chapter 4

Interpretation

Just like the word meaning, and many other words that are central to us in this book, the word interpretation exhibits a process/product ambiguity. “Meaning” may be what someone does, in uttering or displaying a string of words or other signifiers, or it may be the result of such an activity, what the signifying expression, on this occasion, means. As an activity, interpretation is the inverse of the activity of meaning something: it takes the signifier as meaningfully uttered and tries to reconstruct or find out what was meant. The result of the activity is the specification of a meaning for the utterance: an interpretation.

As could be expected, there is a range of different notions of interpretation that runs exactly parallel to the corresponding multiplicity of notions of meaning. To natural meaning there corresponds what may be called natural interpretation – the spots mean measles and the doctor is right if she interprets them as meaning precisely that, and she would be wrong, in this case, if she interpreted them as an allergic reaction. And to different levels of intentionality and intentional meaning there are corresponding forms of interpretation.

A rather special case, though, is the basic level of intentionality or conceptuality itself. We have accepted, along with Husserl and Kuhn, to speak of experience itself as meaningful, in the sense that I can never see or think about anything except by seeing it or thinking about it as something, in terms of concepts that I possess. It comes naturally, at least to many of us, to speak of this as a kind of interpretation – it is presumably along these lines that Nietzsche is thinking, for example, when he says that there are no facts but only interpretations. But interpretations of what? In standard cases of interpretation we make a distinction between the object that we interpret – the text, the painting, the behavior, the signifier in some sense – and the meaning that we look for through the interpretation. But in the notion of a meaningful experience there seems to be no place for a signifier, there is nothing there “before” the interpretation, so to speak, and so how can there still be an interpretation? Such misgivings have been aired by Kuhn himself, with the conclusion that we should, after all, be wary of using the word interpretation in this context, but I will not delve deeper into these misgivings here – suffice it to say that the forms of interpretation that we are concerned with in this chapter are such that they allow for the distinction between a signifier and a signified, between something to be
CHAPTER 4. INTERPRETATION

interpreted and a meaning to be ascribed to it.

4.1 The intentional fallacy

A large part of the debate about interpretation for the last sixty years or so, and in particular about literary interpretation, has been concerned with the role of the author’s intentions in interpretation. The debate was sparked by a now classical article by Beardsley and Wimsatt (1946) called “The Intentional Fallacy” – as the title indicates they were adverse to the very idea of the author’s intention playing any role in the appreciation or interpretation of literature. As the notion of intention has figured so prominently in our account of meaning, and will continue to play a role, although less prominent, in this chapter as well, it may be worthwhile to briefly review and comment on their argument.

The first thing to note about “The Intentional Fallacy” is that the bulk of the article is not about meaning and interpretation at all, but about evaluation. The question is whether the critics evaluation of, e.g., a poem should consider it in relation to what the poet has intended to achieve, and judge its success or failure in relation to that goal, or if it should apply more “objective” standards. Questions about aesthetic evaluation is not our focus here, but it is difficult to understand where the conflict is supposed to be between these two approaches: the usual thing, surely, would be that the artist intends her work to meet some objective standards, connected to the type and genre of work it is, and it is precisely this intention that motivates us to apply these standards to it.

It is not until the end of the article that we get to some remarks about meaning and interpretation that are relevant to our present concerns. The target of the polemics, however, is not any notion of intended meaning in the sense that we have been discussing it, but the practice of seeking biographical readings of literary works – of reading literature for the information it might convey about its author. This may, indeed, in many cases be considered as a fallacy of some sort, but to call it the “intentional” fallacy is surely a very unhappy misnomer. The most pertinent reason to avoid biographical readings of most literary works is obviously that they are not intended by the author to be read in this way, and if, in some cases, a biographical reading is intended, it seems just as appropriate to follow that intention, too.

Despite its reputation in this debate, there simply is no general argument against intentional interpretations in the article by Wimsatt and Beardsley. To point this out is not to say, of course, that every type of interpretation should be looking for intended meanings – we will come back to the merits and limitations of different kinds of “intentionalism” many times in what follows.

4.2 The hidden and the manifest

There are many types of interpretation, but one feature that they share is the contrast between, on the one hand, what is interpreted, the object of
interpretation, and, on the other hand, what is sought through the interpretation, the meaning. We picture ourselves in front of some object, asking the question: What does this mean? To get to more specific types of interpretation we must say something about what sort of a thing “this” might be; what sort of a thing “what” might be looking for; and perhaps what sort of relation “means” might be taken to suggest between them.

I will talk about the terms of the meaning relation as the “manifest” and the “hidden”, respectively. There is no deep meaning to these terms, we just need a couple of words that suggests the contrast between something that we, in a certain context, take as a given and want to interpret, and something less obvious that we want to find out. Depending on how we specify the hidden and the manifest we will get different types of interpretation. I will give a rather extensive list of such pairs, adding some comments as I go along. Here is a first group of possibilities:

<table>
<thead>
<tr>
<th>Manifest</th>
<th>Hidden</th>
</tr>
</thead>
<tbody>
<tr>
<td>Letter sequence</td>
<td>Word</td>
</tr>
<tr>
<td>Word</td>
<td>Concept</td>
</tr>
<tr>
<td>Meaningful Expression</td>
<td>Referent</td>
</tr>
<tr>
<td>Locution</td>
<td>Illocution</td>
</tr>
<tr>
<td>Explicit</td>
<td>Implied</td>
</tr>
<tr>
<td>Literal</td>
<td>Figurative</td>
</tr>
</tbody>
</table>

All of the pairs in this first group have something to do with language and linguistic utterances. This is not to suggest that all interpretation is concerned with language, or that linguistic interpretation should be a model for interpretation of other kinds, but it is a convenient place to start and non-linguistic interpretation will get its due attention as we go along.

To take a mark as a letter of some alphabet is clearly an act of interpretation, in a quite common sense of that word. The very same spatial shape may signify different letters of different alphabets. (I think I was well into my teens before I realized the relation between the fact that the name of the Soviet Union was abbreviated to “SSSR” and the fact that Soviet hockey players had “CCCP” printed on their shirts.) It may be doubtful whether a

---

1In other words, we will get a list of more specific interpretive questions. And we might be tempted to play a trick on “meaning” and “interpretation” which it is often useful to play on troublesome words – namely to try to do without them. Each time we feel tempted to ask what anything means, we might demand of ourselves that we pick something more specific from the list, or else add some new specification if we are not satisfied with what is already there.

2In fact, I think that much of the philosophical debate about interpretation, particularly in aesthetics, has been skewed by over-reliance on linguistic models. This is true both of analytic aesthetics, from Beardsley and Wimsatt to the current debate about intentionalism in interpretation, and of the French structuralist tradition (including post-structuralism).

3Nevertheless, some will probably resist this use of the word “interpretation”. There is usually no interpretive process involved, but just immediate recognition; there is in most cases no room for interpretive disagreement; there is, again in most cases, no phenomenological distinction between the manifest and the hidden – we do not experience the manifest and the hidden as two connected things. These are all contrasts that “interpretation” is used to indicate in different contexts, and where the mark-letter pair seems to fall on the wrong side of the fence. On the other hand, there are similar cases where it is entirely natural to use the word “interpretation” – difficult cases of decipherment, cryptology, etc. However, terminology is not the issue here, so I will just content myself with recording my usage.
piece of scribble on a sheet of paper is an example of somebody’s handwriting, or just a collection of meaningless shapes, and the decipherment of it — if it is writing — may be a laborious interpretive task.4

On the next level, again, we may take it as given that some marks constitute a determinate sequence of letters and still wonder whether they make up a word, and if so, which one. Is ‘rock’ the term for a large stone, or for a certain musical genre or is it the Swedish word for a man’s overcoat? These are not just different meanings of the same word — they are different words comprised of the same letter sequence, and deciding which one is relevant in a certain case is an act of interpretation.

Taking word identity as established, we may go on to ask about what a certain word means, what concept is associated with it on a particular occasion. Does the word “prince” in a certain context carry the significance of “sovereign ruler” or does it mean “son or grandson of a king or queen”? On a still higher level we may go on to ask whether the linguistic expression in question, its sense taken as established, refers to anything and, if so to what. Does the word “prince”, in this utterance, refer to a particular person? And, if the answer is yes, who is it?

Let me stop here for two general observations. The first, and obvious, one is that interpretations of different types go together in hierarchies — what is “hidden” for one interpretation may itself be “manifest” for another interpretation, on another level of the hierarchy. And this leads to the second observation, that interpretive tasks are often underspecified — to speak of interpreting a “text” or an “utterance” often gives very little clue not only about what is taken as hidden, what is sought through the interpretation, but also about what is taken as manifest, the clue from which the investigation starts.

The particular ladder we have climbed so far, has some connection with different aspects of what J. L. Austin, from the side of the speaker, called the “locutionary act”. To perform a locutionary act is, according to Austin’s account, to utter certain sounds as belonging to a certain language, with a certain grammatical structure, and with a certain intended sense and reference.5 The locutionary act itself, however, is not yet a complete speech-act, but only the foundation for further levels of signification. In Austin’s terminology, the next step is to take a complete locution (with its sense and reference) as the manifest, ready to be interpreted as expressing a certain illocutionary force. An utterance of “The ice is thin” (as an English sentence, with a certain meaning, referring in a specific situation to the ice on a certain little pond) may depending on the situation be taken as just a description, as a warning or even as a threat — which ever way one takes it, it will count as an interpretation of the utterance, an interpretation that presupposes interpretations on the previous levels to be already in place.6

4The same thing goes, mutatis mutandis, for sounds in relation to phonemes, of course, but to keep the exposition simpler I will stay with writing and visual shapes for the moment.
5Austin (1962, p 92) distributes these functions over three sub-acts: the phonetic, the phatic and the rhetic acts.
6There are different ways to construe the notion of illocutionary force. Austin reckoned with hundreds, and maybe thousands, of different illocutionary forces, but most of his modern followers seem to be content with some small one-digit number, a sort of pragmatical counterpart of grammatical mood.
The next pair on the list is explicit vs. implied. A lot of interpretation has to do with a movement from what is actually said to some further conclusion that the listener or reader is supposed to draw from the utterance. One type of example of this is what Paul Grice has called “conversational implicature”. Thinking that your food is a bit bland, you turn to your neighbour and say “Can you pass me the salt?”. A perfectly true answer to that question might be a simple “yes”, but that is not the response you are expecting or hoping for – you expect him to interpret what is explicitly only a question about his abilities as implying a polite request that he shall hand you the salt. This sort of “reading between the lines” is ubiquitous in almost any type of linguistic conversation, and from the point of view of the receiver it constitutes yet another level of interpretation, for which a “literal” speech-act is taken as the manifest and what is implied as the hidden.

Another type of meaning and interpretation that piggybacks on an already given layer of manifest meaning is, of course, “figurative” meaning: metaphor, metonymy, irony, and such. Whatever you make of Eliot’s description

When the evening is spread out against the sky
Like a patient etherised upon a table

it surely depends upon your previous understanding of hospitals, anesthesia and operation tables.

4.3 Leaving language behind

So far, all of our manifest/hidden pairs have been, in a broad sense, linguistic, but the same sort of thing can, of course, be said about representations in other “media”. Take, for example, the famous cartoon of Winston Churchill as a bulldog. To understand it you must first take the markings on a piece of paper as indicating a certain spatial shape, second, you must recognize that shape as a bulldog, with the face of a man, third, you must recognize that man as Winston Churchill, before you can, fourth, begin to figure out the point of picturing Churchill precisely as a bulldog, rather than, for example, as a poodle. We will not go into the specifics of how different media manage to represent and communicate, but will go on to levels of interpretation that are largely independent of such differences. To see that there are such levels we need look no further than to the fact that a story as presented in a novel may give rise to largely the same issues of interpretation as the film, the manga, or the theatrical version of the same material. What sorts of meaning is sought through such interpretations, and how can we apply the manifest/hidden distinction to it? Well, let us continue our list with some new pairs:

---

7Grice’s writings on this, as on most other things, are collected in Grice (1989) – the original paper on “Logic and Conversation” (pp. 22-40) is a good place to start.
Again I will comment briefly on each pair, in turn. Theories of how we understand stories and pictures rightly emphasise our ability to “fill in” and “flesh out” what we are actually told or shown, supplying missing details and unmentioned surroundings – we reason from parts to wholes.

To take whatever is concretely told in a story as an example and to generalise it – either in a descriptive or a normative dimension – is one of the most common and striking interpretive modes, both among professional critics and laymen. Indeed, if Aristotle is right, this is the main point of art – to express general truths about reality by means of examples that might not themselves be real.8 All thematic interpretation – “this is a love story”, “this is a story about urban life in the nineties” – is in this line of business.

What about the other direction? Can finding an example to fit a generalisation also be called an act of interpretation? Yes, at least sometimes – in philosophy this is one of the most frequent and cherished ways of pinning down the meaning of abstract and theoretical assertions. And, in the case of literature, one type of “particularisation” that springs to mind is the bringing to bear of insights from a work on the reader’s own case.

With the relation between rules and applications we are still in the dimension of the general and the particular. To find clear examples we may turn to legal interpretation. Fitting the law to particular cases is the bread and butter of lawyers and judges, and extracting a general rule from previous particular decisions is one of the main tools in that process.9

These hidden-manifest pairs clearly go beyond the confines of linguistic interpretation. The manifest, in these cases, is not a linguistic object. For example, when Ally McBeal (of the TV-series) is taken as an “icon of contemporary woman”, she may in some sense be said to be a “signifier”, but she certainly is not an element of language or speech, and the same thing goes for a legal case used as a precedent in a trial.

Although not itself linguistic, the manifest in these cases may, of course, be given or presented through linguistic means – it may be described, or otherwise denoted or referred to, in a linguistic act. But it need not be. The presentation of an example may be effected through some other signifying medium, for instance by a picture or a film. Or it may simply be given in experience.

### 4.4 Interpretation and description

The gist of the foregoing discussion is that the distinction between what is manifest and what is hidden is relative – what is taken as manifest for

---

8I will not stop to elaborate on the details and the different species of this sort of interpretation. Doing so would take us in two directions: towards the epistemological notion of induction, and towards the semiotic function that Nelson Goodman calls “exemplification”.

9Here we are in a domain that many, including H. G. Gadamer, have taken as paradigmatic of hermeneutics in general.
one task of interpretation may itself be the result of a “previous” interpretation. This way of talking runs contrary to ordinary parlance within the aesthetic disciplines. Literary scholars often speak as if there is an absolute distinction between texts and interpretations, and expect interpretations to be supported by what is actually “in the text”. In a similar way, discussions of visual art-works often alludes to a distinction between “description” and interpretation. In both cases, the relevant dichotomy is often assimilated to the purported distinction between fact and interpretation — what is captured by the description, or what is really in the text, is supposed to be a factual matter, while the content of the interpretation is more up for grabs, possible to argue about, maybe, but nevertheless involving a degree of arbitrary choice.

In a specific situation, of course, such ways of speaking may be entirely harmless, because the context makes it clear where the relevant boundary is supposed to be drawn, between what is taken as given and what is sought through the interpretation. Nevertheless, it is important to be clear that the distinction is not absolute — otherwise one runs the dual risk of over-factualising the description and under-factualising the interpretation. Literally speaking, there is nothing “in the text”.

Let us take an example. Here is a perfectly normal piece of text:

道可道非常道
名可名非常名
無名天地之始
有名萬物之母

I suspect (and even hope) that most of my readers will not be able to make much out of this writing.\(^\text{10}\) Perhaps we recognize it as an array of Chinese characters, but even that recognition depends on background knowledge and assumptions. To make the text say anything at all, we need to draw on vast supplies of information that is not present on the page where we see the scribble — information that the reader in more normal circumstances is expected to bring along to the encounter. Being able to read modern Chinese, we can perhaps supply a phonetic reading of each character. Assuming that the text should be read from left to right, starting upper left, rather than from top to bottom, starting upper right, we can stitch these readings together and make a string “dao ke dao fei chang dao ...”. Applying a lexicon we can have a go at supplying meanings for these words, and employing a grammar we can try to make out how the words fit together, making up sentences and assertions, perhaps. But what lexicon and what grammar shall we use? To decide that, we need yet more background information. Someone tells us that it is the first stanza from *Dao De Jing*, the

\(^{10}\)If you happen to be a fluent reader of classical Chinese, please imagine another example that works for you.
book of Tao, and so we know that it is the lexicon and grammar of classical Chinese (roughly the same language as in Mencius and Zhuang-zi) that is relevant—but even with that much granted, there is an on-going interpretive discussion about the syntactic analysis of the first two lines, with important repercussions for the philosophical import of the whole.

I could go on and on about this example, and will come back to further aspects of it later on, but I hope the present point is clear. There is no meaning “in” the text, and the only reason that it often seems as if there were is that the requisite background knowledge is so ready to hand, and so automatically applied, that we do not even notice it.

One noteworthy thing is that the relevant background information falls into two distinct categories. On the one hand we need general knowledge of various kinds that is brought to bear on the object, in this case, for example, about the Chinese writing system and the syntax and semantics of classical Chinese. On the other hand we need specific historical knowledge about the object to be interpreted, that gives us reason to apply precisely these semiotic systems to it. It is the historical hypothesis that this is a classical Chinese text, with probable origin during the “Warring states” period rather late in the Zhou dynasty, that motivates reading it according to the rules of Classical Chinese—and the interpretation of specific points may well depend on quite detailed historical assumptions, for example about the chronological relations between this text and other texts from the same tradition, like the writings of Zhuang-zi or Han Fei-zi.11

In this way, we also see that the common opposition between “conventional” and “intentional” readings of a text is mostly spurious. It may, of course, be the case that a certain author uses words in an unconventional way—something which an interpreter generally would wish to take into account—but even when this is not the case, it is generally the intention of the author to write or speak according to a certain convention that motivates our reading a certain text in the corresponding way. It is our belief that Shakespeare wrote Elizabethan English that motivates us to read him accordingly.

A word also about the ordering of the different levels. I have mostly spoken as if we have to do with an hierarchy, where higher levels presuppose and build upon lower levels. A syntactic analysis, for example, seems to presuppose a prior identification of linguistic elements, a semantic interpretation presupposes that the syntax is already in place, a figurative interpretation builds upon a more literal understanding, and so on. From the point of view of the interpretation as a product, this seems to be correct: an interpretation as a specification of the hidden presupposes that which it is an interpretation of, the manifest. But from a process point of view things are more complicated: it will often be the case that a primary reason to accept a lower level interpretation is that it helps us to make better sense at a higher level. As, for example, Gadamer has emphasised this runs all the way down to the identification of the linguistic text itself: an editor will of-

11 As usual in such contexts, there are a host of other problems connected with the establishment of the text itself. What we actually have is a modern rendering based on a certain textual tradition, going back, in this case, to the text of Wang Bi, from the third century CE. The very establishment of something as “the text” often involves complex questions of interpretation on different levels.
ten appeal to the fact that it “makes better sense” as a reason to prefer one textual variant to another, and may even suggest otherwise unsupported emendations of the text in order to clear up difficulties of understanding.\footnote{In line 3 and 4 of the Chinese quote there are two expressions usually translated “heaven and earth” and “the ten thousand things”, respectively. In most contexts these would be synonymous, both designating the world, no more and no less. But are they synonymous in this passage as well, so that the alternation is merely stylistic, and they could without change of meaning be substituted for each other, or the same expression could have been used in both places – as indeed it is in a certain very old manuscript? An answer to that question about word meaning can only be argued on the level of an overall philosophical interpretation of the work – a perfect example of the hermeneutic circle.}

A related point can be made about the relative importance of the different levels. In discussing ordinary action we touched upon the notion of action descriptions being differently “thick” in different dimensions: Gavrilo Princip kills the Archduke by firing the gun by pressing the trigger by bending his finger.\footnote{Cf. page 85.} In ordinary speech acts, as in ordinary actions, the highest level generally takes precedence: lower level actions are not important except as contributing to higher level accomplishments. Once it is clear what promise I have given, it doesn’t matter what precise words I used or even what language I employed. But in some circumstances more than one layer may be severally important. In ritual action, for example, it often as important how something is done as that it is done, and it has often been taken as characteristic of aesthetic communication that lower levels do not serve only as means for higher level expression: in a poem the acoustic qualities of the words and the precise nature of the verbal imagery may be as important as the overall meaning.\footnote{Ref to Goodman (1976) on repleteness.}

4.5 Reading and claim

I have several times alluded to the false dichotomy between facts and interpretations. There is nothing inherently non-factual about interpretations – for many types of interpretation there are perfectly clear standards of correctness. But the same standards are not relevant for all types of interpretation, and for some types of interpretations it does indeed seem out of place to speak of correctness at all.

One way to get a grip on these problems is to distinguish between two dimensions of what is usually taken as one interpretation: between the reading and the claim.

Let me revert to Dao De Jing. One of the pivotal words in the quoted passage, and of course in the whole book, is the word ‘dao’ itself. If you look it up in a dictionary it has a wide range of meanings, including ‘road’ or ‘way’ in both literal and metaphorical senses, ‘teaching’ or ‘doctrine’, and the verbal meaning ‘to say’ – all the way to being taken as a designation for an ultimate metaphysical principle, the unifying One or perhaps Nothingness underlying all things. Depending on the way you take it, you get a range of different readings of the passage. On one reading, for example, the focus is on different doctrines of philosophy (“daos”) – such as Confucianism, legalism, moism, etc. – and the main point would be that none of these
is invariably correct, independent of circumstances. On another reading “dao” is used as a somewhat paradoxical name for something that cannot be named, the ultimate inexpressible foundation of reality. Is any of those readings correct? Well, not in themselves. The question of correctness does not arise until one makes a claim with regard to the reading, and different readings may be correct with regard to different claims. There is good reason to take the metaphysical reading as correct with regard to how Dao De Jing has been understood in the later Chinese tradition, at least from Wang Bi’s commentary and onwards. If I claim that the metaphysical reading captures Wang Bi’s understanding of the passage I am probably correct. But I may still be wrong, as some have argued, if I claim that the metaphysical reading captures the original understanding of the passage, relevant to whoever first formulated it – perhaps the first, non-metaphysical, reading is correct with regard to that claim.

Both of the indicated claims are historical: they concern historical facts of a certain type, namely how a certain text has been understood by a certain person or a group of persons in the past. The indicated situation is typical of most “classical” texts: that we have a multitude of ways to understand them that each may be of historical interest in their own right. How did Plato understand his own dialogues? How did Aristotle take them? How were they read in the later Academy? What did they mean to the scholastics? To Leibniz? To Hegel? To answer any such question involves the dual task of elaborating a reading and justifying a claim that the reading fits this or that historically given understanding.

With regard to historical interpretations, in the indicated sense of readings associated with historical claims, it seems unproblematic to speak of correct and incorrect interpretations. This does not imply, of course, that we always have the means to actually know if a given interpretation is correct or not. We do not even know who the original author of Dao De Jing was, nor if there is actually a single person behind the whole text, nor what the precise wording of that text would be. But that does not stop a hypothesis about the historical origin and original meaning of the text from being true or false, nor does it exclude that we can present evidence and arguments for and against such a hypothesis. There are situations where we are epistemically even worse off (think about stone-age archaeology, for example) but the fact that the evidence is meager and allows for a wide range of possibilities does not rule out that some interpretations correspond to how it really was, and some do not – nor does it rule out that new sources of evidence may come to light and help us decide which is which.15

But all interpretations are not aimed at historical accuracy, and for those which are not it is pointless to expect or demand that they be historically accurate. Clear examples are furnished by the performative arts. Contrast the task of a scholarly interpreter of Shakespeare with the challenges facing a director and crew trying to stage a version of The Merchant of Venice relevant to a modern audience. Or think about musical interpretation. For the interpreter of Bach, for example, the task is to take the written score and transform it into sound, a process involving numerous choices of

---

15With regard to temporally or culturally remote material we also have the question whether we have the conceptual resources to express something like the original understanding in our own language – such problems will loom large in chapter 6.
tempo, phrasing and so on. There are many different goals for which one can strive in the planning and execution of such an interpretation, and the result can be judged and appreciated on many different grounds. Usually it is the musical pleasure of real or potential listeners that are in the foreground, but in recent decades claims to historical accuracy has played a major role in the rendering of Baroque music – the use of original instruments has been one means in the project of reconstructing the original sound. Clearly, there is no contradiction between judging one rendering to be more historically accurate and simultaneously finding another one more aesthetically pleasing.

The contrast between interpretations aiming for historical accuracy and interpretations that have other more value-oriented goals is not relevant only for interpretations in the performative arts. In writings on the history of philosophy it shows up as the contrast between historically and philosophically oriented interpretations. In jurisprudence and theology, taken by Gadamer as paradigmatic interpretive disciplines, questions about historical plausibility are tempered by considerations of reasonable application to present situations. And literary critics, just like ordinary readers, are often more focused on finding something interesting in the text than on what the author may in fact have been after.

When interpretations are presented, the emphasis is normally on the reading, while the claims made on its behalf are left vague or contextually determined. Reconstructing a plausible claim, or a set of claims, for a given reading is a good exercise, but the result will often be a complex mix of historical claims and value-oriented claims. One complication is that different claims will often be relevant for different layers in the meaning sandwich. Typically, one will strive for historical accuracy with regard to the wording and linguistic meaning of a text, even if one simultaneously argues for a thematic interpretation or a generalisation mainly on the basis of their intrinsic value or interest.\footnote{It is an interesting question why we make such a distinction. If one is prepared to disregard the author’s intention on some levels, in order to have a more interesting reading, why not go all the way and adjust the wording and linguistic meaning of words as well? My own hypothesis is that those layers that are most governed by arbitrary conventions are, perhaps paradoxically, the least tempting to tamper with – as a word can arbitrarily be assigned any meaning at all, the exercise of actually doing so becomes entirely pointless. Extracting a moral lesson from a tale, on the other hand, is a thoughtful argumentative exercise where anyone can have her say, and where the reader can reasonably consider herself on a par with the author.}
CHAPTER 4. INTERPRETATION

4.6 Interpretation and rationality

It has often been said that interpretation presupposes the rationality of what is interpreted. Hans-Georg Gadamer has argued that interpretation starts from an “anticipation of perfection” (“Vorgriff der Vollkommenheit”) – including assumptions not only of the meaningfulness, consistency and coherence of the text, but even of its truth. In a similar vein, recent theories of interpretation in analytic philosophy has emphasized different versions of “the principle of charity”, obliging interpreters to prefer readings that minimize the amount of logical error and falsity it ascribes to whatever is interpreted. In a slightly earlier epoch, William Dray and others argued that reason-based explanation of action differs from causal explanation in having a normative content, infusing historical explanation with an element of rational justification – to explain an action is to show that it, at least in a certain sense, was the right thing for the agent to do.

It is important to note that these theories not only assert that actions and other objects of interpretation are rational as a matter of empirical fact, but that imputations of rationality is an apriori element in the very notion of interpretation – a view that must, of course, in some way be reconciled with the obvious fact that real people are very often not perfectly rational, and that real texts may be inconsistent, incoherent and false, without ceasing to be at least to some extent intelligible.

We will devote this section to investigate the place of rationality assumptions in interpretation, utilising a few representative views from different traditions as our point of departure. The main backdrop will be the hermeneutics of Gadamer, but I will also refer to a group of American philosophers in the analytical tradition, in particular to Donald Davidson, but also to Willard van Orman Quine and Daniel Dennett. One reason for making these comparisons is that it is worth noting the striking similarities between theories emerging within such different traditions, but another reason is that differences in scope and emphasis can provide mutual illumination and a fuller picture than each view by itself.

One pertinent difference has to do with what is taken by each as the paradigmatic examples of objects of interpretation. Gadamer’s guiding examples are high-status texts, particularly from the history of philosophy: the dialogues of Plato, the works of Kant and Hegel, and so on. Davidson, Quine and Dennett, on the other hand, take the interpretation of actions as basic, including speech acts, i.e. elementary actions of uttering something in specific circumstances. For simplicity, I will largely frame the exposition as if texts were the primary object of interpretation, but the underlying idea is that most of what is said should have a much more general field of application – problems and prospects for that generalization will be addressed as we go along.

17 Ref to Wahrheit und Methode.
18 ref
19 Ref to Davidson, Quine, Dennett
20 It is curious and noteworthy that he shares this predilection for the history of philosophy with several other of the most preeminent participants in recent debates on interpretation, for example Jacques Derrida and Paul Ricoeur – the possibility of generalizing from this basis to other kinds of interpretation, for example to the interpretation of literature, is seldom explicitly thematized and discussed.
4.6.1 Meaning holism and the hermeneutic circle

The founding insight of the hermeneutic tradition is the doctrine of the “hermeneutic circle”. Although obviously important, the basic idea is not very precise, and may be formulated in different ways, that often bears no more than a family resemblance to each other. As originally conceived, the hermeneutic circle referred to the mutual dependence between the meaning assigned to parts and wholes in interpretation. In reading a text, I go through the words and sentences comprising it, from the beginning to the end, and my understanding of the whole text must in some way depend on my understanding of these elements. At the same time, it is often obvious that my understanding of the parts depends on my understanding of the whole – the meaning of words and sentences depends on the context, as the saying goes.

Examples are ubiquitous. Words can have different lexical meanings, only to be disambiguated by the context. Pronouns, names and definite descriptions have their reference decided by contextual factors. Circumstances may force a certain expression to be read metaphorically, with the precise interpretation of the metaphor again to be influenced by context; whole episodes that could be read literally in isolation may be open to allegorical or symbolic interpretation within the context of a novel; what would only be a piece of factual information within a report or an inventory, may demand to be taken as a clue within the context of a crime story, and so on. Within a prolonged argument, an ordinary indicative sentence may be taken as a straightforward assertion, or as a supposition to be disproved through a reductio ad absurdum, or as a potential objection in a counterargument, to be rejected in favor of the main thesis – understanding how to take it in a specific case presupposes a grasp of the relevant argument as a whole. And while all of these examples highlight the dependence of (the interpretation) of parts on (the interpretation of) wholes, they also make obvious the dependence in the other direction: without understanding the words and sentences of the text, we would never be able to understand the story or the argument as a whole.

Obviously, the examples are not only manifold – and the list of similar phenomena could be extended indefinitely – but they are also very different, so different that superficial similarities could well turn out to be fundamentally misleading. Nevertheless, I will proceed on the assumption that there is something to be said on this general level about part-whole interdependence in interpretation, and only go into the differences as need arises.

One preliminary observation is that the doctrine of the hermeneutic circle may be taken in a process or a product sense. In the process sense, the doctrine says something about the activity of interpretation as it evolves over time, along the following lines. Interpretation always starts from a more or less rough idea about the meaning of the whole – of the kind of object or text one has to do with, and the kind of meaning it may have – and proceeds to read the parts in the light of that preliminary understanding. As one goes along, one either finds interpretations of the parts that fit one’s pre-understanding of the whole, developing it and supplying more detail, perhaps, but not contradicting or otherwise calling it into question; or one fails to come up with such harmonious interpretations of the parts, and is
forced to revise or even totally reject the anticipated meaning of the whole. In the latter case, one will eventually return to the details of the text with a new set of meaning anticipations, again encounter success or failure, and find occasion for new revisions and elaborations. Ideally, the process comes to an end in a final understanding that integrates a rich interpretation of the details with a plausible understanding of the whole. It is this understanding of the hermeneutic circle as a process, that has prompted some authors to criticize the metaphor of the circle as misleading – going around a circle you return to the same places again and again, but the process of interpretation never takes you back to exactly the same point, as you are constantly revising your interpretation of the parts and the whole, in the light of each other. The process of successful interpretation, the argument continues, does not resemble a circle so much as a spiral, scaling ever new heights for each completed revolution.21

There is no doubt that viewing interpretation as a process, guided by anticipations of different kinds and most importantly by the anticipation of a total meaning, is very important for Gadamer. But this does not exhaust the significance of the hermeneutic circle. Thinking only of the process is to view the circle as a feature of the context of discovery, as a positivist might frame it, and to expect it to fall away and to lose its importance in the result, in the finished interpretation. But Gadamer is adamant that the circle is not only a methodological tool, insisting instead that it also and more importantly has an "ontological" significance – that it is also constitutive of the result, of the meaning obtained through the process of interpretation.

In the product sense, there is no temporality to the hermeneutic circle, and hence no implication of a movement. In this aspect the doctrine of the circle expresses a form of meaning holism: the mutual dependence of the meaning of the parts and the meaning of the whole, even within the finished interpretation.

Here is a comparison with a jigsaw puzzle, that might be helpful. In an ordinary jigsaw puzzle, the search for a solution is guided by a preconception of the final picture: as you go along you guess where this or that piece might fit in the projected result, which in this case maybe printed on the box, but which you may also just see in your mind’s eye. By trial and error, you test your suggestions, and when you are ready, you have the satisfaction to see all the pieces tightly locked in place and forming the anticipated picture, in glorious detail. Now, for an ordinary puzzle, when the last piece is in place the projection of the result has played out its role – once the puzzle is laid, the configuration of the pieces determines the overall picture. But suppose instead that there were several different solutions to the

21The emphasis on process is also behind Dagfinn Føllesdal’s (1994) well known thesis about the equivalence of the hermeneutic circle and the hypothetico-deductive method. All research, according to Føllesdal’s description of the hypothetico-deductive method, involves a to and fro movement that starts from a preliminary hypothesis, which one proceeds to test by means of empirical evidence, which may in turn force one to revise or reject the hypothesis, before going back to test the new hypothesis against further evidence, and so on. Taking the overall interpretation as the hypothesis and the struggle with the details of the text as the confrontation with evidence, there is an obvious similarity between the two “methods”, but the equivalence thesis also neglects many other aspects in which the interpretation of “meaningful material” may have substantial specific traits in relation to other scientific pursuits, and, in particular, it neglects the product sense of the doctrine of the hermeneutic circle.
puzzle, that one and the same collection of pieces could be laid either as a pastoral scene or as an urban landscape, for example. In such a case, the anticipation of the whole would not have played out its role even when the puzzle is finished, it would amount to a fundamental decision, that partly determines which role each piece will play in this particular outcome. This is how Gadamer views interpretation – as a puzzle where the result can vary not only with the pieces at hand but also with the projection of the result, with the important proviso that the relevant projection will seldom be the result of a conscious choice, and more often determined by preconceptions beyond the conscious control of the interpreter.

It may be worth our while to dwell a little longer on the analogy. Once one admits the possibility that the same set of pieces may be used to form different pictures, one is naturally led to think about possible differences of degree with regard to the constraints that the pieces may impose on the overall image. At one extreme we find the ordinary jigsaw where there is only one possibility, and the whole is ultimately determined by the parts. At the other extreme we might imagine a set of totally malleable pieces, that may be pressed to serve any preconceived total image, without imposing any constraints at all. In between, we would find cases where the nature of the parts would admit some different solutions but exclude many others, so that the "interpreter" might be frustrated and have to revise her preconceptions, perhaps many times, before being able to fit the given material into a satisfying picture. Clearly, it is some such intermediate scenario that Gadamer envisages for real life interpretation.

Taken in this way, the doctrine of the hermeneutic circle is connected to a sort of relativism in interpretation, where the outcome of even the most meticulous effort to ascertain the meaning of something will to some extent depend on contingent choices or features of the interpreter. We will come back to the precise nature and extent of that relativism, below, but will just note in passing that it is one reason why Gadamer thinks that the circle is a suitable metaphor for interpretation. There is something in the process and results of interpretation that associates it to circular definitions and circular proofs – through an elaborate procedure one seems just to extract at the end what one has inserted at the beginning. It remains to show why some such circles may be good rather than vicious.

4.6.2 Pre-understanding

On a general level, the hermeneutic circle seems to be a simple enough idea, but, as we will see, almost everything that Gadamer has to say about interpretation is already contained in it, as a precondition or a consequence. One such consequence, that we have already touched upon, is the importance of pre-understanding. To even begin the process of interpreting anything, you must already have some understanding of it.

Pre-understanding comes in several forms. The sort of pre-understanding most often emphasized by Gadamer has to do with the anticipations of total meaning that gives the basic context from which to start interpreting the elements of a text. But of course, to properly enter the hermeneutic circle it is just as important that one has a preliminary understanding of the elements that make it up, and what they can be expected
to signify outside the present context. Without an independent idea of the meaning of the parts, they would put no constraint on the resulting total interpretation, and there would be neither circle nor spiral to the process of interpretation – just the tiresome repetition of the anticipated meaning of the whole.

Simple examples are furnished by ordinary textual interpretation. The modern reader of a Platonic dialogue, say the *Theaetetus*, would usually have a rather good idea of what to encounter: expecting it to be a piece of philosophy, belonging to the ancient Greek tradition, perhaps expressing views we would recognize as Platonic. Armed with this pre-understanding the reader is prepared to classify pieces of the text as questions and suggested answers, arguments and counterarguments, refutations of competing views, and so on. But obviously one would not get very far without also knowing classical Greek: being able to read the alphabet, recognizing the words and having a good idea about what they mean and how they fit into various syntactical patterns. If one is a sophisticated reader, one will be prepared to challenge almost any part of these assumptions as one goes along, including the inherited Greek lexicon – maybe some of the entries in it are based on misunderstood passages from exactly this work – but one could not get going at all without assuming something on each of these levels, and more.

Such anticipations are closely tied to the individual work that one is trying to interpret – about what kind of work it is, what kind of meaning or message it may be expected to convey. But Gadamer also talks of pre-understanding in a much broader and more general way, referring to the whole set of categories, opinions and habits of thought that the interpreter brings to each encounter with something to be interpreted. For pre-understanding in this more general sense, Gadamer often uses the metaphorical term *horizon*, for example when he describes the goal of an interpretation as a “fusing of horizons”. To continue the example of a modern reader of Plato, on this level she would not only bring specific preconceptions about Plato, his aspirations and his language, to the act of reading, but also her own understanding of what philosophy is, what types of questions and answers that count as philosophical, an understanding that, in turn, is rooted in a total view of knowledge and the world, presupposing modern variants of the differences between science, religion, philosophy and art, for example.

Clearly, the importance of such horizons is not empty speculation, but an empirical fact that helps to explain otherwise puzzling facts about interpretation. It has often been observed that each generation seems to need its own version of the classical philosophers, for example. Every few decades, there seems to be a new Plato stepping out of the shades, not only adjusting his views in relation to his predecessors of the same name, but also changing the direction of his philosophical interest: having been absorbingly concerned with epistemology during the early decades of the 20th century, his attention seems to turn more and more to the philosophy of language after the World War II, and lately he seems to be swinging back towards his traditional preference for metaphysics. Clearly, such swings are better explained by changes in the agendas of readers than by actual posthumous vacillations on the part of dead authors.
There is one more thing to note before we leave the notion of preunderstanding, for the moment. One word that Gadamer likes to use in this connection is “Vorurteil”. A neutral and literal translation of it may be “pre-judgement” – a fitting term for the preliminary opinions that the interpreter brings to the process of interpretation. But the standard English equivalent is “prejudice”, and this is also what the word means in ordinary German. Now, this is not an association that Gadamer shirks. To the contrary, he takes one of the important lessons of hermeneutics to be that a prevailing empiricist epistemology has brought the prejudice too much into disrepute. He does not deny, of course, that some prejudices are harmful and inimical to the acquisition of knowledge, but the correct response to that is not to try to get rid of prejudice altogether and to start from a clean slate – as empiricist philosophers sometimes seem to believe – but simply to be on one’s guard and try to spot the bad prejudices as they make their effects felt. On some level, this is surely right – new knowledge is always acquired against the background of old knowledge – but Gadamer’s way of expressing this also points to a pervasive streak of conservatism in his general outlook, as he never tires of emphasising the importance of deferring to tradition and authority.

4.6.3 Relevant totalities

One question made urgent by the doctrine of the hermeneutic circle concerns the nature of the relevant “wholes” that are supposed so be so all important for the interpretation of the parts. In order to properly bow to the necessity of starting the process of interpretation with an anticipation of the ensuing total meaning, one must also anticipate what is the relevant totality to which the parts belong.

In many cases, the appropriate totality seems to be more or less self-evident. Words belong to sentences, sentences belong to paragraphs, paragraphs belong to chapters, chapters belong to books, and so on. But the more one thinks about it, the trickier the issue becomes. As we will see, differences on this point make up a large part of the disagreement between the traditional hermeneutics represented by Gadamer and its postmodern critics, like for example Jacques Derrida.

One preliminary point demands clarification, because it is obscured by the talk of words and sentences. Words and sentences are primarily types: the same word and the same sentence may occur indefinitely many times, as uttered or written by indefinitely many authors. But questions of context concern tokens or occurrences rather than types: we are interested in the utterance or inscription of a word as being part of the utterance or inscription of a sentence, rather than of abstract words being parts of abstract sentences. Similar provisos apply throughout the following discussion – in time we will also come back to the distinction between the internal, linguistic, context of an utterance, and its external context, or situation.

In Gadamer’s work there is not much discussion of what constitutes a relevant whole for interpretation, and the general impression is that he thinks within the framework of what we may call the romantic hierarchy. Central to that way of thinking is the author – the unity of the text derives from the unity of the producing subject. The basic interpretive entity, in
terms of which smaller and bigger unities can be defined, is the work: the novel, the story, the treatise, the poem etc., which is conceived, created and issued by its author as one. Below the work there are subdivisions intended by the author as such: paragraphs, sections, chapters, episodes, and the like. Above the particular work are collections of stories or poems made by the author, and, most importantly, the collected works, the oeuvre of a particular author, always available as a hermeneutic resource. And still further up in the hierarchy we find the context provided by an epoch, and a literature carried by a nation or a language, until we reach what is, in all cases considered by Gadamer, the ultimate context for any serious work belonging to intellectual history: the Western tradition reaching back to the Greeks.

Calling this the "romantic" hierarchy is, of course, meant as a reminder of its roots in the aesthetics of romanticism, where all meaning is taken as expressive of a subject, preferably a genius, but where the individual subject is nevertheless taken to be essentially rooted in a historical tradition, in the first instance associated with a language and a nation.

There are, of course, numerous ways in which this sequence of purportedly natural unities may be, and has been, questioned. One way is to point out that the notion of an author is itself a historical and conventional construction, that has little application outside certain specific epochs and genres. As an illustration, we may consider the request made by Swedish writer Lars Norén in an epilogue to his eleventh book of poems, that he should from this book on be considered "a new author". He has, according to what he says, considered changing his name or using a pseudonym, but finally decided to content himself with simply stating his decision and asking for the compliance of the reader. In one sense, of course, he is simply wrong – his later books are (hopefully) written by the very same Lars Norén as his earlier works. So what does his gesture mean? Presumably, that he does not want his later work to be read or interpreted in the light of his earlier production – and in demanding of his readers that they shall abstain from that contextualization, he highlights that the formation of such hermeneutic totalities is contingent and a matter for decision.

The notion of the author in literature has its counterpart in the artist of the other arts, who we tend to think of in much the same way – as the unifying mind behind the work, disregarding the possible use of assistants and readymade content as inessential. And even in cases where the fit is obviously very uneasy, as in the notoriously large scale team-work that goes into the making of a movie, we try to apply the same schema – casting the director as the artist, and preferring "auteur" films, where the identity of the artist is not in doubt, as objects for highbrow interpretation. But again, other epochs have had different ideas, concentrating on the resulting object rather than on the creative mind – the workshops of early modern artists like Rembrandt or Rubens were more like trademarks than they were guarantees of individual provenance, hence the modern problem of establishing the "authenticity" of individual works. And once you start looking for them there is still no shortage of genres where the notion of an author as the source of an oeuvre makes little sense – just think of copywriters in adver-

---

22 And what about works from other traditions? We will come back to that.
interpreting or the speechwriters of politicians.

In the more general case of the interpretation of action, the idea of an author has its counterpart in the notion of an agent, as the person responsible for what is done. The delimitation of the relevant totality for the interpretation of a particular action is both intricate and interesting. A vivid case is provided by law courts, that are routinely asked to infer the motives and underlying beliefs of particular actions. To do so, they have no other choice than to place the relevant act in the context of other actions by the same agent, and of what would be a reasonable interpretation, from the agents point of view, of the situation at the time, including his or her interpretation of the behavior of other agents. Without any such context it is simply impossible to question or evaluate what the agent now says that he or she believed or intended.

One of the founding fathers of the hermeneutic tradition, Friedrich Schleiermacher, suggested that the correct totality in which to evaluate any individual meaningful expression is the entire psychological life of the agent. This is a natural idea, but it is only viable under certain conditions that are worth thinking about. The context can only influence the interpretation of a particular expression under the presupposition that it is connected to the expression in certain ways, that it forms a unified whole to which the particular expression belongs. Applied to Schleiermacher’s suggestion, the presupposition seems to be that an individual action can only be interpreted to the extent that it is performed by an agent who is not just rational in the moment, so to speak, but whose whole life forms a rational totality. Take a simple example. I understand somebody to be lying, on a certain occasion, by comparing what he says then to what he says on other occasions, and to his actions. Perhaps he tells me that the ice is safe after having warned other people of it and then refuses to go out on it himself – my conclusion is that he does not really believe what he tells me, i.e. that he is lying. But perhaps he changes his mind a lot, so that he actually believed that it was safe when he told me, but believed that it was dangerous shortly before and shortly afterwards? Clearly, unlimited recourse to such possibilities would endanger the whole idea of a reasoned interpretation of somebody’s actions. On the other hand, it would be just as devastating to conclude that hermeneutic principles demand that nobody ever changes her mind, so a compromise between meaning-destroying fickleness and total ossification must be found. Again, there is a natural suggestion for the compromise: a rational, i.e., interpretable, agent only changes his or her mind for a reason, as a response to new information or further thought, in a way that can itself be understood by the interpreter.

So Schleiermacher’s suggestion leads to an intriguing conclusion: the interpretation of an action presupposes the idea of a rational agent, an agent controlled by reason and in control of her actions. And, of course, saying, writing and painting things are themselves actions, so the notion of a rational agent would by the same argument be fundamental to all interpretation. The problem, of course, is that the assumption seems to be obviously false – we often act on impulses that go against our ordinary ways and our reasoned intentions, we change our minds for no apparent reason, perhaps just forgetting that we ever believed otherwise, we make simple mistakes in our reasoning, we behave in incompatible ways in different circumstances,
and so on.

If we assume, at least for the moment, that there is a strong connection between interpretability and rationality, we can sketch four different ways to react to the apparent irrationality of real agents – reactions that, not by accident, correspond to major currents in the theory of interpretation. Let us take them in turn:

The first option is to deny the irrationality, to argue that it is only apparent, and that there is an underlying reason even to actions and beliefs that look arbitrary or irrational on the face of it. A good example of this strategy is Freudian psychoanalysis. Despite its persistent reputation for attacking the notion of a unified rational subject, psychoanalysis is in fact in staunch defense of it, against the seemingly overwhelming counter-evidence. It is true, Freud admits, that we often act against what we avowedly believe and desire, and even against what we ourselves think that we believe and desire. But this just goes to show that our real beliefs and desires are inaccessible to us – underlying our conscious thoughts, feelings and behavior is a coherent system of unconscious beliefs and desires that makes perfect sense of what we say and do. As we would expect, this broadening of the scope of the rational brings along a corresponding broadening of the realm of the meaningful and interpretable – dreams, slips of the tongue, episodes of forgetting and other seemingly meaningless phenomena become keys to unlock the secrets of the self.

Psychoanalysis is not alone, however, in taking this route. Another example would be the rational choice theory associated with modern economics. According to this theory, agents act so as to maximize their expected utility, roughly to best satisfy their preferences in the light of their beliefs. Again, purported counter-examples are not taken as falsifications of the theory, but as evidence that the agent’s real beliefs and preferences are not what we, or even the agent herself, would otherwise take them to be. Both theories impress their followers by yielding seemingly powerful and sometimes surprising explanations of what people do and say, but they also face the same challenge from their opponents: that in their hands the hermeneutic circle indeed becomes a vicious circle, that the interpretation of the elements is always moulded so as to fit the preconceived shape of the whole.

The second option, is to admit that reality does not always live up to the hermeneutic demand for rationality, but maintain that the rationality assumption is nevertheless a useful idealization, a theoretical model, perhaps in analogy to the frictionless surfaces or perfectly elastic particles one may encounter in a physics book. Agents are interpretable in so far as they are rational, and a lot of less than rational behavior may best be explained as deviations from a rational model. Max Weber’s notion of understanding as based on “ideal types” belongs in this camp, as well as Dennett’s theory of “intentional systems”. As we will see, both Gadamer and Davidson may also be understood as advocating this option, where full rationality is taken as conceptually basic for the notion of understanding, while less than rational meaning phenomena are accommodated as explainable deviations from that ideal.

A third way to react to the conflict between the demands of rationality imposed by hermeneutic principles and a recalcitrant reality is to
completely separate them. Understanding and explanation, reasons and causes, belong to different ways of approaching reality, and, properly understood, they should never be mixed. The reason for this is often held to be that understanding is inherently normative, to give the reasons for somebody’s action, is in some (perhaps rather weak) sense to show that it was justified, at least from the agent’s point of view.23 For the domain of historical explanation a view of this kind was forcefully argued by William Dray and the idea was popular among English speaking philosophers in the Fifties.24

All of the first three solutions to the problem of understanding the irrational agree that the notion of a unified rational subject is basic to interpretation and understanding, they just differ about its status: is it a real entity, a useful model or a normative ideal? The fourth option is much more radical, and takes the subject entirely out of the equation.

Claude-Levi Strauss, a towering figure in the first generation of French Structuralism, has described how he finds it hard to remember the content of his own books, because he never has the feeling that he is actually writing them. Instead, he has the experience of the tip of his pen being merely a focus point where all the influences of his life come together, to spread out again on the paper, in new combinations, and then to dissipate further into the lives of his readers. One may perhaps doubt the absolute sincerity of this avowal, but it is a nice emblem of the structuralist and post-structuralist notion of the subject. Instead of being an autonomous agent, fusing all of my experience into a synthetic whole before expressing it again in my work and action, I am merely a transit hall where content from different sources shows up for a while before traveling on to other destinations – interacting and changing in the process, to be sure, but never being reassembled into a harmonious whole with my stamp on it.

This perspective leads to a heterodox approach to textual interpretation, perhaps first evinced in the work of Roland Barthes, but most completely championed by Jacques Derrida. Such readers do not deny that classical texts are, so to speak, built to give the impression of a meaningful unity under the control of the author, but take this subjective unity to be a mere rhetorical device, used to gloss over the cracks underneath, that bear witness to the actual etiology of the text.25 The canons of traditional hermeneutics, though valuable up to a point, make the reader an accomplice in this cover-up, helping to suppress other and perhaps more subversive messages that might be heard by an ear more attuned to dissonance – relishing the value of pieces that do not really fit the puzzle, to return to the previous analogy.

23Perhaps a common-sense example can be used to introduce the idea. In a domestic argument, the question “Why did you do it?” is often best answered by coming up with a good reason to do it, rather than by delving into a messy account of what actually brought it about.

24It received a famous drubbing by Davidson in his 1963 paper “Actions, Reasons, and Causes”, now in Davidson (1980), from which it never fully recovered. The introduction to von Wright (1971) still gives a vivid and useful background to the debate.

25Once this perspective is taken on board, it opens the possibility of texts that abstain from even the pretense to unity, wearing their fragmentary and disorganized nature on their sleeves. While this idea is being championed, and to some extent practiced, by Barthes and Derrida, it, perhaps paradoxically, leaves little leverage for their specific modes of interpretation.
CHAPTER 4. INTERPRETATION

4.6.4 Relevance relations

We have so far talked rather vaguely about the elements of a text comprising a “whole”.

4.6.5 Perfection and charity

4.6.6 Radical interpretation and language learning

4.6.7 Dialogue and triangulation

4.6.8 Imperfect rationality

4.6.9 Horizons and relativism
4.7 Interpretation and translation

4.8 The death of the author?
Ontological questions – questions about what there is and about the nature of reality – have played a an increasingly large role in methodological debates within the cultural sciences for the last few decades. It has even been suggested that a distinct, more relativistic, conception of reality is one of the defining traits of the cultural sciences. The pivot for this discussion has often been the notion of a “social construction”, and here I will take that notion as a clue, structuring this chapter and the next around different interpretations of it.

The span for different conceptions of social construction is, in fact, very large. One useful way to start thinking about that span is with the help of the titles of two seminal books that have done much to shape the field. One of them is, of course, *The Social Construction of Reality* by Peter Berger and Thomas Luckmann (1967). Although relying heavily on earlier traditions, rooted in German idealism, Marxism and phenomenology, it sparked what became known as “social constructivism”, an expression that soon became a sort of centre towards which related ideas from other sources began to gravitate. The title suggests a thesis about all reality, saying that it is socially constructed.\(^1\) The other book is *The Construction of Social Reality* by John Searle (1995). Searle chose his title as a deliberate contrast to Berger and Luckman’s, vigorously denying that all of reality is socially constructed, and concentrating instead on social reality, which he argues is indeed a construction in a specific and interesting sense which he goes on to specify.

Neither of these models actually covers all of the purported examples of social construction, and in particular they fail with regard to those examples that have generated the most controversy. The first model is too inclusive, and not able to accommodate the specific features that might warrant treating some phenomena in contrast to others as social constructions. The second model is too domesticated, and not able to account for the moral and political edge of the discussion. But they are clear-cut enough to serve as foils for more complicated cases, where other ingredients enter the mix, and where we will have to refer to other sources, like Judith Butler, Michel

\(^1\)The book itself does not quite endorse that claim in its most radical form, being, in fact, rather careful to distinguish a “quoted” notion of reality from a “philosophical” notion, and disclaiming having anything to say about the latter. But such *finesse* has largely gone unnoticed in the subsequent debate.
CHAPTER 5. FRAMEWORKS AND RELATIVISM

Foucault and Ian Hacking.

I will start at the general end, beginning with Immanuel Kant, the Godfather of constructivist metaphysics. Kant invented modern constructivism, and although his brand of the doctrine is neither social nor relativistic he laid the foundation on which later more radical constructivists have built – or, perhaps, he started digging the hole that they have turned into an abyss. After following some of the ramifications of the Kantian heritage in anthropology and the philosophy of science, I will turn to Searle's notion of a social fact, and then try to see what must be added to that idea to account for the juicier cases discussed by more radical thinkers.

5.1 Relativism and constructivism

Questions about constructivism, about the different ways in which different kinds of things depend upon the human mind – upon our concepts, beliefs and attitudes – are often conflated with question about relativism, and in particular about ontological relativism, i.e., relativism about what there is. There are, of course, important connections between relativism and constructivism, and we will comment on them as we go along, but as I am using these words there is nothing inherently relativistic about constructivism. That money, to take an uncontroversial example, is a social construction does not stop it from being objectively real and amenable to scientific investigation. There would be no money, to be sure, without human thought and action, but as there really is human thought and action there is plenty of room for money to be real, too.

5.2 Reality as a construction

The world as we know it consists of objects that have properties and stand in relations to each other – spatial, temporal and causal relations being among the most important. Relying on skeptical conclusions that David Hume had derived from empiricist premises, Kant argued that such structural features of the world could never be given to us through the senses. If sensory input was all there is to experience, as empiricists assume, we would know nothing about causality and would not even have an inkling that there is a world of objects in space and time, beyond our fleeting sensory states. In fact, there would not be any knowledge at all, because we would have no way to distinguish between our impressions and what they are impressions of.

Kant never questioned that we need experience to learn the details of how the world is, but in order to have experience in the first place we must already have a general outline of the experienced world. Experience is, so to speak, the projection of our subjective states onto a preconceived scaffolding, a structure a priori that we bring along and actively impose on the deliverance of our senses.

In the complete Kantian picture there are different components to that a priori structure but the most important thing for us is the role of concepts. What is a concept? We will need to return to that question again and again, below, because it is the hub around which the whole debate about
constructivism turns. The function of concepts that first springs to mind is, presumably, that of classification: we collect many similar things under one concept. We think of many four-legged animals of a certain size, shape etc. as belonging together and we collect them under the concept of a horse. But concepts have other functions as well, and most central to Kant is not classification but integration. We need concepts not only to collect given unities into classes, but to make it come about that there are any unities in the first place – the Kantian word for this is “synthesis”.

Coming into a room where I have never been before, I throw a glance around me and immediately notice many of the things that are there: chairs, tables, a pile of magazines, a few pairs of shoes near the door. I hear music, and I have no problem identifying a piece of stereo equipment as the source of it. But what is a stereo? Two loudspeakers, an amplifier, a radio-tuner, a CD-player, maybe a turntable if you’re a vinyl buff, and of course the cables of different types that connect them to each other and to the outlet in the wall. What is it that makes all this diverse stuff into parts of one thing? Well, the concept of a stereo, presumably. Without having that concept, for sure, we could not see them as one thing, and if nobody had that concept they would not be one thing.

As the example shows, the process of conceptual integration works both ways: parts are collected into unities and unities are divided into parts. The lawn contains a football field, which, in turn, has two halves, two penalty areas with their six-yard boxes, a centre circle, and so on. One deck of cards is 52 cards, 13 different values, 4 suites. It seems rather obvious that without the relevant concepts there would be no such things to be counted. And the same thing goes for temporal division and integration. Is that the same stereo that you had three years ago? No, it was stolen, but I liked it so much that I bought another one just like it. Yes, I just had to get a new CD-player when the old one broke down. Without concepts, things could not be identified, differentiated, re-identified or counted – in short, there would be no things. “No entity without identity”, says Quine, and to have identities we need concepts.

You might be tempted to express the upshot of such considerations as an existential dependence of things upon their concepts, as if things of a certain type must wait for the relevant concept to exist. But you have to be careful not to fall into absurdities here. In one sense, perhaps, there would be no trees if there was no concept of a tree, but in another sense, of course, there were trees long before there were any concepts. Later on, we will need to distinguish between different kinds and degrees of dependence of things upon human thinking, because this is where most of the action is, in the hassle about social construction, but for now we just note that nothing of what we have said is meant to imply that the world was not there before us.

Kant emphasized the role of concepts in shaping our experience. To see dogs, stereos, trees and puddles of water in the world, I need to have the relevant concepts. But he did not think, of course, that the entire conceptual structure that we bring to bear upon our experience is a priori in the

---

2The examples are meant to get the idea across, not to argue for it in any rigorous fashion, of course. The stereo, by the way, is a favorite example used by Nelson Goodman, for example in (1984).
sense that we must possess it before we are able to have any experience at all. Whatever formative use we ultimately make of them, concepts like dog, stereo, tree, puddle and water are empirical concepts, that are in some sense abstracted from experience – they are learned, and learning is itself an experiential process. But learning takes place in a world that is already there for us, a world of objects that we learn to distinguish and relate in ever more subtle ways as we trot along, and for that to be possible we must have some concepts to begin with. For Kant, these primary concepts are the very abstract ones that define the notion of an object in space and time – a something to which I can direct my attention.

5.3 What is reality?

We have pointed to the role of concepts in the construction of reality, but so far we have said nothing about what it is to be real. We have concepts not only for real things, but also for unreal things. Some of them maybe fictions, like Santa Claus and Donald Duck, or the hobbits, the elves and the orcs of Tolkien’s stories. Some of them have once been thought to be real, and only later discovered not to be, like ghosts and phlogiston. And some concepts, no doubt, that we now take to correspond to realities will be discovered by later generations to be as empty as the concept of a ghost.

Empty concepts, concepts that have no real application, is one thing. But ever since Plato first gave us the concept of a concept, philosophers have also been interested in another discussion concerning the reality of the conceptual. Some of the classifications that we use seem to be more real than others – they seem to “cut nature at the joints”, as the saying goes, while others seem to be arbitrary inventions, imposing distinctions that may be interesting to us but have little to do with what is really out there.

Let us take another of Nelson Goodman’s favorite examples: star constellations. What shall we say about the Big Dipper – is it real or is it just our invention? Well, it is not a fiction, of course. We readily identify it in the sky, we teach our children to see it, we have a tradition of observing and talking about it that goes thousands of years back into history. And it is not a non-entity either, like phlogiston: when I point out the Big Dipper to my children, I’m not wrong. But even so there still seems to be a question left about the reality of the constellations, where the answer is not as clear cut – they are real, alright, but are they really real?

The Big Dipper comprises seven stars that we bring together under one name. But from the point of view of the world, so to speak, our criteria for picking out and collecting precisely these seven stars seem disturbingly arbitrary and with very little connection to the stars themselves: perceiving them from the angle and the distance where we happen to be located, we can, with some benevolent effort, see them as forming a pattern similar to a cup with a handle. We make those seven things into parts of one thing, but in reality they have nothing to do with each other. Without us – our point of view, our interests, our history, our ways to perceive and think, these seven stars do not belong together more than any other seven stars anywhere in the universe.
Let us, before we stop to think some more about this, go on with another example. Are there stars? Intuitively, I think that we are more inclined to answer yes to that question than to the former one. Stars are more real than constellations. One reason for that intuition, I suppose, is the asymmetrical ontological dependence between them: there may be stars without constellations but not the other way around. But we also have the feeling that the stars are more natural unities than the constellations.

But should we trust that feeling? Can we not say exactly the same things about a particular star that we said about the Big Dipper? It is a collection of many smaller things, atoms, that we take as parts of one thing. But exactly where in space and time shall we draw the border between those atoms that belong to a certain star and those which do not? Is it not all up to us in this case, too: our ways of seeing, our interest in details only down to a certain level, and so on?

And what about the atoms - are they more or less real than the stars? Here we seem to have at least two different ways to think about things, that pull us in different directions. On the one hand, we have the asymmetrical ontological dependence that seems to make the components more real than the collections – there may be atoms without stars, but no stars without atoms. On the other hand, we have that thing about naturalness: we see stars with the naked eye; when we get close enough, as with the sun, we feel the heat and the light emanating from them – stars have a robust reality for us that is hard to deny. Atoms seem much less concrete: millions of people go through life without ever suspecting their existence, our reasons to believe in them are very indirect, they are “theoretical” entities.

5.4 Two criteria

I said above, that we owe the concept of a concept to Plato, and he is also the first to ask for a criterion to decide which of our concepts correspond to realities. In the dialogue The Sophist he lets the main protagonist suggest a criterion that we may call the criterion of causality, which he formulates something like this:

To be real is to affect and be affected by other real things.

How would our three examples – the constellations, the stars, the atoms – come out under this criterion? The constellations would not fare well, in full accordance with our intuitions. Science has no use for the constellations, when it comes to explaining things. No real events anywhere in the universe are causally influenced by the Big Dipper, no theories of astronomy or physics refer to it. The stars seem to come out as much more real, according to the criterion. Among many other things, stars are dominating sources of gravitational attraction, and such is the influence of gravitation that we have strong reason to believe in the existence of stars that we cannot otherwise observe, because we need them to causally explain the behavior of things that we can observe. The atoms come out even better: we use them to explain all kinds of things, and thanks to the asymmetrical ontological dependence, we can even avoid talking about the stars and see them as mere aggregates of atoms.
As we see, the criterion of causality is closely linked to science. It is more or less tantamount to the idea that science decides what is real: those entities that we need to assume in our basic scientific theories of the world are the ones we take to "really" exist.

The causality criterion will be the main foil for our discussion of social reality, but we should also note the other criterion that we touched upon in the previous section. We talked about "naturalness", about robust realities being present to our senses: what we see, feel, hear, smell and taste. When we are in a scientific mood we realize that there is more than a touch of arbitrariness about using the human senses as an ultimate standard of reality: the information they are attuned to is only a few slices of what is actually available, and the way they present it, with the qualitative distinctions between the sensory modalities, for example, seems idiosyncratic to say the least. But it is hard to doubt the tangible presence of perceptual reality for very long, and there is a sense in which it really seems to be fundamental: whatever we come to know about other things must in some way be connected to the evidence of our senses. And it is not all about knowledge: the things that we see are also the ones that we handle, the primary objects and instruments of our bodily actions. Through a microscope or a telescope I can see the very small or the very large and distant, but the instruments themselves are of the same scale as my body. As Quine puts it: what is paradigmatically real for us are the "medium sized physical objects" that we immediately observe and act upon. Just to have a label I will refer to this way of grading reality as the "observability criterion".

From this perspective many of the things that came out on top according to the causality criterion, seem to have only a "derived" type of reality. We may even come to think of the fancy entities of science -- the elementary particles, the genes, etc. -- as little more than useful fictions, things that we postulate or assume in order to account for what we observe and handle, but that actually borrow whatever reality they possess from the "life-world" in which we immediately live and act.

5.5 Intentional mediation

Let us go back for a moment to the constellations and the causality criterion. Is it really true that the constellations have no causal influence on other things? What about astrology -- some people run their whole lives according to a theory of the constellations and their influence on us.

There are two observations to be made here. One of them is just a passing remark, while the other one will be important for what follows. The passing remark is that belief in astrology is a sort of confirmation of the causality criterion. What people who believe in astrology believe, and what people who don't believe in it deny, is precisely that the constellations have a causal influence on us -- the imputed reality goes with the belief in the causality.

Given that astrology is actually false, as it is, the more important consideration concerns the influence that the constellations have on human affairs through the mediation of people's belief in them. Just as you may be afraid of ghosts even though there are no ghosts, and let that fear prevent...
you from sleeping in a certain house, say, you may avoid setting out on a
trip on a certain date because you think it is an unlucky day, or you may
refuse to marry someone because he is a Sagittarius.

In previous chapters we have used the word “intentional” as a common
label for mental states that have a content, states like belief, fear, hope etc.
Using that vocabulary we may distinguish between two kinds of causal “in-
fluence” that something may have. On the one hand we have direct causal
influence, the type of influence that the causality criterion is about. On the
other hand we have intentionally mediated influence, where the actual cau-
sation belongs to an intentional state that has the relevant phenomenon as
part of its content – and, as we know, the intentional object of an intentional
state does not have to exist.

Sometimes, intentional mediation, as in the case of ghosts and astrology,
is through false beliefs, but there are other cases as well. A falling rock may
 crush me – a clear case of direct causal influence – but if I see it coming I
may manage to get out the way, thus adjusting my behavior after the rock
through the intentional mediation of a true belief.

Intentional mediation will be at the center of our analysis of social con-
struction, and in particular we will come back to the notion of nested inten-
tionality that we used to analyze meaning phenomena in chapter three. In
the process we will be looking for other classes of influential beliefs, apart
from the straightforwardly false and the straightforwardly true – namely
types of beliefs that are in a sense “made true” by the fact that we have
them.

But before we come back to that we have some things to add to our
background of Kantian constructivism.

5.6 Reality as a social construction

This chapter is about constructivism and relativism, but so far we have
focused only on constructivism. What about relativism – the idea that dif-
ferent things may be real from different perspectives, that reality depends
on what “conceptual scheme” one employs?

As I said, Kant’s own brand of constructivism is neither social nor rela-
tivistic. The forms that we bring to experience, and which makes it possible
for us to encounter a world, are built into the human subject as such. Indi-
vidualism and universalism go hand in hand: we are all capable of building
the world on our own, but being essentially similar in epistemic respects
each of us will build essentially the same world, which we can then go on
and investigate together.

Kant is an Enlightenment thinker and both the universalism and the
individualism are basic to the Enlightenment credo, but after Enlighten-
ment comes Romanticism and already during Kant’s lifetime relativistic
versions of his scheme were being proposed. Conceptual relativism – which
is our subject for the moment – comprises two main theses: one about de-
pendence and one about variation. It is the dependence thesis that Thomas
Kuhn wants to express by saying that “concepts shape the world to which
they are applied”, and so far Kant is still with him. To have relativism we
must also think that we may have different concepts that shape the world
differently, and that conceptual systems vary with some interesting factor, like language, culture or history.

In various forms, relativistic constructionism has been enormously influential during the last two centuries. Roughly, we may distinguish two main currents in the river of relativism. On the one hand we have brands of ethnological and historical relativism, where the units, so to speak, are the life-worlds of different cultures or historical epochs. On the other hand, we have scientific relativism, where the source of variation is supposed to be something like a “paradigm”, carried by a scientific group. At least in some of its forms, scientific relativism is a special case of ethnological relativism: the scientific group is taken as a small society or “tribe”, and the paradigm is taken as a specification of its specific culture. Whatever the similarities, when thinking about science the distinction is important: ethnological and historical relativism are about the subject matter of ethnology and history, while scientific relativism is a thesis about the scope and limitations of science itself, which at its most radical may even amount to a form of skepticism.

5.7 What is relativism?

As has already been hinted, there are many forms of relativism. Different relativisms may be wildly divergent with regard to their content, their plausibility and their importance, but they all share a certain form: they involve the claim that something X is relative to something else Y:

\[ X \text{ is relative to } Y \]

So in order to describe a specific form of relativism, we should begin by indicating the relevant values of X and Y – what is supposed to be relative to what. Short names for brands of relativism, however, usually indicate just one half of the relation, leaving it to context or background assumptions to supply the other half. Speaking of "moral relativism" we usually think of morality as the X being relative to some unspecified Y, while "cultural relativism" indicates that some unspecified X is relative to culture as the Y in the formula.

A better practice would be to always mention both parties to the relation. Subjective moral relativism, would then be the claim that what is morally right or valuable (X) is relative to the subjective moral opinions of individual agents (Y) – so that for each agent there would be no difference between thinking that something is morally right or good, and for the same thing actually to be right or good. If we keep the same X, but take general acceptance within some cultural group as Y, instead of individual opinion, we get cultural moral relativism as a result.

As I said, the main topic for the rest of this chapter will be something that I will call conceptual relativism, but there will also be some remarks on

---

3There is a whole literature of ethnological studies of science, mostly following the tradition of being based on “field work” in a small local community, letting a laboratory or a department play the traditional role of a village in anthropology. Perhaps the most famous (or infamous) work in this genre is Latour and Woolgar (1986).
epistemological relativism. In the light of what we just said about the general form of relativistic claims, both of these labels are ambiguous. Conceptual relativism obviously involves concepts and epistemological relativism involves knowledge – but on which side of the relation? Do they play the role of the X or of the Y? Is conceptual relativism the claim that concepts depend on something else, or that something else depends on concepts? Both alternatives are possible, and indeed prevalent. It is for example often claimed that concepts are relative to language – that concepts depend upon and varies between different languages – a thesis that may also be called linguistic relativism, after the other side of the proposed relation. But when Kuhn asserts that "concepts shape the world" he is obviously pointing in the other direction: it is the world, what there is, which is supposed to depend upon and vary with different concepts – to emphasize the other side of this relation we may speak of ontological relativism, but the full name should rather be conceptual ontological relativism. To complicate matters even further, relativistic claims may also be chained, so that one thing (X) may be asserted to be relative to another (Y) which in turn is relative to a third thing (Z) – ontology may be relative to concepts which in turn may be relative to language.

The same sort of ambiguity and possibility of chaining is present in talk about epistemological relativism. On the one hand, an epistemological relativist may be taken to assert that knowledge is relative to some Y – for example to a culture, a language, an episteme, or a scientific paradigm. On the other hand, epistemological relativism may be taken to assert that something else – usually truth or the way the world is – is relative to knowledge. On the analogy of conceptual ontological relativism, we might call this latter doctrine epistemological alethic relativism. Or epistemological relativism may be taken as chaining the two possibilities: truth is taken to be relative to a language, culture or paradigm, by mediation of knowledge. In summary, we are interested in two, possibly interdependent, forms of relativism: conceptual ontological relativism which claims that things, what there is, depend on concepts, and epistemological alethic relativism which claims that truth, the way things are, depends on epistemological principles. In both cases, the relevant doctrine derives it interest from the further claim that concepts or epistemologies vary in an interesting way with social or historical circumstances.

5.8 What is a concept?

To fruitfully discuss conceptual relativism, we need to say a little more about concepts. We have already talked about concepts, when discussing intentionality in chapter 3, in particular in sections 3.5.1 and 3.5.2. The main focus there was to introduce and understand the kuhnian thesis that

---

4In recent debate the term conceptual relativism has largely been associated with the work of Hilary Putnam, who obviously uses it to mean ontological relativism in this sense.
5Such is, for example, the relativism of Benjamin Lee Whorf (1956).
6What is at stake here, of course, is not what people actually happen to know, which trivially varies with factors like the above, but the principles of knowledge – what is sometimes called "an epistemology" – which are supposed to determine what can be known in a certain context.
7After the Greek word for truth: aletheia.
"concepts shape the world", and so the emphasis was on the role of concepts in perception. Now, when we want to assess the implied relativism of that thesis we need to say a little more about the role of concepts in thinking.

5.8.1 Words and concepts

We usually think about concepts in connection with words, and in particular with descriptive words like common nouns, adjectives and verbs. We take the concept of a tree to be connected to the word ‘tree’, we take the word ‘purple’ to stand for the concept of purple, and the concept of running to be indicated by the word ‘run’. But the relationship between words and concepts is not straightforward or always one to one. The same concept may be covered by different words, in the case of synonyms within one language or between different languages – in French the word corresponding to the concept of a tree is ‘arbre’ and in German it is ‘Baum’. Inversely, ambiguous words stand for a variety of different concepts, sometimes totally unrelated, as in the ubiquitous example of the English word ‘bank’, sometimes connected in more or less obvious ways, as when we use the word ‘run’ about running water without implying that water has legs. Some concepts require complex linguistic expressions, as in ‘better than average Swedish marathon runner’ while other concepts possibly lack linguistic counterparts altogether.

In chapter three, I sketched the distinction between the reference or the range of application of a word, roughly, the things in the world that it is used for, and the meaning of the word, equally roughly, the way that those things are characterized by the use of the word. We must now take that distinction a bit further, and describe both sides to this distinction in more detail.

5.8.2 Concepts: classification

One of the most important and obvious dimensions of concepts is that they are used for classification, to group things together. We will talk about the set of things to which a concept applies as the extension of the concept.

Normally, we presume that such groupings are not arbitrary, but that there is some shared property, a similarity or even an “essence”, that connects everything within the extension of one concept. Sometimes we take that shared property as being obvious at first glance – red things fall within the extension of the concept red precisely in virtue of being red; and sometimes we take it to be the task of science to discover what the underlying essence of a classification is – different puddles belong to the extension of the concept of water by virtue of sharing the same chemical composition. But, as we shall see, this presumption is not always fulfilled, and there is often room for controversy over how natural or arbitrary a certain classification is.

As an antidote to the presumption that classifications are natural, Michel Foucault is often quoted as quoting, in turn, a reference by Jorge Luis Borges to a “certain Chinese encyclopedia”:

---

8 Distinction between reference and denotation.
In its remote pages it is written that the animals are divided into: (a) belonging to the emperor, (b) embalmed, (c) tame, (d) sucking pigs, (e) sirens, (f) fabulous, (g) stray dogs, (h) included in the present classification, (i) frenzied, (j) innumerable, (k) drawn with a very fine camelhair brush, (l) et cetera, (m) having just broken the water pitcher, (n) that from a long way off look like flies. As Foucault immediately observes, the problem is not the subclasses themselves, but precisely the idea that they are subclasses of anything, that a single concept can play the host to such a comic disarray of otherwise unconnected phenomena. Borges goes on to draw the radical conclusion that “there is no classification of the Universe not being arbitrary and full of conjectures”.

---

9Reference to Foucault and Borges. Borges, in turn, credits the discovery of the encyclopedia to doctor Franz Kuhn, who is a real person, versed in Chinese literature, but not known from any other source for precisely this piece of information.
Chapter 6

Social construction

The previous chapter was about the thesis that all reality, at least in so far as it is accessible to human knowledge, in some sense depends on the conceptual and epistemological frameworks of the knower. We will now turn our attention to what looks like a much more restricted notion of social construction, which says about some types of phenomena that they are social constructions in a way that other things are not. It is important at the outset to note that these types of constructivism are not just restricted versions of the relativized Kantianism that we have been discussing until now, but rely on very different ways that facts, what there is, can depend upon the human mind.

We may put the difference this way. The Kantian line of thought makes reality depend on the cognitive resources of the knower: on the concepts and the epistemological principles that the knower brings to experience. There can be no stars, atoms or witches except in relation to a cognitive framework containing the corresponding concepts and the criteria for applying them. But the framework itself does not decide whether anything really fulfills the criteria – just like the fact that I have the concept of a unicorn does not guarantee the existence of unicorns, the mere possession of the concept of a star or a witch does not bring stars or witches into being. Reality, according to both absolute and relativized Kantianism, does not depend on what I believe. To the contrary: in seeking knowledge I try to adjust my beliefs to reality, by mediation of my concepts and epistemological principles.

Social constructions, in the stronger sense, are things that really depend for their existence upon what people believe about them, or what they take to be true about them – they are things that we, so to speak, create by thinking, we think them into existence by believing that they exist.

We will approach the notion of a social construction with the help of John Searle’s notion of a social fact, starting with uncontroversial standard examples, like games and money. Instructive as it is, it is fairly obvious that Searle’s theory does not account for most of the controversy surrounding the notion of social construction. I will argue, however that many of the more controversial cases can be illuminated by comparison with Searle’s simple model, by noting differences and similarities, and adding further ingredients as we go along.
6.1 Social facts

Let us think again of a game of football. The referee makes a call for offside, and let us suppose that he is right. What has he seen? He has spotted a certain pattern in the distribution of persons in differently colored jerseys on the field, at a certain moment. The pattern itself is a perfectly objective fact, accessible to anyone who would bother to look for it, and possible to register with a camera. But why would anybody care?

We care, of course, because the relevant pattern is endowed with a special significance by the rules of football. Play is stopped, a free-kick is given, possession of the ball is shifted to the other team. The pattern is a brute or natural fact, as Searle would say, while the offside is an institutional or a social fact. The pattern counts as a case of offside, within the context of the game. In the same way a wooden piece of a certain shape counts as the queen in a game of chess, and a certain constellation of five playing cards counts as a full house in a game of poker. This is the structure, according to Searle, of all social facts, they can all be captured by expressions of the form:

\[ X \text{ counts as } Y \text{ in the context } Z. \]

The context \( Z \) is supposed to be some human activity or practice. The practice is defined by a structure of rules and values, that are tied to the relevant facts: there are rules for what happens when an offside is given, for how the queen can be moved on the board, for the value of a full house in relation to other poker hands, and so on. The material support, so to speak, of the queen in chess is to a large extent arbitrary, but it is important that there is some way to tell which piece is the queen and to be able to distinguish it from the pawns, the bishops, and so on.

It is important to note, however, that the supporting facts are not arbitrary in the same way and to the same degree in all cases. The form of the queen is only limited by practical considerations of size and weight, and chess may even be played entirely in the head, without any material support. A deck of cards must admit of random shuffling – poker can be played on a computer but not in the head. And think about the actual ball in a game of football: it has an important social role in the game, of course, that in a concrete case distinguishes it from the idle balls waiting by the sidelines for their fifteen minutes of fame, and there is room within the social role for some variation in its natural properties, its exact weight and size for example – but it would hardly be a good idea to replace it with a heavy metal cube.¹

Games are particularly simple and instructive examples of social facts, and we will come back to them in what follows, but Searle’s own favorite example is money. Just like chess-pieces, the currency of a certain country must be recognizable if it is to be of any use, but it is not the brute properties

¹The same thing goes, mutatis mutandis, for the different roles of the players within the game plan, so to speak. There is nothing either in the rules nor in nature that precludes assigning any role in the team to any player, but it may still be more effective to use a quick but short, left-footed player on the left side of midfield, rather than in the center of defense or as a goal keeper.
– the material, the weight, the size, the engravings etc. – of coins and bills that makes them into money.

As a first stab, we might say that a certain piece of metal is turned into money by being seen as money, by a certain collective belief that it is money. To be eligible for the belief it must have certain natural properties (including a certain history, even perfect counterfeits are not really money), but the belief goes way beyond the natural properties. Without any brute change whatsoever, something may cease to be money overnight, by a political decision or by people simply losing confidence in it.

As I said, it is important to note the difference between this kind of dependence on our minds and the kind that is involved in Kantian constructivism. Even if there is a sense in which it is true that something may be a dinosaur only in relation to our concept of a dinosaur, the dinosaurs did not have to wait for us to come around, in order to exist – they even managed to become extinct long before the concept was invented. But money, chess-queens, poker-hands and offsides really had to wait for us to think them into existence. They depend not only on our conceptual and epistemological resources, they depend on what we actually think.

For Searle, it is also important that social facts presuppose natural facts – not every fact can be a social fact in this sense. Why not? For the counts-as formula to work it must be possible to pick out the X antecedently and independently of its being Y – the whole point of the formula is to connect the game-facts to the world outside the game, so to speak.

But even if we admit that there must be brute facts for there to be social facts, it is also important to note that brute facts that carry social facts tend to be highly artificial, and so in another sense not “natural” at all. Coins and playing cards are manufactured to have the “brute” features that we use to recognize them, and even if you may apply the offside rule in a football game played on an ordinary lawn with jackets for goal-posts, or play chess with pebbles of different sizes and colors, the relevant natural properties would never be salient except in the context of the game. One of the things that separates more controversial cases of social construction from the well-behaved examples of money and games, is that the relevant brute facts seem less artificial, and more naturally tailored to serve as foundations for the relevant social properties.

6.2 The power of shared belief

When discussing reality criteria we distinguished two forms of causal influence: direct causality and intentional mediation. The real causal factor in intentionally mediated influence is an intentional state, normally involving belief. We also pointed out that both true and false beliefs have real

---

2 This difference seems obvious enough, but there are still those who deny, or perhaps overlook it. Bruno Latour (1998) famously denies that Ramses II could have died of tuberculosis because the disease was not “invented” before the 19th century. This should be reformulated – Latour’s actual position is less clear and perhaps more sophisticated.

3 It is not the case, however, that every fact that can play the X role must be a natural fact. Think again of a poker game: this money counts as the stake in the context of this game. But it seems natural to assume, and thereby to concede Searle’s basic point, that such chains must ultimately be anchored in something that is not itself a social fact.
effects, and that even things that do not exist, by means of this detour via the mind can have real, albeit intentionally mediated, “effects”. It seems clear that the effectiveness of social facts is of the intentionally mediated variety: money does what money does by way of us thinking about it in the way that we do. The specific difference between social facts and the earlier forms of intentional mediation seems to be that the relevant beliefs are neither independently true nor independently false: they are made true by the fact that we believe in them. But how is this done?

We must not exaggerate the creative powers of beliefs, of course. I cannot make a piece of metal into a coin by thinking that it is money – it has its social status largely independent of my individual mind. To be effective, the belief itself must be socially shared in a certain way. It is for example not enough that we all think that this is money, we must also think about each other that we all think that it is money, and we must possibly think that we all think that we all think that this is money, and perhaps we could go on in this way forever. Social facts not only depend on our intentional states, they depend upon complex forms of nested intentionality.\footnote{The basic idea behind this way of thinking about the social goes back at least to Hegel (1807), but the modern discussion of shared beliefs and its role in establishing social facts was started by David Lewis (1969).}

There are also important distinctions within the field of social facts. Money, in its fully developed form, does not depend only on free-floating conventions, in the sense of shared belief, but is tightly regulated through specific formal institutions, even if these institutions themselves ultimately depend on being recognized as legitimate by the relevant group of people. A clearer example of a convention in its pure form is language. The meaning of a word is determined by its use. I can be individually wrong about what a certain word of my native language means, but we cannot all be wrong: collectively speaking the word means what we think it means. And my individual usage takes part in deciding what is collectively correct – if a usage that is originally a misunderstanding spreads far enough it simply becomes the norm.

Clearly, the possibility of creating social facts by collective thinking is very useful to us, and permeates our lives. But how did it come about? Reasonable speculation, I think, points in two different directions.

The first direction is towards make-believe and imitation games. To make-believe that you are something that you are not and to make this the foundation for an activity of cooperative play is very similar to establishing a social fact. There is just one difference: make-believe does not make the pretense come true. No matter how deeply they go into the fantasy of Star Wars, and how skillfully they yield their plastic light sabers, my youngest boy and his friends will never make it real that they are Luke Skywalker or Darth Wader - nor would they ever think so. But there seems to be a sort of gradual transformation from, for example, fictional war-play to the game of chess. When a child learns to play chess the fantasy of the war may be an important part of the fun, and gradually the reality of the game takes over and becomes its own end.

The second, and perhaps more important, speculation points towards the intentional efficacy of false beliefs. Human societies are complex networks of positions and relations, defined by institutions that for the outside
observer clearly meet Searle’s criteria of the social, and that in reality rest on nothing but the shared belief that this is real and this is how it ought to be. But as we know, this has seldom been enough, but important social institutions are usually surrounded by powerful mythologies, providing anchor-points outside the social – in nature or in the divine. Kings are thought to come from sacred lines, anointed by God and perhaps with divine ancestors. Slaves are thought to belong to inferior races and being by nature incapable of caring for themselves. Human rights are not only something that we recognize and are willing to grant to each other – we take them to belong to us by nature.

The thought that our social institutions rest on nothing but shared belief and common trust may be scary even if you are in total accord with how the institutions works – just think about the sense of vertigo engendered by the recent financial crisis. If the institution is itself unfair and biased to serve the interests of some at the expense of others, it will be all the more in need of a natural alibi and the resistance to give up the mythology will be even stronger.

To see social institutions as social institutions is not easy or natural for us, one might say, but a difficult and perhaps never-ending task, begun by the Sophists in ancient Greece, continued by the thinkers of the Enlightenment, and inherited by the social sciences when they first came into being. A radical social constructionist, as we shall see, is a person who thinks that we still have a long way to travel down this road.

### 6.3 Is gender a social fact?

Probably the most discussed and most hotly contested examples of social construction have to do with sexual identities, and in particular with what it is to be a woman or a man. Can this controversy be illuminated by Searle’s notion of a social fact?

The received way to approach the ontological status of sexual identities is through the distinction between sex and gender, and at first blush it seems tailor-made for Searle’s formula. Sex is supposed to be a natural fact, you are born a female (“une femelle”, as Simone de Beauvoir has it), while gender is a social role, you are not born but has to become a woman (Beauvoir; 1949). So is this all there is to gender being a social construction?

Sex X counts as gender Y in the context Z.

There is something to this, of course, and you can apply Searle’s formula to a host of similar cases, in each case highlighting what really seems to be an important aspect of the respective phenomenon. Motherhood being a social construction is partly captured by the formula:

Giving birth counts as becoming a mother in societies Z.

– where the concept of a mother comprises the specific set of rights, responsibilities and attitudes that come with giving birth in the relevant societies. In a similar way being an adult is a specific social status that comes with
the natural fact of achieving a certain \textit{age}, and so on. In each case we have to distinguish a set of brute properties on which a certain society has chosen to hang a set of socially enforced norms and attitudes – very often, of course, without making any terminological distinction between the two, as in:

\begin{quote}
Being made of gold (natural substance) counts as being gold (social status) in societies Z.
\end{quote}

As far as it goes, there is doubtless something illuminating in this way of looking at things, but it is also clear that it leaves at lot of important differences between the different cases in the shadows.

Let me introduce a potentially misleading but suggestive piece of terminology here, and speak of the social role as a \textit{meaning}, something \textit{signified}, which is being attached to the underlying brute fact as a \textit{signifier}. In a certain cultural context giving birth “means” becoming a mother, in a way that is to some extent similar to the way that the brute properties of certain pieces of metal “mean” that they have certain values, in relation to a certain monetary system. Speaking this way, it is clear that a coin is a \textit{conventional} and \textit{arbitrary} sign, in the sense that it is nothing other than a social convention that links precisely this signifier to its meaning, the value. And it seems about as clear that the relation between giving birth and motherhood is not arbitrary in exactly the same way. There are several pertinent differences here, and I will point to two of them.

The first difference is \textit{conceptual inclusion}. Even if becoming a mother involves a lot more than giving birth to a child, giving birth is still a \textit{part} of what it is to become a mother, in a way that carrying a picture of George Washington is not a part of the concept of being a dollar bill.\footnote{I am not saying, of course, that we \textit{must} have a concept of motherhood that works this way, or that such conceptual connections are not amenable to change. With new pro genital techniques, the connection between the social roles of parenthood and different aspects of the biological process may well weaken to the extent that the traditional linguistic usage will seem inconvenient and cumbersome. After all, in connection with adoptions, for example, we already make the distinction between being the “biological mother” and being the mother in charge.} In itself this has nothing to do with the nature/culture distinction: the same thing is true of the relation between kicking the ball and taking a corner. Taking the corner just is kicking the ball in a certain socially defined context, and it could not be accomplished in any other way.

When we discussed meaning in Chapter 3, we distinguished between thick and thin descriptions of an action, using “he took the corner” as an example of a description that is convention thicker than “he kicked the ball”. In the same vein, we can distinguish between thick and thin concepts: the concept of a corner adds layers of meaning to the concept of kicking the ball, and the concept of motherhood adds extra layers of meaning to the concept of giving birth. Thick concepts are economical, of course, allowing you to say much in few words, but they are also \textit{persuasive}, as they allow you to abstain from explicit assertion in favor of conceptual presupposition. Instead of explicitly \textit{saying} that someone that has given birth has specific duties towards the child, including the duty to feel in a certain way, etc., we
pack it all into the concept of motherhood. The effect is that a whole package of norms and theory that we connect to giving birth is silently brought to bear on the case at hand, and it takes a conscious effort to foreground, and maybe to question, specific parts of it.\footnote{There is also a specific type of \textit{ambiguity} associated with thick concepts, allowing us to slide imperceptibly between thin and thick interpretations of a word in the course of an argument.}

The second difference is \textit{motivation}. Let us start with an uncontroversial example: the connection between the social role of gold and the eponymous natural substance. There is no necessity here, of course – something other than natural gold might very well fill the social role of gold, and natural gold might very well have been culturally indifferent to us. Nevertheless, gold has some natural properties that make it fitting for the social role: it is comparatively scarce, it is malleable to a suitable degree, it is relatively immune to oxidation and other forms of corruption, and it is aesthetically pleasing to us.

This is, of course, where the real action is, in most debates about social construction. Is there anything about giving birth to a child that motivates assigning the specific set of duties, rights and attitudes that come with motherhood to the very same person that gave birth? Is there anything about having male or female sex that makes one especially suitable for the social roles traditionally assigned to men and women? The traditional answer, I presume, is a resounding yes, but it is precisely the foundations for this yes that are questioned by social constructionism.

\subsection*{6.4 Kinds of rules}

Searle ties the difference between brute and institutional facts to another important but elusive distinction, between two kinds of rules: constitutive and regulative rules.\footnote{Searle’s most elaborate treatment of the distinction is in (1995) but he discusses it in many places, going back to (1964) and (1965). It is formulated in almost exactly the same terms by Midgley (1959). John Rawls (1955) makes a similar but not exactly equivalent distinction between “summation rules” and “practice rules”.} The basic idea is that regulative rules tell you how to perform activities that exist prior to and independently of the rules, while constitutive rules define or constitute the activities they apply to. Paradigmatic examples of regulative rules are rules of table manner, like “hold the fork in your left hand” or “chew with your mouth closed”. The activity of eating is obviously possible without those rules – one can eat in accordance with other rules or without adhering to any rules at all. Paradigmatic examples of constitutive rules are game rules – the rules of chess do not regulate an antecedently existing activity of playing chess, to play chess simply \textit{is} to act in accordance with the rules, and if you change the rules it is no longer a game chess.\footnote{The possibility of tweaking the rules of the same game, different versions of the same game etc. – a vagueness to be ignored here.}

Now, it may be thought that strict adherence to this distinction would undermine the analogies between clearly institutional facts, as in the case of games and money, and other alleged cases of social construction, like gender and motherhood. Is it not obvious that being a mother is a brute
fact, independent of all norms, and that the rules of motherhood are “only”
regulative rules, norms about the proper ways for mothers to behave? Let
us take a closer look.

The only examples that Searle gives of constitutive rules are counts-as
rules of the type we have been discussing above, and you easily get the
impression that all constitutive rules are of this form. The reason that
he concentrates so exclusively on counts-as rules is, presumably, that they
seem to be the special locus of the ontological creativity of institutions – it
is by means of them that brute facts are transformed into social facts. But
all the rules that define a game or an institution are obviously not of this
form. Here are some examples from chess:

White makes the first move, and after that the players take
turns.

Bishops may only be moved diagonally.

When the king is checked, it must be moved or protected so that
it is no longer checked. If this is impossible the game is over.

The board is a grid of 8 x 8 squares in two alternating colors,
called “white” and “black”.

When we say that chess-facts are defined or constituted by the rules of
chess, we are referring to the whole system of rules. To say of a piece of
wood that it is the Queen means that it has a certain role in a game, relating
it to all the other pieces, to the board and to the allowable sequences of
moves. The rules of chess actually say nothing about which real things are
queens or bishops, but only about what permissions and obligations that
are allowed in relation to them once they are taken as such. We may think
of the ordinary rules of chess as internal rules giving rise to internal game-
facts, like the fact that the King is checked on a particular occasion. That
a certain real piece of wood is taken as the King in a particular game, in
contrast, is an entrance convention – it sets up a prerequisite for the game
but it is not part of the game as such.9

It is a special feature of chess that the entrance conditions are very
loose and not even regulated by the game rules. More common, in the case
of games, is that the rules give necessary conditions for entrance: to be al-
lowed as the ball in a game of football an object must have a certain shape,
size and weight, but it is still a decision in relation to a particular (physical)
ball that this is the match ball of a game – other exactly similar balls lying
around the pitch are not “dead”, as the match ball would be in the same po-
sition, but simply not part of the game. One of the things that distinguish
social practices that are not games from games, in the literal sense, is that
there are sufficient conditions for entrance: if something fulfills the condi-
tions for being money, as specified by the rules of the monetary system, it
simply is money, there is no further decision to be taken. The same thing
is true of traditional gender systems – to be a woman (gender) is not an
option for a person of female sex but an obligation; and for judicial systems

9The distinction between being the Queen of an actual game, and being the Queen as part
of a chess set.
– if someone fulfills the brute conditions for being guilty of a specific crime he not only may but should be treated as such.

Are the rules of the penal system regulative or constitutive? Are they regulations of a prior and independently existing practice or do they constitute a new form of social reality? Not easy to say. On the one hand, the penal system clearly gives rise to roles and offices that are only defined and so “constituted” within the system itself – the judge, the defense, the prosecutor, the police, the jailor, etc. - and many crimes themselves presuppose specific social institutions, like the institutions of property. On the other hand, it seems pretty obvious that penal systems historically arise as regulations of pre-existing practices of revenge and corrective treatment, and the notion of punishment extends well beyond the confines of legal systems, for example into the family. Even Searle’s paradigmatic example of money is less clear-cut than he makes it out to be. The full blown institution of money can be seen as part of a system that regulates the antecedent activity of exchanging goods and services, and it is rather obvious that it historically arose in precisely this way, through many intermediate stages, involving the use of specific goods, like gold, as general intermediaries.10

In the same way, the conventions of motherhood can be seen as regulating an independently existing practice of caring for children, by, among many other things, specifying realms of responsibility for mothers, indicating the proper attitudes for mothers to take in different circumstances, etc. But the way this regulation works, is precisely by setting up an institution with conventionally defined roles for particular people to identify with, to conform to, or to resist. The autonomy of the institution is shown by its cultural variation, by the external and internal pressure it exerts on people to conform to it, and by its incorporation into the meaning of the relevant labels – we have no problem with understanding the stereotyped announcement in a novel by a father to his rebellious offspring that “you are no longer my child”.11

6.5 Fetishism, reification and alibi

In a famous chapter of Das Kapital, Marx discusses the relation between the real nature of the (exchange) value of a commodity and our everyday conception of it. In reality, the value of a thing is a social relation, having to do with the readiness of people to exchange different kinds of goods in definite proportions. But we do not experience value in this way, but instead take it as a real property of things – instead of seeing through the notion of value to the social relations packed into it, we tend to see the relations as natural consequences of the thing being valuable. Value is, according to Marx, turned into a fetish, something with magical powers to engender

---

10 As mentioned in a previous footnote, the idea of constitutive rules goes back to Rawls (1955), but his contrast is not with regulatory rules but with “summation rules”, which are rules of thumb, generalizing principles of good or correct behavior from specific instances. Standard examples of regulatory rules, like rules of etiquette or traffic regulations, are obviously not summation rules in this sense, and it is not clear whether Rawls would count them as ‘practice rules’ or would take them as lying outside the classification.

social relations.\footnote{This process, of relations being “reflected” as properties of things, is a major theme in the philosophy of Hegel.}

Marx’s idea is taken up by George Lukács (1923) and made the basis for his concept of \textit{reification} (Verdinglichung).\footnote{Honneth (2005) is a sensible (perhaps a little too sensible) discussion of Lukács’ notion from a modern perspective.} To reify something is to make it into a thing, and for Lukács there is a strong tendency towards reification in human thinking, and particularly in capitalist societies. There are two facets to reification. One of them is the reduction of relational complexity: the tendency to see relations as \textit{properties}, and complex relations as simpler than they really are. The other facet is the tendency to see what is really intentional and social as \textit{natural}. Commodity fetishism, in the Marxian sense, exemplifies both features: we project the social relations of an exchange economy into the thing as a property, and we view that property as a natural reality, something that explains the social role.

Why would we have such a tendency? What makes reification an attractive option for us? There have been many answers to this question, some of them on a deep existential level, but in the present context it is natural to return to Roland Barthes’ (1957) notion of the \textit{alibi}, that we touched upon in chapter 3.\footnote{The existential themes I am thinking about is Heidegger’s (1927) discussion of “Verfall” and “Das Man”, and Sartre’s (1943) notion of “mauvaise foi”.} To let a semiotic relation masquerade as a natural connection absolves the sender from being responsible for his utterance, and from the burden of argument. In the same way, a natural fact does not need legitimization, it is just there to be reckoned with, while a social institution is liable to questioning and may have to be defended by explicit arguments.

This, of course, is the next component in thinking about gender as a social construction: it is not just a social fact, in Searle’s sense, but a \textit{reified social fact}. And the reason for its reification is, presumably, the need to deflect questions about legitimacy. Game institutions do not need legitimization through reification because taking part is voluntary; the institution of money does not need it either, because it is obviously useful for everybody and not in itself obviously unfair. But gender relations, at least in most societies, are obviously unfair and so in need of a rational justification that seems hard to come by – what could be more convenient than to let nature step in and cover it all in a cloak of necessity?

But can it not be true that the natural connection really is there? Look at motherhood, for example – is it not rather probable that there is a “maternal instinct” which is in some sense a natural foundation for our cultural stereotype of motherhood? And is this not an empirical question that cannot be settled by conceptual speculation? Sure, but there are at least three comments to take on board before leaning back with a sigh of relief in the bosom of Mother Nature.

a) The ubiquity of the stereotype means that it is easy to take it as more natural than it actually is – there just is no control group unaffected by it. This means that the relevant empirical investigations will be difficult, yielding only ambiguous results, which are always liable to be contested.

b) Motivation by causal factors is not itself direct causation. Maybe there is a natural tendency to violence in humans, that makes it in some
sense “natural” for us to hit our children when angry – is that an argument to allow or encourage us to do just that? The purpose of many social institutions is, precisely, to curb natural tendencies and channel behavior in directions that go against our reflexes.

c) When a purported natural fact seems to support an unfair social arrangement, there is a strong bias towards believing in it, on the part of those who benefit or derive support from the arrangement. Wishful thinking is one of the strongest mechanisms of belief formation, and contravening wishful thinking is a correspondingly important function of science.

6.6 Identification

We are trying to get to the heart of some controversies about social constructions, by means of a comparison with Searle’s relatively uncontroversial notion of a social or institutional fact. In the last section we added the component of reification, saying that a social construction is a social fact masquerading as a natural fact. In this section we will look at one more thing that distinguishes, for example, sexual identities from money – namely precisely their being “identities”. A coin has no inkling of the social status that we are attributing to it, but part of my being a man is the fact that I identify myself as a man, and that the awareness of my gender influences the way that I am and the way I behave. There are two facets to identification in this sense: the first one is the self-application of the relevant concept, the other is its function as a model or ideal in relation to which I see myself.

Self-application is a widespread phenomenon, of course, in the sense that anything that may be true of me is in the range of what I may believe about myself. But some concepts also have the more interesting feature that they cannot be true of someone unless that someone applies them to herself. I know my approximate weight, but I would weigh the same if I did not. But could I be a university professor without knowing it? It may happen, of course, that I have just been appointed and have not yet heard the news, and so on, but to actually take up the responsibilities and start performing the actions that come with the job is impossible without applying the relevant concept to myself. In the same way, to be a police officer is to have certain rights and duties that you cannot exercise unless you know that they apply to you.

In Searle’s terminology we may say that being a police officer or a professor is a social fact that one must recognize as applying to oneself for it really to apply. I must identify myself as a professor in order to be one. And identifications of this kind usually come with a set of more or less implicit norms that go much further than the minimal definition of the role. Being a professor I also want to be a real professor, and so I will form myself in accordance with, and sometimes in conscious opposition to, a certain socially recognized ideal of my profession. I will strive for expertise in my subject, I will try to be impartial and fair in relation to students, and, less consciously, I will perhaps take up some bodily habits and speech patterns that I associate with my status: a certain way to walk, with a more self-assured busy gait than before, and talk, cutting down on swearing, for example.
That being a professor is a social status is evident to all, but the same mechanisms are at work in relation to less obviously social identities. At birth, you are declared to be a boy or a girl — a quick visual inspection is usually enough to settle the matter. From that moment, this classification will shape how others treat you, in countless respects: what kind of name you get, how you are dressed, what toys you are given, how your hair is cut, how and when you are taken up and held, how one reacts to your displays of aggression or tenderness, etc. And as soon as you are capable of it, you will be made aware of what class you belong to, and of your duty not to stray too far away from the stereotype of your sex — real girls (boys) don’t do that! And soon you will become your own most diligent supervisor, taking care not to fall short of the internalized ideal of your sexual identity — most of the time, of course, without consciously thinking about it, and just doing what comes “naturally” to you. The way you dress, smile, move and talk is shaped by the stereotype until it becomes second nature. Being a man, to go out wearing skirts and a top leaving the shoulders bare is not just something that I choose not to do, but it would take the utmost determination and courage even to try.\footnote{Hacking (1986) is about the way in which new kinds of people are “made up” by the invention of new labels to identify with.}

The debate about social construction is about the border between nature and culture in us all. In a way, it seems evident that there must be such a border — after all, we are animals shaped by society, there is no point to deny that double nature — but where is it? Take being drunk, as another example. Alcohol is a drug that has definitive physiological effects on the human body, a little different in different persons, depending on genetic factors and habituation, but clear and measurable effects, all the same. If you drink a certain amount, you will get drunk — but is being drunk simply an alcohol-induced physiological state? Rather obviously not: the cultural variation of behaviors associated with intoxication is much too wide for that. The bodily state may be the same, but the range of actions and reactions that it makes available to the drinker and his or her associates is bound by social norms, that differ between societies and with social context (and that also depend on the drinker’s social status in other respects, notably gender). Being culturally drunk is a social role available to drinkers in virtue of having drunk a certain amount of alcohol. To take up the role you must identify yourself with it: you must believe that you are drunk and know what the stereotype allows and demands from you. And, of course, it is essential that the role is reified and not taken as a role — the whole point of being drunk is that it to some degree absolves you from responsibility, that it is not really you but the drug that makes you do what you do.

It is tempting to draw out the Alexander sword here and cut the Gordian knot. Sartre, at least in some of his moods, would argue that the category of causation is simply not applicable to human action. To view the natural setting for your actions as a set of causally determining factors is a type of self-deception or “mauvaise foi”. Natural factors only influence actions by detour of their meaning for the agent, they are part of his “situation” — which is not simply a factual background, but the product of the agent’s “interpretation” from the standpoint of his “projects”. There seems to be an
important insight here, but it can easily be taken too far. Being drunk is not just a role that you take up – you may get drunk against your will, and playing drunk is not an easy thing.

### 6.7 The body as artifact

So far in this chapter, the emphasis has been on the contrast between natural and social kinds of things. Roughly, natural things are things that do not depend for their existence on human intervention, like trees and mountains and bacteria, while social things are things that can only exist by being socially recognized, like money and university professors. Now we must turn to a third category of things, artificial things, or artifacts. Like social things, artifacts are the products of human activity, and therefore in some sense of human thought and planning, but they do not depend for their existence on social recognition.

One is tempted to think of artifacts as in a sense ontologically intermediary between natural and social things. With natural things they share a certain robustness – with luck, they can survive their makers and the social context in which they came into being, and be rediscovered by archeologists thousands of years after their original production. With social things, they share the dependence on human activity and purpose – to know what they are, one must relate them to a human context of makers and users.

In general, artifacts are partly defined by their function. To be a chair is to be an one-person sitting instrument, to be a toothbrush is to be a brush for cleaning the teeth. Not all functions are as clear-cut as these, and the function itself is often not enough to determine a kind of artifact, but the general idea seems right – artifacts are things that we make for purposes, because we need them in order to do something. This does not imply that only artifacts can have functions, of course. Even if we disregard biological functions (like the function of the heart to pump blood), there is the fact that natural objects can be taken up and given functions in relation to human activity – I may use a sharp piece of stone for cutting, or keep afloat by clinging to a suitable piece of wood. Indeed, the whole idea of an artifact being a thing that is “made” for a certain purpose is a bit misleading: we do not have the power to make something out of nothing. What we do is to take something that is already there and modify it to suit our purposes better, where the degree of modification may go from almost nothing to complete transformation.

Above, we noticed that standard examples of social facts are supported by artificial facts, by things that are expressly manufactured to carry the relevant roles, like coins and chess pieces. At first blush, this seemed to be an important point of difference in relation to more controversial cases – like gender, where the social role is supported by the biological sex of the natural body. But is this the right way to look at it? As, among oth-

---

16 Natural and social kinds - is mountain a natural kind? Probably not, but that does not stop mountains from being natural things.

17 But are not some social things robust in the same way – do archaeologists not find money in the ground, just like they find arrowheads and necklaces? Well, no, actually they do not find money, they find only coins – certain artifacts typically used as money.
ers, Michel Foucault (1975) and Judith Butler (1993) have forcefully emphasized, even our bodies are not completely natural things, but heavily modified in different ways to suit our purposes and ideals – indeed, human bodies are themselves artifacts of a kind.

This is obviously true on the level of what we may call bodily design. Plastic surgery is the most radical form of bodily design, but there are less drastic ways to modify the ways our bodies look: from gym work and diets to hair-cuts and shaving. We usually think of such measures as “aesthetically” motivated, we submit to laborious, sometimes painful and even dangerous, disciplines in order to look “better”. But the improvements we seek are generally motivated by gender-related stereotypes: women seek bigger breasts, fuller lips and slimmer waists, while men yearn for better defined muscles and perhaps a bigger penis. It is as if we do not really trust nature to distinguish the sexes well enough, but must help ourselves to become what we supposedly already are.

Butler has even argued that the very binarism of biological sex is a projection of our gender expectations – the supposedly natural fact that there are exactly two sexes is itself artificially induced. To be sure, in the majority of cases it is easy enough to apply the boy/girl classification, but the interesting thing is what happens in those cases where it is not. The idea of actual intermediate possibilities and real vagueness seems intolerable when it comes to sex, and a decision is always taken, usually by the medical staff assisting at the delivery, and if need be it is reinforced by surgery.

Organization by binary opposition is a powerful feature of our common-sense worldview. Not only do we tend to view the world in terms of polar opposites and to forget the existence of intermediate or neutral cases – if you are not my friend you are my foe – but we tend to further reduce complexity by rigidly associating different oppositions with each other. According to Butler, the most evident and potentially devastating example of this tendency is the “hetero-sexual matrix” – a thought pattern where sexual preference, personal character, career choices and ways of life, are all sorted according to the male/female opposition and related to the supposedly basic distinction of biological sex.

But there are more subtle ways than external bodily design in which we are formed into what we are. The rearing period of our lives is largely devoted to drill and training in which patterns of action, reaction, perception and feeling are ingrained as bodily habits and automated reflexes. Look at language, for example. Confronted with a language we do not know, we don’t hear foreign words that we don’t know the meaning of – we are at a loss even to hear what the words are and what to make of those strange noises that flow incessantly from the lips of strangers. With our mother tongue the problem is reversed. It is impossible for us not to hear and understand what others are saying, and often impossible to hide our reactions to it – of shame, joy, pride, disgust, or whatever.

Let us look again at the simple example of a game. Understanding the rules is a very small part of learning to play football. To be any good at it you have to start out as a child, learning to kick the ball in different ways, controlling its speed, direction and height, learning to receive the ball in whatever way it comes at you, learning to run with the ball, and then learning to do all these things while in bodily contact with an opponent who
wants to take the ball away from you, learning when to make a pass and how to place yourself in order to receive one, and so on. And apart from the manual skills we have the perceptual and intellectual skills of seeing and understanding what happens on the pitch, the skills of concentration, the value system connected with winning, losing, acceptable behavior, fair play, team spirit – we could go on forever. And all of this has to be drilled until it becomes “second nature” – when the game is on there is neither time nor place for questions and deliberations about basics.

Obviously, a similar story can be told about almost any field of human activity, each has its characteristic habitus, as Pierre Bourdieu (2003) calls it. Being alert to the importance of habit formation and internalization is an important corrective to the overly intellectual attitude that it is otherwise easy to take in relation to human actions and social institutions. Looking at games from the perspective of the rule-book, at institutions from the perspective of laws and regulations and viewing actions from the perspective of deliberation and calculation of interests, we tend to forget how ingrained and inarticulate most of our patterns of thought, feeling and behavior really are.18

### 6.8 Performatives and the division of linguistic labor

We will come back to identification and bodily forming below, but first we need to say a little bit more about how social facts come about. There is an analogy between beliefs that constitute social facts and a certain type of linguistic utterances, called performatives. And not only that: it also seems that performative utterances play an important role in many social institutions – performatives are, typically, more or less ritualized speech acts that create social facts.

The term “performative” was introduced into philosophy by J. L. Austin.19 It has since been taken up and used in different and partly surprising ways by other thinkers, notably Jacques Derrida and Judith Butler, but we will start from Austin’s original idea. His basic contrast is between performatives and constatives. A constative is a normal descriptive utterance, which is made true or false by whatever the utterance is about. I utter “I was born in Liverpool” and my utterance is true if that is where I was actually born, and false otherwise. Whichever way it is, the truth or falsity of what I say is not affected by the fact that I say it. Now, look at this sentence instead (as said in a normal situation):

I promise that I will come to your party tomorrow.

It also seems to state a fact about me (namely that I promise to come) and one would expect it to be true or false depending on how the world stands with regard to that purported fact. But my promise seems to be a very peculiar fact: it doesn’t exist until I say that it does, and when I say it, it seems

---

18Bourdieu has branded this tendency to intellectualization as the “scholastic fallacy”, and sees it as a specific bias of scholars, caused by the dominance of books and thinking in their own habitus. Again (Bourdieu; 2003) is a good source.

19Austin discusses performatives in many places, but the most extended treatment is in (1962).
that I cannot be wrong. Afterwards, I can speak about my promise in the past tense, as about anything else. I can say (truly) that I did promise, or I can say (falsely) that I did not – lying about it, or perhaps having forgotten – but the original present tense utterance cannot be false. Austin’s conclusion was that saying “I promise ...” is not the description of a promise, but the making of a promise. In uttering the words that seem to describe the act, I am actually performing it.

Once the discovery was made, Austin saw performatives everywhere. Here are some of his examples:

I sentence you to 4 years in prison. (Said by a judge.)
I name you Benjamin. (Said by a priest.)
I pronounce you man and wife. (Said by a suitable official.)

It is pretty clear why we have explicit performative utterances of such kinds: many types of social status are so important in their normative ramifications that it must be unambiguous whether they apply or not. There should be no hesitation whether two persons are married or not, or if a certain punishment is applicable to someone, and so specific ritualized forms have evolved to put such things beyond doubt.

The distinction between performatives and constatives may seem clear enough, but there are in fact intermediate cases that are even more pertinent for understanding the social construction debates. Consider this question, in relation to a football game: when is it a goal? There seems to be two very different, but each in its own way correct, answers to that question:

When the entire ball has crossed the goal line.
When the referee says so.

For all practical purposes the second answer seems to be the more important. The referee decides when to award a goal, and once the decision is made it is final, and there is no appeal. Nevertheless, the first definition is clearly the more fundamental – it is on this criterion that the referee is supposed to base his or her judgement, and without it there would be no game of football. The referee may be wrong, awarding a goal when in fact there was none – but it still counts, and, as many frustrated football fans know, one may win or lose a game on account of a goal that was never scored.

The speech act of the referee has the effect of a performative, but it shares the possibility of being mistaken with ordinary constatives: I will speak of this type of act as a decisive speech act.

It is just as easy to see the importance of decision as of full-fledged performatives. In the context of a game or other social institution, a certain brute fact may “count as” a certain institutional fact, but to apply the rules of the game we have to decide in real time whether the relevant brute fact

---

20 For the sentence to take effect it is not enough, of course, that it is uttered by a judge, but there is a host of other conditions that must also be fulfilled. The same comment applies, mutatis mutandis, to all the other examples given in the text, and I rely upon the reader to supply the necessary provisos in each case.

21 I don’t know anyone who has pointed out the importance of “decisive” speech acts, in this sense, and their relation to performatives – if the reader knows better, please inform me!
is actually there. In fact, if we go back to the judge in the law court, we will see that her most important function is not the performative act of sentencing, but the decisive act of judging whether the accused is guilty or not. The accused does not become guilty by being declared as such by the court, and may in fact be innocent, but after the verdict he is being counted as guilty for the purposes of the legal system.

We may put it this way. Being guilty is both a natural fact and a social status, much like the two aspects of being gold that we touched upon earlier. The function of the judge is to decide if the social status should be awarded, based on her assessment of the natural facts. Notably, there will often be other considerations guiding her as well, for example having to do with on which side it is better to err. It is, for example, inscribed into the legal system that if there is reasonable doubt the option of not guilty should be preferred – in other cases, as we will see, the “right side to err” may be the positive one.

One noteworthy fact about most performative and decisive speech-acts is their connection to what Hilary Putnam (1975) has called “the division of linguistic labor”. Almost anyone can make a promise, to be sure, but declaring marriages, naming children, judging criminals and refereeing football games are the prerogatives of specially appointed officials. Putnam’s own examples are less formal, however, and it is interesting to compare them to the ones that we have already touched upon. Take the capacity to use the word gold:

/. . . / everyone to whom gold is important for any reason has to acquire the word “gold”; but he does not have to acquire the method of recognizing if something is or is not gold. He can rely on a special subclass of speakers. The features that are generally thought to be present in connection with the general name – necessary and sufficient conditions for membership in the extension, ways of recognizing if something is in the extension (“criteria”), etc. – are all present in the linguistic community considered as a collective body; but that collective body divides the “labor” of knowing and employing these various parts of the “meaning” of “gold” (p xx).

Clearly, Putnam is primarily thinking of another socially important function than the judge, namely the function of the expert. To be an expert, in the pure form of the concept, is simply to be outstandingly competent with regard to something. Such competence makes the expert socially useful, she becomes a resource that others can draw upon to gain access to knowledge that they do not themselves possess – including the possibility of correct classification. Another one of Putnam’s examples shows the role of the expert in almost pure form. Most of us have a lot of names of biological species in our vocabulary, that we would be hard put to apply correctly in a real context. Putnam himself admits to not being able to tell an elm from a beech tree, but that does not make the two labels synonymous for

22 Whether it is actually the judge that performs this function depends on the legal system involved – in the Anglo-Saxon traditon it is, for example, the jury.
him. He knows very well that they are different species of tree and if he needs to know more he will simply consult someone savvier.

The case of gold, however, is already less simple. Precisely because gold has such an important role for us, the function of the expert starts blending with that of the judge and we get social control and diplomas to guarantee the reliability of experts. The same type of blending occurs in many other cases as well, and a particularly rich source of examples is the practice of medicine. Many medical conditions have institutionalized social consequences. A psychiatric diagnosis may lead to enforced treatment and confinement, sometimes as an alternative to a prison sentence. A physical or psychological disability may entitle a person to specific types of support, financially or otherwise, and it may deprive one of specific rights, such as driving a car or taking responsibility for one’s own economy. The doctor, in such cases, is clearly asked to blend the functions of the expert and the judge. She is supposed to base her verdict on her scientifically grounded expertise, but a verdict it must be, on which the social consequences can be hung – an assessment of the evidence and a tentative conclusion is not enough. And just as for the judge, considerations beyond those of the brute facts will very likely come into play – having to do with the desirability of the projected consequences and with “the right side to err”.

### 6.9 Knowledge and power

From Putnam’s perspective the linguistic division of labor is by and large a good thing – we are all empowered by pooling our resources and taking advantage of our different fields of expertise. Language is a great machine that we operate together and which enables each of us to punch well above our individual weight. But, of course, there is also another side of the coin: society is not just cooperation and reciprocity, but also competition and unequal distribution of power. This more sinister side to the division of linguistic labor is a recurrent theme in the work of Michel Foucault.

According to Foucault, there is a strong tendency in the modern era to disguise social norms as natural facts, letting science replace morality and religion as the primary source of authority. In a series of historical investigations he has described how phenomena that in earlier epochs were viewed primarily in moral or religious terms – madness, crime, sexuality – during the modern era are transferred to a scientific, primarily medical, context, where notions of norms, crime and punishment, in a moral sense, are replaced by ideas about normality, deviance and corrective treatment. This obviously connects with the theme of reification: the role of the expert relying on scientific knowledge is, according to Foucault, used as an alibi for that of the judge imposing social norms.

In *The Archeology of Knowledge* (1969) Foucault introduces the notion of a “discursive formation”. We would normally think of language as primarily a system where subjects that are given before and outside of language communicate about objects that are also essentially prelinguistic. Relativistic Kantianism has prepared us to think about the objects of discourse as in some sense constituted by the language that we use. Foucault takes this one step further and argues that the subjects of discourse are no less de-
pended on linguistic schemes. Speech acts are only performed by embodied concrete agents, speaking from specific social positions defined from inside a discursive formation. Within the discourse about mental illness there are roles for doctors, nurses, patients, relatives, etc., and perhaps also for neutral outsiders. Even if they to some extent share the same linguistic forms, the preconditions and the effects of their utterances are totally different: they do not have access to the same range of possible speech acts. To take part in the discourse, you must not only know the linguistic forms and their meanings, you must also know the possible subject positions and correctly identify your own role.23

Within this framework, Foucault describes the constitution of objects as a multi-layered process, involving a complex distribution of linguistic, epistemological and practical authority (1969, chapter 3). The first stage is the “emergence” of a possible object; it is the stage where someone, to continue with the same example, becomes a candidate for the status of mentally ill. The “surfaces of emergence” are widespread. Foucault mentions the family, religious communities, the school system, work groups, military service – groupings “which all have a margin of tolerance and a threshold beyond which exclusion is demanded”. But although the question of mental illness may be raised in these contexts, one does not have the mandate to settle them there. Questions of diagnosis, explanation and treatment are delegated to the “authorities of delimitation”: doctors, hospitals, expert panels and even law-courts. At this stage, what is originally just deviant or mad, in a vague sense, gets submitted to detailed theoretical elaboration, which serves as the basis for the practical application of different, more or less coercive, treatments, but which also has the function to underpin and legitimize the authority of those who measure them out.

Foucault’s point is that the very notion of a mental illness only makes sense within this network of knowledge and power. A family of behaviors that would not have been classified together in other epochs are taken to have something in common, which motivates for the people involved to be delivered to the theoretical and practical authority of a specific system of social institutions, ranging from the clearly medical to the penal system, where they are subject to various types of reformation and exclusion from ordinary society.

I take it as obvious that Foucault’s description of discursive formations is illuminating and interesting in many ways, and that there is an important insight in the general thesis about the tendency to conflate the roles of the judge and the expert, using the authority of science to naturalize social norms. But does it follow from this that mental illnesses, like schizophrenia, do not really exist or that there is something wrong with current psychiatric practices? Let us look at a couple of other examples.

First, the medieval practice of the Inquisition. It seems to fit Foucault’s model perfectly. The express purpose of the Inquisition was to detect, diagnose and treat cases of heresy. As in the case of madness there was a widely distributed vague awareness of heresy which allowed for the emergence of possible cases, which could be brought to the attention of the re-

23 Patients, sometimes, will be likely to try to transgress their own role, and perhaps not to understand the rules of the game, but this will only serve to defeat their contribution, transforming it from speech act to symptom.
evant authorities of delimitation, i.e., the courts of the Inquisition. The courts then could apply specific procedures to decide whether a purported case of heresy really was one, using a highly developed taxonomy for the theoretical elaboration of different kinds and degrees of heresy, defined in terms of the conceptual resources of contemporary Christian doctrine. And, as a final stage, punishments and treatments were measured out, ranging from a rich variety of capital punishments to milder forms of penitence.

As a second example, take the classical culture surrounding tuberculosis, as it is described for example in Thomas Mann’s novel Der Zauberberg. Again, we have all the ingredients of Foucault’s model: the surfaces of emergence, the authorities of delimitation, the theoretical elaboration with the different forms and stages of the disease, the institutional framework for exclusion and treatment, the different subject positions, ready for identification and habituation.

From a certain perspective, the examples seem exactly parallel. There is just one difference. Tuberculosis is a real disease, and the culture that was built around it is reasonably seen as a way to cope with a dangerous infection against which one had no efficient treatment, before the advent of antibiotics. Herey, on the other hand, is social all the way down: the Inquisition was nothing but an instrument of power and domination, on many levels, from the small scale uses of accusations against a neighbour to exact revenge, to the large scale politics of church power and the establishment of central government and the national state.

Now, is schizophrenia more like tuberculosis or more like heresy? I will let the reader decide, but for now the important thing is that no amount of discourse analysis will settle that dispute. To find out, one must answer all sorts of questions about the etiology and treatment of a range of mental and physiological states and human behavior – as always by elaborating theories, collecting evidence, weeding out sources of bias and error – but very few of these questions belong to philosophy, the history of ideas or the sociology of science.

To point this out takes nothing away from the importance of Foucault’s analysis. Claims to knowledge and scientific status are instruments of power, both within science and in relation to society at large, and it is important to understand that social function. In pursuing such an understanding it is often useful to distance oneself from questions about the validity of the relevant knowledge claims and the associated norms and practices, in order just to describe how the system works. But to do that is not in itself to deny those claims or supply reasons against them – the sociology of knowledge is not a master science where every question is finally resolved.\footnote{In saying this, I do not mean to imply that Foucault himself commits this particular mistake – in fact, I think that he is often remarkably good at staying neutral, and that what provokes his critics is often not the valuations that he makes, but the valuations that he does not make.}

### 6.10 Knowledge as a social status

The important function of performatives and decisives is to secure general recognition of something as a fact, in order to get on with the practices
that hang on it. We need to know “in real time” who is guilty and who is innocent, whether a vote has been cast, whether a goal has been scored – and so we put specific procedures in place in order to resolve all doubt.\footnote{In his discussion of Austin, Derrida (1972) points out that, strictly speaking, all doubt can never be removed in this way, because we can always ask for further procedures to certify that the conditions of the first ones are actually fulfilled.}

But the need for socially accepted agreement about how things stand is not possible to delimit in advance – there are countless circumstances in which we need to presuppose that something or other really is the case, in order to base further action or investigation on that presupposition.

Now, what is the general name for opinions that we have the right to take for granted without further discussion, in a certain social context? Is that not what we call “knowledge”? So what shall we say – is knowledge itself a social construction, a social status bestowed by society on certain opinions? This looks rather similar to the sociological concept of knowledge that we gave such short shrift in Chapter 1. Should we reconsider our view? Is the sociological conception of knowledge really basic after all?

What would it mean to say that knowledge is a social status? There is a wide array of different cases. Compare the case of money, where some things are arbitrarily assigned a social status in order to serve a certain social purpose, with the case of being the Olympic champion of the 100 m sprint. To be an Olympic champion is surely a social status – it exists only in the context of the Olympic games – but there is nothing arbitrary about it: it is bestowed upon you as a recognition that, on a certain occasion, you ran faster than all the competition. Now, if we admit that being knowledge is in some sense a social status – is it more like being money or more like being an Olympic champion? At least \textit{prima facie}, it surely seems a lot more like being an Olympic champion. The status of knowledge is bestowed on some opinions as a recognition that they fulfill certain conditions that are not themselves social statuses, of which the most important is the condition of being true, and if it should turn out that they do not fulfill the conditions, the status is withdrawn.

We may admit that there is a social status of being counted as knowledge, without admitting that knowledge is just a social status, or that it is primarily a social status. To think that it is, is a mistake on a par with thinking that to be a goal in football is just to be declared to be a goal by the referee. We delegate the decision whether a goal has been scored to the referee, and we agree to abide by that decision, but the primary criterion of being a goal is the one that we expect the referee to apply, namely that the whole ball shall have passed the goal-line in the course of play. In the same vein, we allow ourselves and each other to take a certain stock of received opinions for granted as “knowledge”, but only because we expect them to have been already cleared for the primary criterion of knowledge: that there shall be sufficient reason for thinking that they are true.

\section*{6.11 Two fallacies}

The time has come to draw some threads together and tie things up a little bit. We took off, in this chapter, from the comparatively simple notion of a
social fact, using examples like games and money, that wear their sociality on their sleeves, so to speak. We then introduced the ideas of reification, identification and habituation, allowing us to understand the ambiance of falsity and deceit that surround the more controversial (alleged) cases of social construction, like sexual identities, and we may half-seriously venture a definition along these lines:

\[
\text{Social construction} = \text{Social fact} + \text{reification} + \text{identification} + \text{habitation}
\]

To say that being a woman, for example, is a social construction, in this sense, is to say that being a woman is a social role masquerading as a natural fact, imposing itself on its bearer through identification, to the point that it becomes “second nature”.

Why is it important whether something is a social construction or not? Obviously, it has to do with the space of possibilities in which something is embedded, with the dynamics of possible change. We think of natural facts as given, as just being there to be reckoned with, and not to be evaluated in moral or political terms. Nature is not unchangeable, of course – for good and for worse there is not much in our environment that has not been touched or modified by human activity – but to say that a state of affairs is “natural” nevertheless works as a sort of defense of it, a strong hint that it should be accepted as it is. To say that something is a social construction, on the other hand, is to say that it depends on how we think, feel and act in relation to each other – that it is, at least in principle, within the realm of our freedom and responsibility. Perhaps it is even a sort of duty to reflect and take a stand on it?

To map this space of what is humanly possible is one of the main functions of research in the humanities, most obviously in history and anthropology. The notion of a social construction is a tool for that work, but it is also important to see that it is a rather blunt tool, and that to classify something as a social construction is not in itself to say very much about it. In particular, there are two tempting but dangerous fallacies to be on one’s guard against.

First, we have the “nothing but” fallacy. Important as it may often be to point out the conventional or normative component of something, and to challenge the air of natural necessity in which it may be cloaked, the tempting step from ‘X is a social construction’ to ‘X is nothing but a social construction’ is in most cases misleading and sometimes dangerous. Almost anything touched by the human mind is an intricate weave of nature and culture, held together by links of causality and motivation, different from case to case. To pin a label like “social construction” on something should not be the end of analysis and investigation, but the beginning.

Second, we have the closely related “just like” fallacy. Behind the concepts and theories of this (as of any other) area is a network of analogies and metaphors. Behind a glossy distinction like that between constitutive and regulative rules are a few examples of each kind, and the optimistic projection that other examples shall fall neatly into one of the two similarity classes suggested. Social institutions are like games, natural facts are signifiers of cultural signifieds – sure, but what kinds of games and what
kinds of signifiers in each case? Anything is in some respects like anything else, but not just like it. Analogies illuminate but they do it both ways: differences are as important as similarities.\textsuperscript{26}

\textsuperscript{26}This may be a fitting place to refer to Ian Hacking's *The Social Construction of What* (1999), where he, with the help of carefully elaborated examples, forcefully argues the point that there is no general theory of social construction, but that each case must be treated as new.
Bibliography


155


