Ingvar Johansson, Niels Lynøe

Medicine & Philosophy

A Twenty-First Century Introduction

Ingvar Johansson, Niels Lynøe

Medicine & Philosophy

A Twenty-First Century Introduction

ontos verlag

Frankfurt I Paris I Lancaster I New Brunswick

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

North and South America by
Transaction Books
Rutgers University
Piscataway, NJ 08854-8042
trans@transactionpub.com

United Kingdom, Ire, Iceland, Turkey, Malta, Portugal by
Gazelle Books Services Limited
White Cross Mills
Hightown
LANCASTER, LA1 4XS
sales@gazellebooks.co.uk

Livraison pour la France et la Belgique:
Librairie Philosophique J.Vrin
6, place de la Sorbonne ; F-75005 PARIS
Tel. +33 (0)1 43 54 03 47 ; Fax +33 (0)1 43 54 48 18
www.vrin.fr

P.O. Box 15 41, D-63133 Heusenstamm
www.ontosverlag.com

ISBN 978-3-938793-90-9

2008

Printed on acid-free paper
ISO-Norm 970-6
FSC-certified (Forest Stewardship Council)

Printed in Germany
by buch bücher **dd ag**

Table of Contents

(The front cover contains a picture of Helicobacter pylori; see chapter 7.7.)

Foreword

This book is meant to acquaint the reader with problems that are common to medical science, medical ethics, medical informatics, and philosophy. Our conviction is that all the disciplines mentioned can benefit from some more interaction (see Chapter 1). In this vein, we offer our book to the readers. We hope that it can be of interest not only to people working within the medical field proper, but to healthcare workers in a broad sense, too. Similarly, we hope that it can be of interest not only to medical information scientists, but to all bioinformaticians (especially Chapter 11).

The book can be divided into three parts. The first part consists of Chapters 1-7, and treat questions concerned with 'philosophy of science and medicine'; Chapters 8-10 are concerned with ethical matters of various kinds, and might be called 'ethics and medicine'; the third part (written by Ingvar J alone) consists only of Chapter 11, and it might be called 'philosophy of classification and medicine'.

To many people learning about medical facts might be like building a 'tower of facts', i.e., using facts as if they were bricks that can be put both side by side and on top of each other. Learning philosophy is not like this. It is more like observing a house in a mist when the mist gradually fades away. Initially, one sees only parts of the house; and these parts only vaguely. But gradually one sees more and more distinctly. We ask readers who are not already familiar with the philosophy of medicine to keep this in mind, especially if some parts of the present book are not immediately understood. We have done our best in order to be pedagogic, but the holistic feature of some learning processes is impossible to bypass. After each chapter we present a reference list, suggesting books for further reading as well as listing some of the references on which our views are partly or wholly based. The historical name-dropping is done in the belief that it makes it easier for the readers to move on to other books in philosophy, medicine, and the history of science.

In order to avoid uncomfortable terms such as 'he/she' and 's/he', and not writing 'she' since we are 'he', we use the personal pronoun 'he' as an abbreviation for physicians, patients, researchers, dabblers, and quacks of all sexes. We regret the lack of an appropriate grammatically neutral

personal pronoun, and we think that to substitute an all-embracing ‘he’ with an all-embracing ‘she’ is not a good move; at least not in the long run.

Our backgrounds are as follows. Ingvar Johansson is professor of philosophy (Umeå University, Sweden) with long experience of teaching and doing research in philosophy in general as well as in the philosophy of science. From July 2002 to March 2008, he has mainly been working as a researcher at the Institute for Formal Ontology and Medical Information Science (Saarland University, Germany). Niels Lynøe is a general practitioner and professor of medical ethics (Karolinska Institutet, Stockholm, Sweden); he has for many years taught the philosophy of medicine for medical students and PhD students in Umeå and Stockholm.

This book of ours has a Scandinavian twentieth century pre-history. In 1992, we published a Swedish book with the same intent as the present one. This book was revised and enlarged in a second edition that appeared in Swedish in 1997 and in Danish in 1999. The present book, however, contains so many re-writings, revisions, and additions, including whole new sections and even a new chapter (the last), that it has become a wholly new book. Among the new things is an even harder stress upon the fallibilism of science. Of course, we regard our own knowledge as fallible, too.

We would like to thank Frédéric Tremblay for very many good comments on the whole book. Several persons have taken the time to read and give useful comments on some parts or specific chapters of the book. They are: Gunnar Andersson, Per Bauhn, Dan Egonsson, Uno Fors, Boris Hennig, Søren Holm, Pierre Grenon, Boris Hennig, Rurik Löfmark, Stefan Schulz, Barry Smith, Andrew Spear, and Inge-Bert Täljedal. For these comments we are also very grateful.

Saarbrücken and Stockholm, February 2008,
Ingvar Johansson and *Niels Lynøe*

Acknowledgement:
I. Johansson’s work was done under the auspices of the Wolfgang Paul Program of the Alexander von Humboldt Foundation, the Network of Excellence in Semantic Interoperability and Data Mining in Biomedicine of the European Union, and the project Forms of Life sponsored by the Volkswagen Foundation.

1. Science, Morals, and Philosophy

Many scientists – perhaps most – regard scientific research as a process that is wholly independent of philosophical problems and presuppositions. Conversely, many philosophers – perhaps most – take philosophy to be an enterprise that is independent of the results of the sciences. In the history of philosophy, some philosophers have placed philosophy *above* science, claiming not only that all philosophical problems can be solved independently of the sciences, but also that empirical science has to stay within a framework discovered by philosophy alone. This is true of Kant and of pure rationalists such as Descartes and Hegel. Other philosophers, especially logical positivists, have placed philosophy *below* science, claiming that, in relation to the sciences, philosophy can only contribute by sharpening the conceptual tools that scientists are using when they try to capture the structure of the world. In both the cases, philosophy is looked upon as being of some relevance for the sciences, but there are also philosophers who claim that philosophy is of no such relevance whatsoever. For instance, we think that the self-proclaimed epistemological anarchist Paul Feyerabend would be happy to agree to what the famous physicist Richard Feynman is reported to have said: 'Philosophy of science is about as useful to scientists as ornithology is to birds'. All these three views make philosophy sovereign over its own domain, and all of them except the Descartes-Kant-Hegel view make science sovereign over its domain too. Our view is different. We claim that science and philosophy are overlapping disciplines that can benefit from interaction. When science and philosophy are not in reflective equilibrium, then one of them, if not both, has to be changed, but there is no meta-rule that tells us what ought to be changed.

Physics, which is often regarded as the pre-eminent empirical science, has from a historical point of view emerged from philosophy. Many scientists and philosophers have falsely taken this fact to mean that embryonic sciences may need to be nourished by the theoretically reflective attitude typical of philosophy, but that, when a science has matured, this umbilical cord should be cut. On our view it should not. But

then nourishment ought to flow in both directions. We are not claiming that science and philosophy are identical, only that there is an overlapping division of labor. Often, the overlap is of no consequence for specific scientific research projects, but sometimes it is. And this possibility is of such a character that all scientists and science based practitioners had better acquaint themselves with some philosophy. One purpose of the present volume is to show that medical science has a philosophical ingredient; another is to show that even medical problems of practical and ethical natures can benefit from being philosophically highlighted.

Prior to the nineteenth century, researchers within physics were classified as philosophers as much as scientists. Why? Probably because the growing specialization had not yet turned the natural sciences and philosophy into administratively separate disciplines. The first European universities of the eleventh and the twelfth century had usually four faculties, one for philosophy, one for medicine, one for law, and one for theology. The different natural sciences were at this time regarded as being merely different branches of philosophy. Newton's chair in Cambridge was a chair in 'Natural Philosophy'. But great scientists such as Galileo and Newton even thought and wrote about purely philosophical questions. Conversely, great philosophers such as Descartes and Leibniz made lasting contributions in the fields of physics and mathematics. During the time when the Arabic culture was the most advanced scientific-philosophical culture in the world, prominent persons such as Ibn Sina (Avicenna) and Ibn Rushd (Averroes) made contributions to both philosophy and medicine.

The difficulty in keeping science and philosophy completely apart sometimes shows itself in class-room situations. Now and then students put forward questions that the teacher evades by saying that they are 'too philosophical'. Of course, this *may* be an adequate answer; there is a division of labor between science and philosophy. However, many teachers seem always to use the phrase 'too philosophical' in the derogatory sense it has among people who think that philosophical reflections can never be of scientific or practical relevance. A change of attitude is here needed.

An important aspect of our interactive message is that neither science nor philosophy can be the utmost arbiter for the other. On the one hand, the philosophy of science should not be given a juridical function, only a

consultative one. Philosophers of science should not be appointed legislators, judges, or policemen with respect to scientific methodologies. On the other hand, scientists should not tell intervening philosophers to shut up only because philosophers are not scientists. As the situation looks within medicine today, we think such an interaction is especially important in relation to the interpretation of abduction and some medical probability statements (Chapters 4.6 and 4.7), the analysis of the placebo effect and the discussion of psychosomatic phenomena (Chapter 7), the fusion of ethics and medical research (Chapter 10), and the handling of the medical information explosion (Chapter 11).

In some respects, a philosopher of science can be compared to a grammarian. As a grammarian knows much about language structures that most speakers do not bother about, a philosopher of science can know much about structures of scientific research and scientific explanations that scientists have not bothered to think about. But neither language structures nor science structures are eternal and immutable. Therefore, in neither case should the studies in question aim at reifying and conserving old structures. Rather, they should be means for improving communication and research, respectively.

The importance of science for society has increased dramatically since the mid-nineteenth century. This is since long publicly obvious in relation to the natural and the medical sciences, but it is true even for many of the social sciences. Nowadays, many economic, political, and administrative decisions, both in public and private affairs, are based on reports and proposals from expert social scientists. From a broad historical perspective, after a millennium long and slow development, science has now in many countries definitely replaced religion as the generally accepted and publicly acknowledged knowledge authority. During the period that for Europe is called the Middle Ages, people turned to their clergymen when they wanted authoritative answers to their questions about nature, man, and society; sometimes even when they wanted answers about diseases, since some of these were regarded as God’s punishment. Today we ask scientists. And most people that are in need of medical treatment and advice go to scientifically educated physicians. The contemporary authority of university educated physicians is partly due to the general authority of science in modern societies.

There is a link between the first phases of modern science and the broader intellectual movement called ‘the Enlightenment’, and there is also a link between the Enlightenment and the attempt to establish free speech, public political discussions, and democracy as being basic to good societies. The combination of science, rationality, democracy, and industrialization is often referred to as ‘modernity’. Modernity contains a hitherto unseen stress on discussion and argumentation. Arguments are intended to convince, not just to persuade. Even though there is a gray zone between true convincing and mere persuading, the prototypical cases are clearly distinct. Threats, lies, and military parades have only to do with persuading, but rational arguments are necessary when one tries to convince a person. In most arguments there is an assumed link to some facts, i.e., in order to argue one should know something about reality.

When enlightenment became an intellectual movement in the eighteenth century, it usually stressed secure and infallible knowledge. As we will make clear, such an epistemological position can no longer be sustained. But this does not mean that rationality and argumentation have to give in; neither to faith, as in the New Age Wave criticism of the sciences, nor to the view that it is impossible to acquire any interesting knowledge (epistemological nihilism), as in much social constructivism and post-modern philosophy. Something, though, has to be changed. Both epistemological fallibilism (Chapter 3.5) and an acceptance of tacit knowledge (Chapter 5) have to be injected into the Enlightenment position in order to enlighten this position even more. If the rationality ideal is made too rigid, neither science nor science-based technologies and practices will be able to continue to develop. Put briefly, although we can understand and agree with some of the criticisms of modernity, we definitely concur with the latter’s basic ideas. Also, we think such an adherence is important if medical science and medical technology shall be able to continue its remarkable development. Even though large parts of the world might truly be said to live in post-industrial and post-modern societies, there is no need to become a post-modern philosopher, only a need to improve our view of science and philosophy. Our view admits the globalization of science.

Fallibilism implies tolerance. Everyone needs to hear criticism of his views in order to keep them vivid, and such an insight might ground some

tolerance. But as soon as one acknowledges the possibility that one may be wrong – partly or wholly – one has to become much more tolerant. Why? Because then criticism might be needed even in order for oneself to be able to improve one's views. Tolerance is necessary not only in religious and political matters, but also in scientific and philosophical.

In Chapters 2-7 we are mainly presenting traditional problems in the philosophy of science and what we think the solutions look like; in Chapter 8, tolerance in medical science is discussed in more detail; and in Chapters 9 and 10 we take into account the fact that ethical problems have become an integral part of modern clinics and medical research. We discuss both the morality of being a good researcher and ethical guidelines related to informed consent in research. In Chapter 11, the last chapter of the book, we discuss taxonomic work and take account of the fact that medical informatics and bioinformatics have become part of medicine. Traditional twentieth century philosophy of science gave taxonomy a step-motherly treatment. It was preoccupied with questions concerned with empirical justifications and theoretical explanations. To classify and to create taxonomies were falsely regarded as proto-scientific activities. Taxonomies have always been fundamental in science, but it seems as if the information explosion and the computer revolution were needed in order to make this fact clearly visible. And not even taxonomic work can be completely released from philosophical considerations.

Philosophy,
medical science,
medical informatics, and
medical ethics
are overlapping disciplines.

2. How and Why Does Science Develop?

There are, and have been, many myths about science and scientists. In particular, there are two versions of the myth of the lonely genius. One version stems from romanticism. It regards the brilliant scientist as a man who, in a moment of inspiration, unconditioned by his social setting, creates a new idea that once and for all solves a scientific problem. The other version disregards the surrounding milieu, but stresses a supposedly calm and disinterested use of a rational faculty. Notwithstanding the existence of scientific geniuses, these views heavily underrate the role played by technological, economic, political, social, and cultural circumstances in the development of science. Even though some famous scientists have in fact had the experience of receiving a revolutionary idea like a flash of lightning, it should be remembered that even real light flashes have their very determinate existential preconditions. We would like to propose an analogy between ‘swimming’ and ‘doing research’.

There are bad swimmers, good swimmers, and extremely good ones. But in order to swim all of them need water in some form, be it a pool, a lake, or a sea. Analogically, there are researchers of various capabilities, but all of them need an intellectual milieu of some form, be it a university, a scientific society, or an informal discussion forum. In order to learn to swim one has to jump into the water sooner or later, and in order to learn how to do research, one has to enter an intellectual milieu sooner or later. Furthermore, as it is easier to swim in calm water than in troubled, innovative research is easier in tolerant milieus than dogmatic. Let us end this analogy by saying that some research is more like playing water polo than merely swimming a certain distance.

Louis Pasteur is often quoted as having said: ‘In the field of observation, chance favors only the prepared mind.’ It is a variation of the more general theme that luck is just the reward of the prepared mind. Normally, in order to invent or discover something new, people must be in a state of readiness. Therefore, even seemingly accidental scientific discoveries and hypotheses can be fully understood only when seen in the light of their historical settings**.**

We will distinguish between the question (i) *how* science develops and the question (ii) *why* it develops, i.e., what causes it to develop.

(i) Does science always accumulate by adding one bit of knowledge to another, or are there sometimes discontinuities and great leaps in which the old house of knowledge has to be torn down in order to give room for new insights? The history of science seems to show that in one respect scientific communities (with the theories and kind of research they are bearers of) behave very much like political communities (with the ideologies and kind of economic-political structures they are bearers of). Mostly, there is an evolutionary process, sometimes rapid and sometimes slow, but now and then there are revolutions. In some cases, historians talk of half a century long extended revolutions such as the first industrial revolution around the turn of the eighteenth century and the scientific revolution in the mid of the seventeenth century. In other cases, such as the French revolution of 1789 and the Russian one of 1917, the revolutions are extremely rapid. In science, most revolutions are of the slow kind; one case of a rapid revolution is Einstein's relativistic revolution in physics.

(ii) When those who search for causes behind the scientific development think they can find some overarching one-factor theory, they quarrel with each other whether the causes are factors such as technological, economic, political, social, and cultural conditions external to the scientific community (externalism) or whether the causes are factors such as the social milieu and the ideas and/or methodologies within a scientific community (internalism). We think there is an interaction but, of course, that in each single case one can discuss and try to judge what factor was the dominant one.

The 'How?' and the 'Why?' questions focus on different aspects. This means that those who think that (i) *either* all significant developments come by evolution *or* all significant developments come by revolutions, and that (ii) *either* externalism *or* internalism is true, have to place themselves in one of the four slots below:

	Pure *evolutionary* view	Pure *revolutionary* view
Pure *internalist* view	1	2
Pure *externalist* view	3	4

We want everybody to think in more complex terms, but we will nonetheless for pedagogical reasons focus attention on merely one or two slots at a time. But first some more words about creative scientists.

2.1 Structure and agency

The discussion between externalists and internalists is a discussion about what kinds of causes, correlations, or structures that have been most important in the development of science. Externalists and internalists oppose the romantic and the rationalist views of the scientist, but even more, both oppose or avoid in their explanations talk of freely creating scientists. This denial should be seen in light of the millennia long debate about determinism and free will in philosophy and the corresponding discussion in the philosophy of the social sciences, which has been phrased in terms of structure and agency. In our little comment we take the so-called ‘incompatibilist view’ for granted, i.e., we think that it is logically impossible that one and the same singular action can be both free and completely determined.

In our everyday lives, it seems impossible to stop altogether to ask, with respect to the future, questions such as ‘What shall I do?’ and, with respect to the past, questions such as ‘Why did I do that?’ Also, it is hard to refrain completely from asking questions that bring in moral and/or juridical dimensions of responsibility and punishment, i.e., questions such as ‘Who is to blame?’ and ‘Who is guilty?’ Normally, we take it for granted that, within some limits, we are as persons acting in ways that are not completely pre-determined by genes, upbringing, and our present situation. Implicitly, we think we have at least a bit of freedom; philosophers sometimes call this view *soft determinism*. Science, however, even the science of the history of science, looks for the opposite. It looks for causes, correlations, and structures; not for freedom and agency. When it looks backwards, it tries to find explanations why something occurred or what

made the events in question possible. When it looks forwards, it tries to make predictions, but freedom and agency represent the unpredictable.

Disciplines that study the history of science can philosophically either admit or try to deny the existence of agency within scientific research. The denial of agency is very explicit in the so-called ‘strong program’ in the sociology of scientific knowledge (e.g., David Bloor and Barry Barnes), but even historians of science that admit human agency have to focus on the non-agency aspect of science.

Accepting the existence of agency, as we do, social structures have to be regarded as being at one and the same time *both constraining and enabling* in relation to actions. A table in front of you put constrains on how you can move forward, but at the same time it enables you easily to store some things without bending down; the currency of your country or region makes it impossible for you to buy directly with other currencies, but it enables you to buy easily with this very currency. Similarly, social structures normally constrain some scientific developments but enable others. The philosopher Immanuel Kant (1724-1804) has in a beautiful sentence (in the preface to *A Critique of Pure Reason*) captured the essence of constraining-enabling dependencies: “The light dove, cleaving the air in her free flight, and feeling its resistance, might imagine that its flight would be still easier in empty space.” Air resistance, however, is what makes its flight possible. Similarly, brilliant scientists may falsely imagine that their research flights would be easier in a room emptied from social structures and critical colleagues.

It is as impossible in scientific research as in everyday life to stop asking agency questions such as ‘What shall I do?’ Since experiments and investigations have to be planned, researchers have to ask themselves how they ought to proceed. If an experiment does not give the expected result, the experimenters have to ask ‘Did we make anything wrong?’ Agency comes in even in relation to the simple question ‘Is there anything more that I ought to read just now?’ Moral questions always bring in agency. Therefore, agency pops up as soon as a research project has to be ethically judged (see Chapters 9 and 10). Even if the acting scientist is invisible in his research results, his agency is just as much part of his research life as it is part of his everyday life.

2.2 Externalism and internalism

According to the pure externalist view, scientific developments are the results *only* of technological, economical, political, social, and cultural factors external to the scientific community. That such factors play some role is trivially true and easily seen in modern societies. Mostly, the state allocates resources for research; each and every year the government presents a research policy bill to the parliament. Also, many big technological businesses and pharmaceutical companies house complete research departments that can be given quite specific research directives.

The external factor can also be illustrated historically. The ancient river valley civilizations of Mesopotamia and Egypt acquired much knowledge of the movements of the stars. Why? They were agrarian societies based upon well-organized irrigation systems, and they needed an almanac by means of which they could predict the floods. But, in turn, a precondition for a functioning almanac was some knowledge about the positions of the heavenly bodies, i.e., some astronomical knowledge. However, the astronomical knowledge they acquired went far beyond what was necessary for estimating the phases of the year. In these cultures, by the way, there was no distinction made between astronomy and astrology. The constellations of the stars and the planets were also regarded as being of importance for the interpretation and prediction of societal events and relations between humans.

These societies did not have any scientists in the modern meaning. It was clergymen who, from our point of view, were at the same time astronomers. Religion, science, and technology were, we might retrospectively say, tacitly seen as an integrated whole. Therefore, even though the ancient agrarian societies of Mesopotamia and Egypt were not centrally interested in obtaining knowledge based on empirical evidence, they did nonetheless produce such knowledge. It was left to the Ancient Greek natural philosophers (e.g., Thales of Miletos, ca. 624-546 BC) to be the first to adopt an exclusively theoretical and scientific attitude towards knowledge of nature. Yet, astronomy was still intimately related to practical needs until much later. When Ptolemy (ca. 90-168) constructed his theory about how the planets and the stars move around the earth, such knowledge was of importance for sailing. At nights, sailors navigated by means of the positions of the heavenly bodies.

Andreas Vesalius (1514-1564) was the foremost in the first generation of physicians after Galen (129-200) that tried to study the human body and its anatomy in detail; he was also one of the first to present his findings in detailed figures. He got his new knowledge partly from dissections of corpses of executed people. But his scientific curiosity was not the only factor. In the early Italian Renaissance, such dissections became allowed. In order to draw and paint the human body in realistic detail, even artists such as Leonardo da Vinci (1452-1519) and Michelangelo (1475-1564) studied the anatomy of the human body also by means of corpses (Figure 1). Galen had been a physician for gladiators, and he had made public dissections on living animals, but he had not really dissected human bodies.

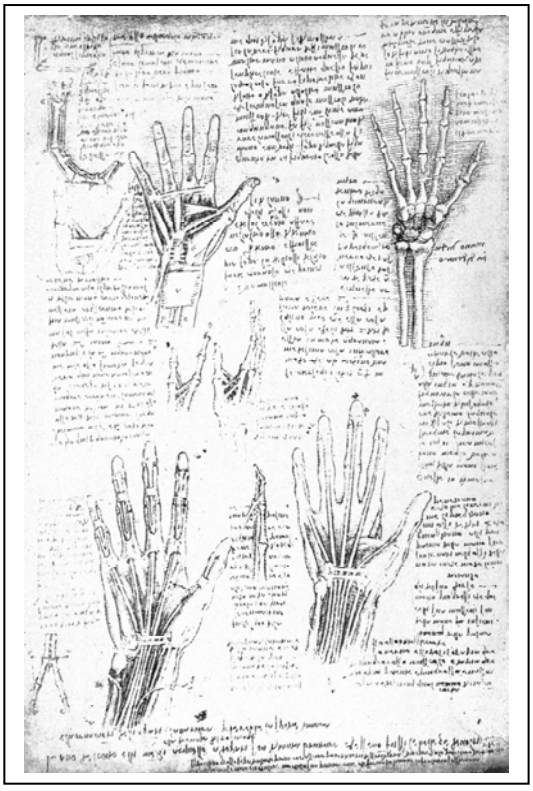


Figure 1: *Anatomical structures drawn by Leonardo da Vinci*

The interaction between external factors and science is easily seen in relation to technology. Just as the emergence of new scientific theories may be heavily dependent on a new technology, new technologies can be equally strongly dependent on new scientific theories. Without the

nineteenth century theories of electromagnetism, the inventions of the electric engine and the electric generator are hardly thinkable, and without twentieth century quantum mechanics, the inventions of the laser and some of the modern computer hardwares are hardly thinkable. In these cases, the direction goes from science to technology. But without the eighteenth century invention of the steam engine, the science of thermodynamics would have been next to impossible to discover. Here, the theory that made it possible to explain the function of the engine came after the invention of the engine. Another conspicuous example where a new invention creates a new science is microbiology. Without the invention of good microscopes, microbiology would have been impossible. It should, though, be noted that the microscope was only a necessary condition. It took more than 200 years before it was realized that the micro-organisms seen in the microscopes could be causes of diseases (see Chapter 2.5). The invention of the microscope, in turn, was dependent on a prior development of techniques for cutting glasses.

Back to social structure. During the seventeenth and the eighteenth century, the European universities focused mainly on teaching well-established knowledge. Research, as we understand it today, was not part of a university professor's obligations. This is the main reason why scientific societies and academies, like the famous Royal Society in London, arose outside universities. Eventually, the success of these scientific societies forced universities to change their internal regulations and to integrate research within the teaching task.

According to the pure internalist view, scientific development can be understood without bringing in factors from outside of the scientific community. If agency is admitted into internalism, one can note that some researchers consciously try to *make* internalism true. They try to forbid external views to influence research. As a tumor biologist once said: 'I am not interested in developing new treatments – I am devoted only to understanding what cancer is.' The American sociologist of science Robert K. Merton (1910-2003) imagined a group of scientists proposing in this vein a toast: 'To pure mathematics, and may it never be of any use to anybody.'

Even advocates of pure internalism are of course aware of the fact that, in some way, be it by taxpayers or by firms, full-time working scientists

have to be supported financially. With respect to Ancient Greece, this point can bluntly be made by saying that it was slavery that made it possible for interested free men to conduct research. Some outstanding researchers have managed to be both economic entrepreneurs and scientists. According to the pure internalist view, such economic preconditions are wholly external to the way the *content* of science progresses.

In relation to internalism, the Copernican revolution (in which it was discovered that the sun, not the earth, is at the center of the planetary system) has an interesting feature. Copernicus' heliocentric theory had tremendous repercussions on the general worldview and on the various Churches' interpretation of the Bible. However, and astonishingly from an externalist perspective, it had no immediate consequences for navigation techniques (which still relied on the positions of the stars). Despite the fact that Ptolemy's astronomy is based on the false assumption that the sun and all planets are moving around the earth, his theory was at the time sufficient for the practical purposes of navigation. Thus science can progress of itself beyond the contemporary practical needs of society.

The fact that there can be interaction between factors that are external and internal to the scientific community can easily be seen in relation to disease classifications too. Since diseases, and what causes them, figure in many insurance contracts, it can be of great economic importance for many people (a) whether or not their illnesses are symptoms of a disease and (b) whether or not their diseases can be tracked to some specific causes such as accidents or workload conditions. Let us exemplify.

In the 1980s, the American Psychiatric Association declared that there is a certain mental illness called 'post traumatic stress disorder' (PTSD). It was discovered in the aftermath of the Vietnam War, and the diagnosis includes a variety of symptoms such as anxiety, depression, and drug or alcohol dependence. According to anthropologist Allan Young, the symptoms should be considered to result, not from an actual traumatic event, but from the recovered memory of an event. Mostly, the mental breakdown began only when the soldiers came back to the US. It is the delay in reaction to the trauma that sets PTSD apart from the so-called 'shell shock' suffered by many soldiers in the First World War. This classification was wanted not only by the psychiatrists, but also by many veterans and their supporters. Probably, the latter wanted this

psychiatrizing of their symptoms not for scientific reasons, but because it meant free treatment and economic compensation.

In retrospect, it is easy to find examples where medical classifications seem to be almost wholly socially conditioned. The once presumed psychiatric diagnoses ‘drapetomania’ and ‘dysaesthesia Aethiopica’ are two cases in point. These classifications were created by the Louisiana surgeon and psychologist Dr. Samuel A. Cartwright in the 1850s. The first term combines the Greek words for runaway (‘drapetes’) and insanity (‘mania’), and it was applied to slaves that ran away from their owners. In ordinary language, the classification says that such slaves suffer from a psychiatric disease, an uncontrollable desire to run away. Dysaesthesia Aethiopica means ‘dullness of mind and obtuse sensibility of body that is typical of African negroes’. It was thought to be a mental defect that caused slaves to break, waste or destroy (their master’s) property, tools, animals, etc. Both ‘diseases’ occurred in the American South, and the diagnoses eventually disappeared when slavery ceased. Similarly, the eugenically based sterilizations conducted in many European countries and in the US in 1920-1960 were more influenced by social ideologies than scientific reasoning. In Nazi oriented medicine, being a Jew was perceived as a genetic disease or degeneration. Homosexuality was in many (and is still in some) countries seen as a psychiatric dysfunction. It is obvious that this kind of disease labeling cannot be understood if one does not take the historical and social context into account. What is mainly socially conditioned in contemporary science is for the future to discover.

The scientific discipline ‘sociology of knowledge’ emerged in the 1920s with works by Max Scheler (1874-1928) and Karl Mannheim (1893-1947), and within it pure externalism has been a rare phenomenon. It was, though, explicitly advocated by some Marxist inspired historians of science in the 1930s (e.g., Boris Hessen (1893-1936) and J.D. Bernal (1901-1971)) and some in the 1970s; Bernal later in his life stressed interaction, and Hessen had no chance to change his mind since he was executed by Stalin. Pure internalism has mostly seen the light in the form of autobiographical reflections from famous scientists, and it has perhaps become a completely outmoded position. But it has been argued even among contemporary philosophers of science (e.g., Imre Lakatos, 1922-1974) that it would be a

good thing to write the history and development of science *as if* it was a case of the application of one overarching methodology.

2.3 Evolution and revolution

According to the pure evolutionary perspective, scientific knowledge *always* grows somewhat gradually. However, this succession can be understood in either a Lamarckian way (new features are added/acquired and then inherited by the next generation) or in a Darwinian way (new features come by mutation and survive if they fit the surrounding better than the competitors). Positivism has the former view and Karl Popper (1902-1994) the latter (see Chapters 3.4 and 3.5, respectively). The 'surrounding' is constituted by both competing theories and empirical observations. All kinds of evolutions, Popper claims, can be seen as processes of trial and error where only the fittest survive. That is, both amoebas and scientists learn by trial and error; the difference between them is mainly one of consciousness. This difference, however, brings with it another and more important difference. Often, an animal or a species dies if it fails to solve an important problem, but a researcher who fails does not normally die, only his hypothesis does.

A scientific revolution is not like a mutation adapting to an ecological niche or a change of some epigenetic conditions. It is more like a radical change of the whole ecological system. The worldview, the fundamental values, the way of conducting research, and the way of thinking and presenting results are changed.

The development of natural-scientific knowledge during the seventeenth century is often referred to as 'the scientific revolution'; the emergence of the social sciences and the 'scientification' of the humanities take place mainly in the twentieth century. The scientific revolution brought with it a new view of nature and, consequently, new requirements on explanations. Since nature was no longer regarded in itself as containing any goals, teleological explanations came in disrepute. Explanations, it was claimed, should be made in terms of mechanical interaction, and laws should be stated in terms of mathematical relationships. 'The book of nature is written in the language of mathematics', as one of the great inaugurators of the scientific revolution, the physicist Galileo Galilei, claimed. With the emergence in 1687 of Isaac Newton's book *Philosophiae Naturalis*

Principia Mathematica the revolution was completed in physics. Nicolaus Copernicus (1473-1543) was a famous forerunner, and apart from Galileo Galilei (1564-1642) and Isaac Newton (1643-1727) we find physicists (astronomers) such as Tycho Brahe (1546-1641) and Johannes Kepler (1571-1630).

The scientific revolution was not confined to physics. Starting in the Renaissance, revolutionary changes took place even in medicine. Andreas Vesalius (1514-1564) and William Harvey (1578-1657) were key agents. Men like Claude Bernard (1813-1878), Louis Pasteur (1822-1895) and Robert Koch (1843-1910) might be said to complete this more extended revolution.

The anatomical work of Vesalius paves the way for Harvey's new physiological theories. Also, Vesalius reinstates the importance of empirical observations in medicine and, thereby, indicates that the old authorities, in particular Galen, had not found all the medical truths, and that these authorities had even made great mistakes. This is not to say that Harvey's theory of the circulation of the blood lacked empirical problems. For example, how the blood managed to circulate through the tissues in the periphery (the capillary passage) of the vessel system was still a mystery. It was only solved later by Marcello Malpighi (1628-1694) who, in 1661, with his microscope, managed to observe the small vessels (capillaries) between the arterial and the venous sides of a frog lung. And it was not until oxygen was discovered at the end of the eighteenth century, that the combined heart-lung function could be fully understood. (We will discuss Harvey's scientific achievement more at length in Chapter 4.8).

The post-medieval revolution in physics and medicine went hand in hand with a radical re-thinking in philosophy. Very influential was the French philosopher René Descartes (Latin: Cartesius, 1596-1650), who claimed that all animals are just like machines, that the human body is also merely a machine, but that (in contradistinction to bodies of animals) it is connected to a soul. Souls have consciousness and free will; they exist in time but are completely non-spatial. Despite their non-spatiality they can interact with 'their' human body in the epiphysis (the pineal gland); and in some way a soul can even permeate 'its' body.

According to Cartesianism, apart from actions and events caused by free will, all internal bodily functions and externally caused bodily behavior

should be explained the ways machines are explained. To Descartes, this means explanations by means of mechanisms where the movements of some parts cause (push) other parts to move in a certain way. All teleology, i.e., everything that has to do with goals, purposes, and design, was removed from the body as such. The clockwork became rather quickly a metaphor for the description and explanation of internal bodily processes (Figure 2). Its purpose, to show what time it is, is externally imposed on it by its designer and users. Once in existence, a clock behaves the way it does quite independently of this externally imposed purpose. Similarly, Cartesians thought that God had designed the human body with a purpose in mind, but that nonetheless the internal functioning of the body should be purely mechanistically explained. It was not until the Darwinian revolution that the concept of a godly or a pre-given natural design left biology for the purely causal concept of 'natural selection'.

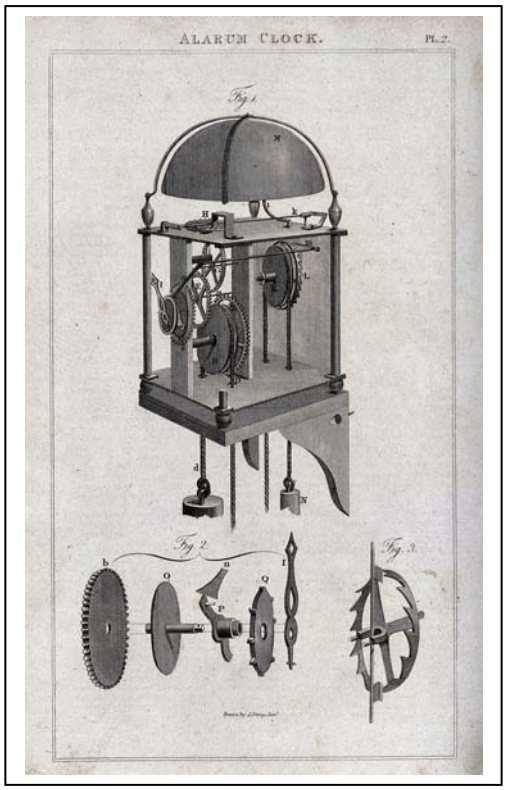


Figure 2: *With its cogwheels, the clockwork early became a metaphor for the mechanistic worldview.*

2.4 The concept of paradigm

After the demise of modern positivism in the 1960s, two philosophers of science came to dominate the Anglo-American scene, the Austrian Karl Popper (1902-1994) and the American Thomas Kuhn (1922-1996). Popper defends a non-positivist but evolutionist perspective, whereas Kuhn stresses revolutions; some of Popper's other views are presented later (Chapters 3.5, 4.2, and 6.3). Kuhn's most famous book has the title *The Structure of Scientific Revolutions* (1962). He divides scientific developments into two kinds, revolutionary science and normal science; in the latter the development is evolutionary. Crucial to Kuhn's position is his concept of paradigm, and we will use this section to explain it. In normal science a paradigm is taken for granted, in revolutionary science one paradigm becomes exchanged for another. Kuhn was an historian of physics, and he claims that the concept of paradigm is needed if one wants to understand the history of physics. In our opinion, the concept can also be applied to the history of medical science. As a matter of fact, using examples only from the history of medicine, the Pole Ludwik Fleck (1896-1961) put forward ideas similar to those of Kuhn before Kuhn. Fleck's most famous book has the title *The Genesis and Development of a Scientific Fact* (1935). Instead of *paradigms* and *scientific communities* he talks of *thought-styles* and *thought collectives*.

Before we continue with our exposition of paradigms, we ask the reader to bear in mind that our presentation brackets the conflict between epistemological realism (the view that we can acquire at least partial knowledge of the world) and social constructivism (the view that all presumed knowledge pieces are merely human conceptual artifacts). This issue will be dealt with in Chapter 3.5. Let us just mention that Fleck is a social constructivist, Popper an epistemological realist, and that Kuhn is ambiguous. In the best-seller mentioned, Kuhn says: "I can see in their [Aristotle, Newton, Einstein] succession no coherent ontological development. […] Though the temptation to describe that position as relativistic is understandable, the description seems to me wrong (Kuhn 1970, p. 206)." In a late interview he says:

> I certainly believe in the referentiality of language. There is always something about referential meaning involved in

> experience that tells you whether it is used to make true or false statements. There is a sense, a deep sense, in which I absolutely believe in the correspondence theory of truth. On the other hand, I also believe it's a trivial sort of correspondence (Kuhn 1994, p. 166).

When in what follows we talk about Kuhn, we will disambiguate him as a realist; social constructivists, on the other hand, try to disambiguate him as being one of them, even as one of their founding fathers.

According to social constructivists, we cannot come in contact with nature at all; according to Popper and Kuhn, we can never come into contact with nature by means of a purely passive reception. Popper calls such a passive view of mind a 'bucket theory of the mind', i.e., he thinks that minds should *not* be regarded as empty buckets that without any constructive efforts of their own can be filled with various kinds of content. Popper's and Kuhn's view is instead that, metaphorically, we can never see, take in, or 'receive' nature without artificially constructed glasses. In science, these glasses consist of partly unconsciously and partly consciously constructed conceptual-perceptual frameworks; outside science, our cognitive apparatus makes such constructions wholly without our conscious notice.

This constructive view brings with it an epistemological problem. Even if one finds it odd to think that *all* the features observed through the glasses are effects of the glasses in the way colored glasses make *everything* look colored, one has nonetheless to admit that in each and every singular case it makes good sense to ask whether the feature observed is a glasses-independent or a glasses-created feature. In fact, when telescopes and microscopes were first used in science, optical illusions created by the instruments were quite a problem. According to the Popper-Kuhn assumption, whereas ordinary glasses can be taken off, the epistemological glasses spoken of can only be exchanged for other such glasses, i.e., there is no epistemologically direct seeing.

Continuing with the glasses metaphor, one difference between Popper and Kuhn can be stated thus. Popper thinks it is rather easy to change glasses, whereas Kuhn thinks that old glasses are next to impossible to take off. Scientific revolutions are possible, he thinks, mainly because old

scientists have to retire, and then youngsters with glasses of a new fashion can enter the scene. This difference between Popper and Kuhn reflects different views of language on their part. Kuhn has a more holistic view of language meaning than Popper, and he thinks that meaning patterns are in their essence social phenomena. Popper gives the impression of believing that it is rather easy to come up with new semantic contents and by means of these make new hypotheses, whereas Kuhn's view (with which we agree) implies that this cannot be so since the new concepts should (a) conceptually cohere with the rest of the scientists' own conceptual apparatus and then also (b) to some extent socially cohere with the rest of his scientific community.

Kuhn distinguishes between two parts of a paradigm: (1) a disciplinary matrix and (2) exemplars or prototypes. A disciplinary matrix consists of a number of group obligations and commitments; to be educated into a researcher within a scientific community means to be socialized into its disciplinary matrix. These matrices have several aspects. One is the rather explicit prescription of norms for what kind of data and problems the discipline should work with. These norms answer questions such as 'Can purely qualitative data be accepted?', 'Are mathematical models of any use?', and 'Are statistical methods relevant?' In the history of medicine, this part of the disciplinary matrix is easily discernible in the microbiological paradigm that arose at the end of the nineteenth century; we are thinking of Robert Koch's famous postulates from 1884. In order to prove that a specific microorganism is the cause of a specific disease, four norms, says Koch, have to be adhered to:

i) the specific microorganism must be found in all animals suffering from the specific disease in question, but must not be found in healthy animals;
ii) the specific microorganism must be isolated from a diseased animal and grown in a pure culture in a laboratory;
iii) the cultured microorganism must cause the same disease when introduced into a healthy animal;
iv) the microorganism must be able to be re-isolated from the experimentally infected animal.

The other aspects of the disciplinary matrix are called ‘metaphysical assumptions’ and ‘symbolic generalizations’, respectively. According to Kuhn, even the basic empirical sciences have to contain assumptions that are not directly based on their experiments and observations. The reason is that experiments and empirical data gathering are only possible provided some presuppositions. Such non-empirical presuppositions do therefore often, says Kuhn, take on the appearance of being definitions and not laws of nature, which means that they also seem impossible to contest. If empirical results contradict them, the natural response is to question the accuracy of the observations and/or the researcher’s skill, not these basic presuppositions. If they are quantitative relationships, they can be called symbolic generalizations. Think of this view: ‘necessarily, velocity equals distance traversed divided by time used’, i.e., ‘$v = s / t$’. Does it state a natural law or a definition? Can we even think of a measurement that could falsify this functional relationship? If not, shouldn’t we regard it as a definition of velocity rather than as a natural law? According to Kuhn, Newton’s three laws of motion were once regarded almost as by definition true; it was, for instance, unthinkable that the second law could be falsified. This law says that the forces (F) acting on a body with mass m and this body’s acceleration (a) are numerically related as ‘$F = m \cdot a$’. To Newtonians, it had the same character as ‘$v = s / t$’.

In order to appreciate the point made, one has to bear a philosophical truth in mind: *necessarily, there has to be undefined terms*. The quest for definitions has to come to an end somewhere – even in science. If one has defined some A-terms by means of some B-terms and is asked also for definitions of these B-terms, one might be able to come up with definitions that are using C-terms, but one cannot go on indefinitely. On pain of an infinite regress, there has to be undefined terms, and the last semantic question cannot be ‘how should we *define* these primitive terms?’ but only ‘how do we *learn* the meaning of these undefined primitive terms?’ The situation is the same as when a child starts to learn his first language; such a child simply has no terms by means of which other terms can be defined.

What Kuhn calls metaphysical commitments may take on the same definitional character as symbolic generalizations do, but they are not quantitative. They can be views, he says, such as ‘heat is the kinetic energy

of the constituent parts of bodies' and 'the molecules of a gas behave like tiny elastic billiard balls in random motion'.

Whatever one thinks about the particular cases mentioned above, it is true that it is impossible to get rid of non-empirical presuppositions altogether. Assume that someone claims (as traditional empiricists and positivists do) that all knowledge apart from that of logic and mathematics has to be based solely on empirical observations. How is this claim to be justified? Surely, it cannot be justified by observations alone.

A disciplinary matrix with its methodological norms, symbolic generalizations, and metaphysical commitments tells its 'subjects' how to do research, with what general assumptions the objects of investigation should be approached, and what can count as good explanations.

What then are exemplars or prototypes, the other part of a paradigm? Kuhn has a view of language and language acquisition that in some respects is similar to Ludwig Wittgenstein's (1889-1951) later language philosophy. We learn how to use words mainly by doing things when talking and by doing things with words. There is no definite once-and-for-all given semantic contents of words. In order to learn even scientific terms such as those of Newtonian mass and acceleration, one has to *do* things with these terms. One has to solve theoretical problems and/or conduct experiments with their help. In isolation from such situations a formula such as '$F = m \cdot a$' is merely a mathematical formula that has nothing to do with physics. An exemplar is a prototypical example of how to solve a theoretical or experimental problem within a certain paradigm. In order to understand assertions made within a paradigm, one has to learn some of its exemplars. In order to understand an obsolete scientific theory, one has to understand how it was meant to be applied in some crucial situations.

Medical people familiar with learning diagnostics can perhaps understand Kuhn's point by the following analogy. At first one learns a little about the symptoms of a disease (the meaning of a scientific term) by looking at pictures (by having this meaning explained in everyday terms), after that one improves this learning by looking at typical real cases (exemplars) and, thirdly, by working together with a skilled clinician one becomes able to recognize the disease even in cases where the symptoms of a certain patient are not especially similar to those known from the medical textbooks and the typical cases. Good clinicians are able to

transcend the learning situations, and the same goes for competent language users. They are able to use their competence in completely new situations. Exemplars, and what one learns through them, cannot be reduced to systems of already defined terms and verbally explicit rules or standards.

Having distinguished between the exemplars and the disciplinary matrix of a paradigm, one should note their connection. Through the exemplars one learns how to understand and apply the terms that are used in the disciplinary matrix. Koch's postulates, for example, are connected with experimental practices, without which they cannot be substantially understood. Koch did not just put forward his abstract postulates he created a certain laboratory practice, too. Over time, the exemplars may change a bit.

Two things must now be noted. First, paradigms and sub-paradigms are like natural languages and their dialects. It is hard to find clear-cut boundaries between a paradigm (language) and its sub-paradigms (dialects) and sometimes even between one paradigm (language) and another. But this vagueness does not make the concepts of paradigms and languages meaningless or non-applicable. Rather, they are impossible to do without. Second, the glasses metaphor has at least one distinct limit. Whereas colored glasses color everything, a paradigm does not. To the contrary, so far in the history of science, paradigms are normally during their whole life-time confronted by some (albeit shifting) empirical data that ought to, but does not at the moment, fit into the paradigmatic framework. For instance, Newtonian mechanics did *never* make accurate predictions of all the planetary orbits. In Chapter 3.5, we will claim that anomalies can, so to speak, be nature's way of saying 'no' to a paradigm.

In Chapter 6, we will present what we take to be the dominant paradigm in present-day medical science, 'the clinical medical paradigm'. Here, we will present the overarching medical paradigm that arose in Ancient Greece and which dominated Europe and the Arab world during the medieval times. Its origin rears back to the famous Hippocrates (460-377 BC). But its most prominent figure is the Roman physician Galen (131-200), hence its name: 'the Galenic paradigm'. Another possible name would be 'the teleological four humors paradigm'.

Why do organs such as livers and hearts behave the way they do? A typical explanation in Ancient times referred to the goals or purposes of the organs. This is the way we normally explain actions of human beings: ‘why is he running?’ – ‘he is trying to fetch the bus’. The action is explained by the existence of a goal inside the acting person. In Galenic explanations, it is as if the organs have inside themselves a certain purpose. Such a view was made systematic already by the Ancient Greek philosopher Aristotle (384-322 BC). According to him, even if some things may be merely residues, most kinds of thing have a certain ‘telos’, i.e., something ‘for the sake of which’ they are behaving as they do when they are functioning properly. The world view was teleological not mechanical. It was thought that nature does nothing without some kind of purpose.

According to Galen, the body contains four fluids and three spirits (Latin ‘spiritus’ and Greek ‘pneuma’), which are distributed via the blood vessels - the veins and the arteries. The fluids are: sanguine, yellow bile, black bile, and phlegm. Sanguine originates in the liver, where material from the intestines, i.e., originally food, is converted into blood. If there is too much sanguine fluid, the surplus is transformed into yellow bile. Black bile is useless sanguine, which is collected in the spleen. Phlegm is associated with the brain. The spirits are: animal spirits (Greek: ‘pneuma psychikon’), vital spirits (‘pneuma zooikon’), and natural spirits (‘pneuma physikon’).

Animal spirits are produced in the brain, distributed along the nerves, and their functions (goals) are related to perception and movement. Vital spirits are produced in the left part of the heart, and they are distributed in the body by means of the arteries. Their function (goal) is to vitalize the body and to keep it warm. Natural spirits are, like the sanguine and yellow bile, produced in the liver, and they go with the liver created blood first to the right part of the heart; their function (goal) is to supply the organs and tissues with nutrition.

The bloodstream (containing all four fluids) is assumed to move in the veins as follows. It goes from the liver to the right part of the heart (see Figure 3 below). Then it moves (but not by means of pumping) out into the rest of the body (organs, extremities, and tissues) where it is absorbed. It moves as slow or as fast as it is produced in the liver and absorbed in the body. The vital spirit is (according to the Galenic paradigm) created in the left part of the heart as a combination of blood from the right side of the

heart (assumed to penetrate via tiny pores in the wall of the heart, the septum) and air from the lungs (assumed to arrive in the left part of the heart via what was referred to as the vein like arteries – today we define these vessels as veins: vena pulmonalis, as the direction of the bloodstream in these veins go from the lungs to the heart).

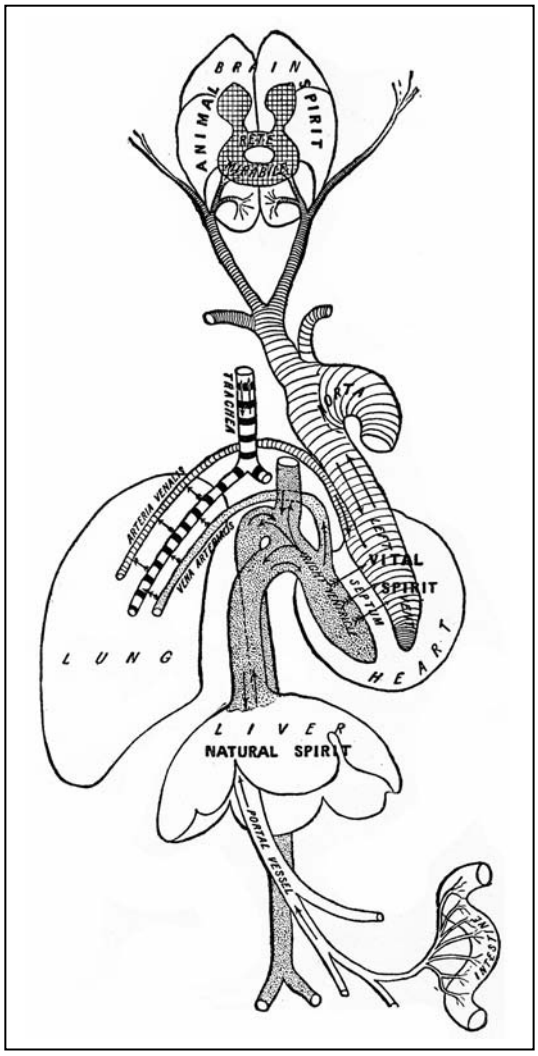


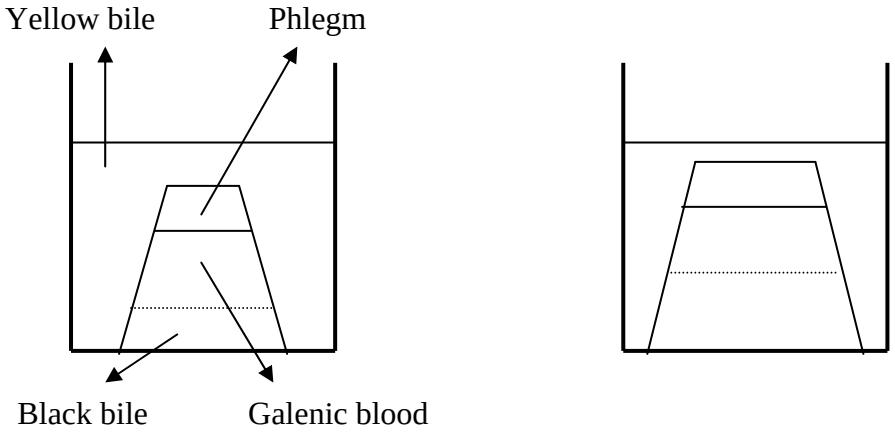
Figure 3: *Here is the Galenic model for the movements of the blood. 'Spiritus animalis' was assumed to be produced in the brain, 'spiritus vitalis' in the left part of the heart, and 'spiritus naturalis' in the liver.*

One goal of the whole body is to try to keep the amount of the four fluids in a certain balance. Stable unbalances explain psychological character traits (temperaments) and accidental unbalances explain diseases. Sanguine persons have too much sanguine or Galenic blood (Latin for

blood: sanguis), choleric persons too much yellow bile (Greek: chole), melancholic persons too much black bile (Greek: melas chole), and phlegmatic persons too much phlegm (Greek: phlegm). The way to cure a disease is to restore balance. Sometimes, the body tries (and often succeeds) to do this automatically; as for instance when we are coughing up phlegm. Sometimes, however, the balance has to be restored artificially by a physician. Independently of which of the fluids there are too much, the cure is blood letting. Bloodletting automatically excludes most of the fluid of which there is too much.

This ‘four humors (fluids) pathology’ is not as odd as it first may seem today. The Swedish medical scientist, Robin Fåhraeus (1888-1968), the man who invented the blood sedimentation test, has suggested that the true kernel of this doctrine might be understood as follows – if it is accepted that the four fluids could be mixed in the blood vessels. If blood is poured into a glass jar, a process of coagulation and sedimentation starts. It ends with four clearly distinct layers: a red region, a yellowish one, a black one, and a white one (Figure 4, left). There is a reddish column in the middle and upper part of the jar; it might be called sanguine or ‘Galenic blood’. As we know today, it consists of coagulated red blood cells that permits light to pass through. The lowest part of the same column consists of sediment that is too dense to permit light to pass through. Therefore, this part of the column looks black and might be referred to as the ‘black bile’. On the top of the column there is a white layer, which we today classify as fibrin; it might correspond to Galen’s ‘phlegm’. The remaining part is a rather clear but somewhat yellowish fluid that surrounds the coagulated column in the middle. It might be called ‘yellow bile’, but today we recognize it as blood serum. But there is even more to be said.

Fåhraeus showed that when such a glass of blood from a healthy person is compared with a similar one from a person suffering from pneumonia (caused by bacteria), the relative amounts of the four fluids differ (Figure 4, right). In the sick person’s glass, the proportions of the ‘black bile’ and the ‘phlegm’ have increased, whereas those of the ‘yellow bile’ and the ‘Galenic blood’ have decreased. Such an observation is some evidence in favor of the view that an unbalance between these four fluids can cause at least pneumonia.



The composition of the four fluids in the whole blood from:
(i) a healthy person (ii) a person with pneumonia

Figure 4: *How blood is stored in different layers when poured into a glass jar. The figure might suggest why Galen thought of diseases as being caused by changes in the composition of the four fluids.*

A scientific paradigm may cohere more or less with a surrounding more general world-view. The doctrine of the four fluids (or humors) and the four temperaments were very much in such conformance with other views at the time (Table 1). Each of the four fluids and temperaments was seen as a combination of one feature from the opposition hot–cold and one from the opposition dry–wet. Furthermore, the same was held true of the four basic elements of dead nature: fire (hot and dry), water (cold and wet), earth (cold and dry), and air (hot and wet).

Planets	*Elements*	*Seasons*	*Fluids*	*Organs*	*Temperaments*
Jupiter	Air	Spring	Blood	Liver	Sanguine
Mars	Fire	Summer	Yellow bile	Gall bladder	Choleric
Saturn	Earth	Autumn	Black bile	Spleen	Melancholic
Moon	Water	Winter	Phlegm	Brain	Phlegmatic

Table 1: *A summary of how the ancient Greeks thought of connections between the macrocosmic and the microcosmic worlds as well as between psychological and organological features.*

The four fluids mentioned were also thought to be influenced by different heavenly bodies and the latter's relative positions. Blood was supposed to be controlled by Jupiter, the yellow bile by Mars, the black bile by Saturn, and the phlegm by the moon. An unfortunate planetary constellation was assumed to cause diseases by creating unbalances between the four fluids. Think, for instance, of the old diagnostic label 'lunatic'. Observe that, just like the moon, several mental illnesses have cyclical phases.

Galen had made extensive and significant empirical studies. For example, he had compressed the ureter in order to show that the kidneys produce urine. He even made public dissections of living animals. Such empirical and experimental research was not continued during the medieval ages. Instead, the doctrines of Galen became dogmas canonized by the church. During the medieval era, roughly, medical researchers sat in their chambers studying the writings of Galen trying to develop it only theoretically. But in the sixteenth century some physicians started to criticize the doctrines of Galen – if only in an indirect way.

We have claimed, with Kuhn, that paradigms have a special inertia because of the holistic and social nature of conceptual systems, but more straightforward external causes can help to conserve a paradigm, too. Figure 5 below illustrates a typical fourteenth century anatomy lesson. In an elevated chair we find the university teacher. He is elevated not only in relation to the corpse but also in relation to the students, his assistant teacher (demonstrator) and the dissector (or barber). The teacher is reading aloud from Mondino de' Liuzzi's (1275-1326) compendium *Anathomia* (1316); primarily it is composed of the writings of Galen on anatomy, which contains few and very simplistic pictures. The demonstrator is pointing to the different organs and structures while the teacher is describing them from the text; the dissector is cutting them out. As explicitly stated by one of Vesalius' contemporary colleagues from Bologna, Matthaeus Curtius (1475-1542), it was beneath the dignity of teachers to approach the corpses. This attitude went hand in hand with the view that it was also unnecessary for already learned men to study the corpse – Galen had already written the description needed.

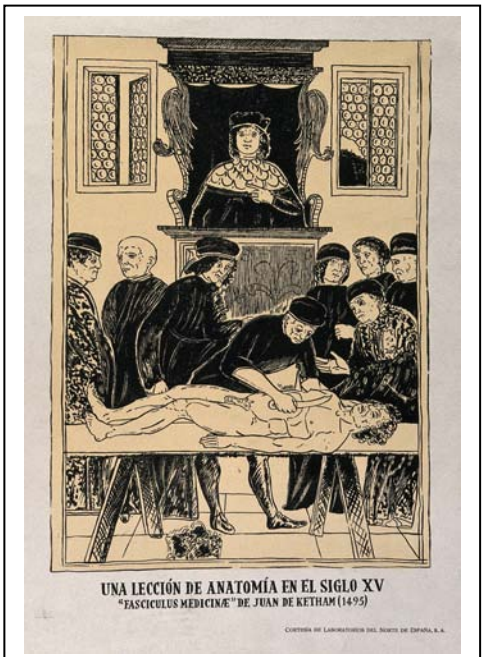


Figure 5: *A medical teacher reading aloud from a compendium. To the right of the picture we find the demonstrator, and in the middle we see the dissector with his knife. All the others are medical students.*

Seen in its historical context, a seemingly simple act performed by Vesalius shows itself to be consequential. Vesalius did not care about the social dignity of the medical teachers; he began to make dissections himself. He wanted to study the human anatomy more systematically, more carefully, and in more detail than before. Also, his anatomical drawings set a precedent for future detailed and advanced anatomical illustrations (Figure 6 below).

When the Galenic paradigm was first questioned, it was so deeply integrated into both the worldview of the church and the values of the secular society that it was hard to criticize. Nonetheless the Galenic views were shown to be confronted by obvious anomalies. According to Galen, there are pits in the walls between the right and left side of the heart. Vesalius stated that he was not able to observe them. Nonetheless, neither Vesalius nor his contemporary and subsequent colleagues made any head-on attack on Galen. The medical revolution started in a gradual way. Even William Harvey, who a hundred years later discovered the blood

circulation, avoided being directly critical of Galen. In fact, often he tried to strengthen his arguments by saying that his view was also Galen's view.

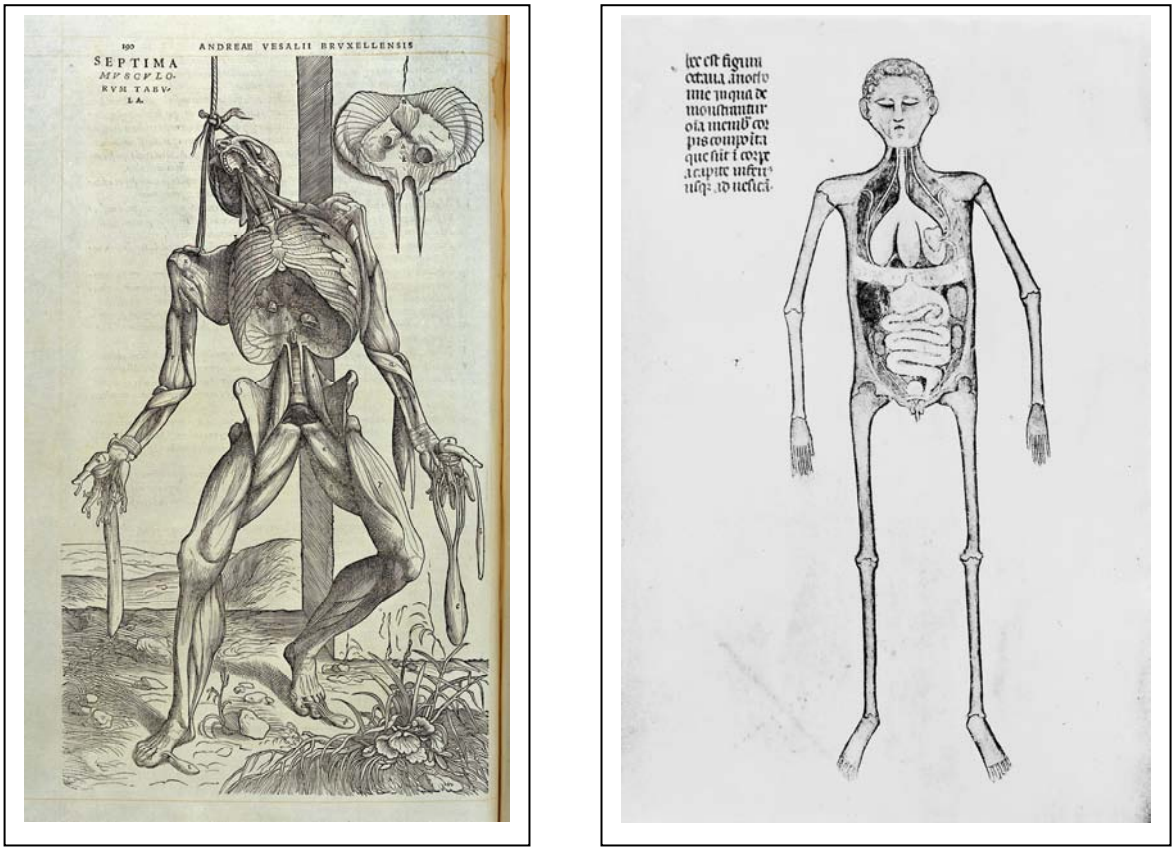


Figure 6: *The left picture from Vesalius' "De humani corporis fabrica" is considerably more detailed than the right one from the compendium of Mondino (1316). The latter is primarily a summery of Galen's anatomical writings, which did not contain any illustrations. Mondino, who held a position at the University of Bologna, was one of the first to present anatomical pictures.*

Independently of each other and before Harvey, the Italian anatomist Realdo Colombo (1516-1559) and the Spanish anatomist Miguel Serveto had observed what we call lung circulation (or the 'small circulation'). But according to Galen there is blood merely in the vein system and in the right side of the heart. The arterial system was supposed to contain a composition of air from the lungs and blood penetrating from the right part of the heart via tiny pores in the heart wall – creating spiritus vitalis. The

windpipe (trachea) was also classified as an artery and was supposed to be directly connected with the arterial system through the lungs.

2.5 Generating, testing, and having hypotheses accepted

Many empirical scientists find it easy to distinguish between a first stage of research where they are merely thinking about a problem, which one day ends when they find or create a hypothesis that might solve the problem, and a second stage in which they are testing their hypothesis. This experience is in some philosophies of science elevated into an important distinction between two kinds of research contexts, 'the context of discovery' and 'the context of justification', respectively. Positivists and Popperians (see Chapters 3.4, 3.5, 4.4 and 6.3) claim that the philosophy of science should be concerned only with the context of justification; the context of discovery is, they claim, only of relevance for psychology and sociology of knowledge. In a sense, with Kuhn, we disagree. According to him, a paradigm supplies at one and the same time both a context of justification and a context of discovery. There is an inner connection between these two types of contexts. A paradigm is fertile soil for certain kinds of specific hypotheses while simultaneously justifying the general structure of these hypotheses.

When a paradigm is taken for granted, the development of knowledge is in an evolutionary phase, and in this phase hypotheses do – just like apples – fall close to the tree-trunk. For instance, as soon as the microbiological paradigm was established (at the end of the nineteenth century), the microbiologists rather quickly both discovered and justified many specific hypotheses about different bacteria as being causes of various diseases. As soon as it was *in principle* accepted that bacteria might cause diseases, many such pathogenic agents were isolated rather promptly by means of the microscope and Koch's postulates. Here is a list:

1873 The Leprosy bacterium	Gerhard A Hansen
1876 The Anthrax bacterium	Robert Koch
1879 The Gonococci bacterium	Albert Neisser
1880 The Typhus bacterium	Carl Ebert
1882 The Tuberculosis bacterium	Robert Koch
1883 The Cholera bacterium	Robert Koch

1883 The Pneumococci bacterium	Carl Friedländer
1883 The Streptococci bacterium	Julius Rosenbach
1884 The Staphylococci bacterium	Julius Rosenbach
1884 The Diphtheria bacterium	Friedrich Loeffler
1884 The Tetanus bacterium	Arthur Nicolaier
1885 The Escherich Coli bacterium	Theodor Escherich
1885 The Meningococci bacterium	Anton Weichselbaum
1888 The Salmonella bacterium	August Gaertner
1889 The Ulcus molle bacterium	Augusto Ducrey
1892 The Haemophilus bacterium	Richard Pfeiffer
1894 The Plaque bacterium	A. Yersin & S. Kitasato
1896 The Brucella bacterium	Bernhard Bang
1897 The Botulism bacterium	Emile van Ermengen
1898 The Dysenteri bacterium	Kiyoshi Shiga
1900 The Paratyphus bacterium	Hugo Schottmüller
1905 The Syfilis bacterium	F. Schaudinn & E. Hoffman
1906 The Whooping-cough bacterium	J. Bordet & O. Gengou

Let us now present some aspects of the pre-history of this rapid development. Hopefully, this can give a vivid view of how many presuppositions there are around in empirical science – both for generating and justifying specific hypotheses.

A Dutch lens grinder and drapery tradesman, Antonie van Leeuwenhoek (1632-1723), is often called the father of microbiology. Improving on both existing microscopes and preparation techniques, he managed to see things that no one had seen before. Among other things, he examined material from his own mouth and observed an entire zoo of small living organisms, provided he had not just drunk hot coffee. He reported his observations to the Royal Society in London. At first, his reports were received as simply interesting, but when he reported his observations of microorganisms, which he called ‘animalcules’ (Figure 7), he was met with skepticism. The legendary secretary of the society, Henry Oldenburg (1615-1677), corresponded with Leeuwenhoek and asked the latter to describe his procedure in more detail; eventually a respected team was sent to Holland to check the observations, which they vindicated.

Leeuwenhoek became a famous and distinguished member of the Royal Society, and many persons came to look through the microscopes in order to see this new micro-world; or, by the way, to convince themselves, just

like Leeuwenhoek, that the hypothesis of spontaneous generation must be wrong. During the period between Leeuwenhoek and Pasteur, many researchers were preoccupied with observing the microbes. They studied how microorganisms proliferate, whether they occur spontaneously or not, under what circumstances they die, and so on. Microorganisms were observed in infected wounds, but they were for quite a time thought to be the effect and not the cause of the infection. The existence of efficient microscopes was a necessary but not a sufficient condition for discovering bacteria and developing bacteriology.

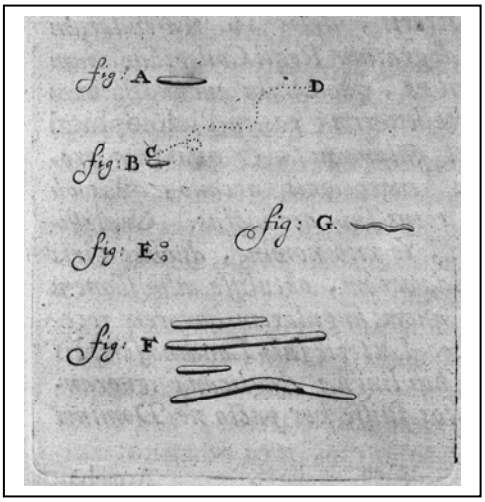


Figure 7: *The picture shows a drawing of Leeuwenhoek's small animals or 'animalcules', which he observed in his microscope.*

During the seventeenth century, Galen's view that blood is the location of the pathogenesis of diseases dominated, but with respect to fever diseases there were competing theories around; mainly, the contact theory and the miasma theory. According to the contact theory, as the name makes clear, a disease might be caused by means of contagion from a diseased person. For some time, this theory was of great practical importance, especially in Italy and the northern Mediterranean. It was the theoretical basis of the quarantine regulations for merchant vessels. Its popularity started to decrease at the beginning of the nineteenth century. Theoretically, it was hard to explain why some patients became ill and some not, albeit being exposed to the same contagion. Speculating about external factors, one can note that the quarantine regulations were very

expensive for the trading companies. ‘Quarantine’ is an Italian word that means forty, understood as the forty days that a ship had to wait before it was allowed to approach an Italian harbor. If no one on board became ill during the quarantine period (mostly, it was smallpox that one was afraid of), it was regarded as safe to let the ship into the port. Viewed from today’s knowledge, it is remarkable how close this quarantine period is to the incubation period for many diseases.

According to the miasma theory, diseases are caused directly by something in the air and indirectly by something in the surroundings. Sick people in slum districts were supposed to have breathed poisoned air, and people living in marshlands were often infected with malaria. ‘Malaria’ is an Italian word that means bad air. In retrospect, it is obvious that the contact theory is more in conformance with modern microbiology, but the miasma theory had many powerful supporters among renowned frontier physicians and scientists at the time. For instance, the German pathologist Rudolf Virchow (1821-1902), strongly rejected the contact theory and later on the microbiological paradigm. His main reason was that microbiology presupposed mono-causality, whereas the miasma theory allowed multi-factorial explanations of diseases to come more naturally.

Among people in the nineteenth century English hygienist or sanitary movement, the miasma theory was popular too. Edwin Chadwick (1800-1890), who was a lawyer and rather skeptical towards the medical profession, maintained in a report in 1842 that the only way to prevent diseases was to eliminate poverty and improve the laboring population’s living conditions, their homes as well as the sewage and garbage collection system. However, in 1854 the medical epidemiologist John Snow (1813-1858), who did not support the miasma theory, presented a report about the hygienic standards around water, in which he claimed concisely that it must have been pollution of the water in one specific pump that was the cause of ninety-three persons’ death by cholera. Snow removed the handle of the pump, the cholera epidemic subsided, and Snow became a hero. Let it be said, that even before his intervention the epidemic had begun to decrease.

Another bit in the medical research puzzle that eventually made Leeuwenhoek’s ‘animalcules’ fit into a contact theory of diseases brings in

the English physician Edward Jenner (1749-1823) and vaccination. But before vaccination there was variolation.

From the Arab medicine of the twelfth and the thirteenth centuries, the variolation technique was in the eighteenth century imported into Europe; probably, it was first discovered in Chinese medicine. In variolation, healthy individuals are deliberately exposed to smallpox (variola) in order to become immune. The preferred method was rubbing material from a smallpox pustule from a mild case into a scratch between the thumb and forefinger. Unfortunately the variolation technique was not safe, and it was met by considerable opposition at the time of its introduction. Also, it presupposes contagion as a disease cause.

At the end of the eighteenth century, Edward Jenner introduced a new and safer technique. Jenner was a general practitioner in rural England. He had spent his childhood in the countryside, and among the rural population it was said that milkmaids that had been exposed to cowpox did never get smallpox. A day in May 1796, Jenner tested this layman hypothesis by inoculating material taken from cowpox-infected blisters from a milkmaid, Sarah Nelmes (Figure 8), into an eight year old boy, his gardener's son, James Phipps. The boy got fever for a few days, but was soon healthy again. Today we would refer to this as an immunization trial procedure.

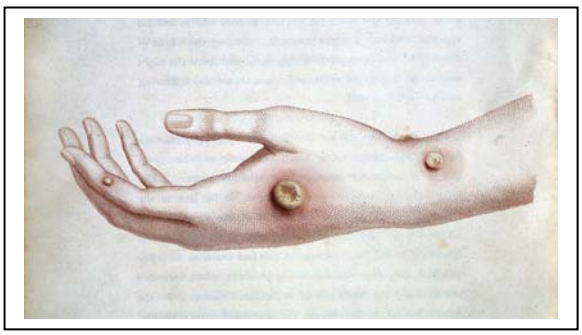


Figure 8: *Cowpox infected blisters from the milkmaid, Sarah Nelmes.*

Six weeks later, Jenner inoculated smallpox material into the boy. Due to the fact that smallpox is rather virulent, one might have expected that the boy would become very ill or die. But fortunately he did not become sick at all. We shall discuss the research ethical aspects of this case further in

Chapter 10. In the present context, we want to stress that although Jenner's experiment supported his hypothesis, he had at first a hard time having it accepted. Being merely a family physician in the countryside, Jenner did not impress his academic colleagues at the universities. But even more, when they tried to repeat his experiment the results were not unambiguous. Today we know that in order to avoid second order effects and erroneous results, the cowpox material must be purified from other microorganisms as well as potentially allergic materials. Otherwise it might result in reactions such as fever, other infections, and skin reactions. Thus Jenner's skeptics had real reasons not to be convinced. However, eventually they succeeded in purifying the cowpox contagion, and the procedure was accepted. Nonetheless, it should be noted, there was still no reasonable theoretical explanation at hand. But the practical use and benefit of the procedure was very significant, especially for the military. At this time, after battles soldiers often died from smallpox or other infections diseases. In 1798 Jenner's 'An Inquiry into the Causes and Effects of the Variolae Vaccinae' was published, and it was soon translated into several languages. Napoleon supported Jenner's views, and had his entire army vaccinated.

Next the most famous man of the contagion story: Louis Pasteur (1822-1895). Pasteur was a chemist, not a physician. But perhaps this made it easier and not harder for him to believe and to show that without microorganisms there are no infection diseases. Since he was not committed to the current medical paradigms, he could conduct his scientific work without the disciplinary matrix of the medical scientific community. It took the latter a long time to accept Pasteur's contributions. It is often the case that anomalies in an old paradigm (here: variolation in relation to the miasma theory) becomes a positive core issue in a new paradigm.

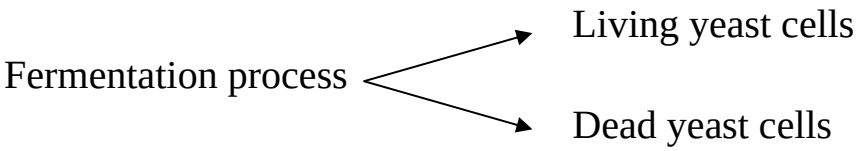
Two events put Pasteur on the right track. First, as a renowned chemist he had been entrusted with the task of examining the fermentation process at a vineyard. Sometimes these fermentation processes did not proceed to the end as expected – and the result tasted bad. Second, he happened to study an epidemic among silkworms.

Unlike many of his chemist colleagues, Pasteur did not find it odd to use a microscope when he studied the wine fermentation process. He made comparative studies of the processes in question, and was able to show that

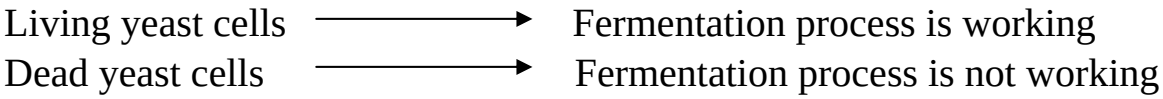
there were living (yeast) cells in the fermentation processes that resulted in normal wine but not in the other ones. Also, he managed to isolate these living yeast cells. He then conjectured that the living fermentation (yeast) cells found are the cause of, or a necessary condition for, the fermentation process. Since the presence of yeast fungi in a vineyard was supposed to kill yeast cells, it ought to be possible to take away the bad processes by reducing the amount of yeast fungi in the tendrils of the vines. So Pasteur did, and good wine production was successfully restored. Pasteur concluded that his hypothesis had been verified, but his colleagues were not convinced.

According to the common view, both microbes and yeast cells were products and not causes of the relevant processes. Yeast cells, be they alive or not, were supposed to play no role in the fermentation process itself. Pasteur turned this picture upside down and claimed that living yeast cells are causes and fermentation an effect.

Fermentation process according to the old theory:



Fermentation according to Pasteur:



Changing the direction of these arrows was also a precondition for Pasteur's reasoning about infectious diseases. When Pasteur studied the silkworms mentioned, he found a phenomenon similar to that in the fermentation processes. First, using the microscope, he saw a certain kind of microorganism in the sick silkworms that he could not see in the healthy ones. And then he managed to isolate even these organisms. In analogy with his fermentation hypothesis, he now conjectured that it was the presence of these microbes that caused the disease among the silkworms.

Pasteur's hypothesis was rejected by a number of his influential colleagues. For instance, the German chemist Justus von Liebig (1803-1873) claimed that only enzymes can regulate fermentation processes; even if there are yeast cells they are not able to influence the process. From today's perspective we can say as follows: Liebig's hypothesis that it is enzymes that regulate the process is chemically correct, but it is yeast cells that produce these enzymes.

Although the reception of Pasteur's hypothesis was not in all corners enthusiastic, the time was ripe for his ideas. His hypothesis spread quite rapidly, and several physicians took his ideas into serious consideration. Among the latter were the rural German general practitioner, Robert Koch (1843-1910) and the Scottish surgeon Joseph Lister (1887-1912).

Robert Koch read about Pasteur's results and began himself to isolate and make experiments with presumed pathogenic bacteria. In his most famous experiment, he infected healthy animals with what is now called anthrax bacteria, which he had isolated from the blood of sick animals, whereupon the infected animals got anthrax. Then, he was able to identify the same type of bacteria in the blood of these artificially infected animals. That is, he really used the 'Koch's postulates' that we mentioned in Chapter 2.3. He showed that it is possible to make animals sick by means of pathogenic bacteria.

Koch's international breakthrough came some years later when he discovered and isolated the tuberculosis bacteria (1884). The fact that cholera bacteria could cause epidemics supported John Snow's views and measures in London thirty years earlier. But more was needed in order to establish the microbiological paradigm beyond doubt.

It was the combination of the fruitfulness of Pasteur's ideas, the carefully conducted procedures of Koch, and the work of their successors in the next twenty years that finally established the microbiological paradigm. But some researchers were die-hards. As already mentioned, the prominent German pathologist, Rudolf Virchow never accepted this paradigm.

An interesting reaction to Koch's views came from a professor in dietetic chemistry in Munich, Germany, Max von Pettenkofer (1818-1901). He requested a bottle of cholera bacteria from Koch's laboratory, got it, and then he claimed to have drunk it without becoming ill; thereby saying

that Koch's views were wrong. We don't know whether he actually drank it or if he merely cheated.

Another man who read about Pasteur's results, and has become famous in the history of medicine, is the mentioned Joseph Lister. He is the man behind the antiseptic treatments of wounds. After the introduction of anesthesia in 1842, one might have expected that the status of surgery would increase, but this was not unanimously the case. The reason was that the new painless surgery also resulted in an increasing amount of surgical operations – with all the by now well known accompanying complications, especially infections. The mortality rate after amputations was high. In Lister's hospital, sixteen out of thirty-five patients died in 1864-1866. In some other hospitals the mortality rate was significantly higher. Lister learnt from his reading of Pasteur that bacteria might also be found in the air, and he concluded (in a kind of synthesis of miasma and contact theories) that it was airborne bacteria that were the main cause of post-operative infections (Figure 9).

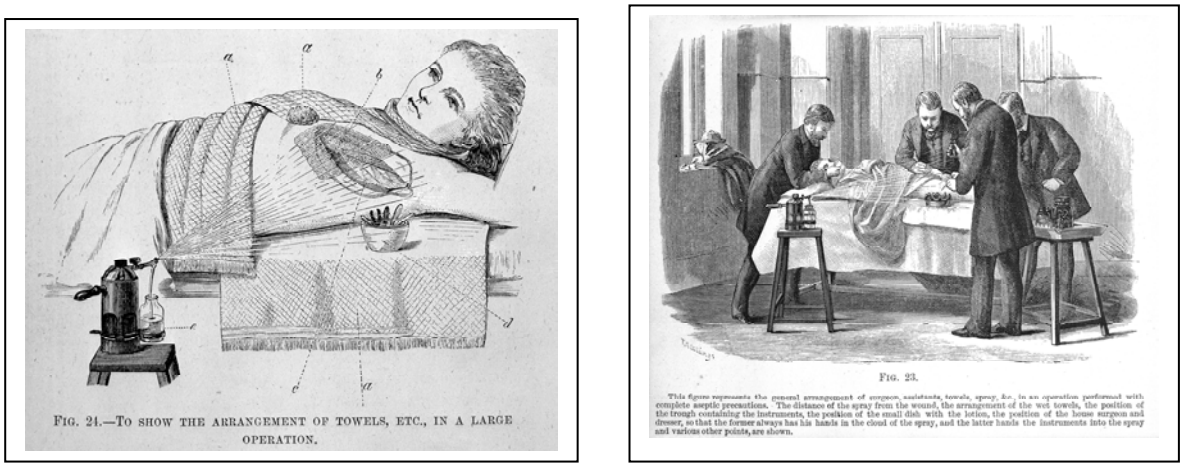


Figure 9: *The left picture shows how the Lister atomizer was supposed to work and the right picture how the antiseptic surgery worked in practice. Notice that the surgeons had no protective gloves and wore their own street clothes.*

In order to prevent pathogenic bacteria from infecting operation wounds, Lister had a carbon acid based suspension sprayed in the operation room – over the operation wound. Apparently he was influencing the air, and accordingly this technology was not in conflict with the miasma theory.

Lister began in 1867, and in 1870 he could report that now only six out of forty patients had died; quite an improvement. Even though Lister's procedure was uncomfortable for the surgeons and all the others in the operation room, his theory and practice were rather promptly accepted.

With these remarks we end our brief history of the emergence of microbiology and the scientific use of the microscope. Later, the microbiological paradigm got its share of anomalies. When the medical community was first faced with the symptoms that we now classify as symptoms of deficiency diseases, the microbiologists continued to search in the microscopes for specific bacteria that might be the causes.

Reference list

Barnes B, Bloor D. 'Relativism, Rationalism and the Sociology of Knowledge'. In Hollis M, Lukes S (eds.). *Rationality and Relativism*. Basil Blackwell. Oxford 1982.

Bernal JD. *The Social Function of Science*. Georg Routledge & Sons. London 1939.

Bernal JD. *Science in History. The Scientific and Industrial Revolution*. MIT Press. Cambridge Mass. 1983.

Bird A. *Philosophy of Science*. Routledge. London 1998.

Bloor D. *Knowledge and Social Imagery*. Routledge. London 1976.

Cartwright SA. < http://www.pbs.org/wgbh/aia/part4/4h3106t.html >

De Kruif P. *Microbe Hunters*. Harcourt Brace & Co. New York 1996.

Farrington B. *Greek Science*. Penguin Books. Harmondsworth. Middlesex, England 1953.

Feyerabend P. *Against Method. Outline of an Anarchistic Theory of Knowledge*. Verso. London, New York 1993.

Fleck L. *Genesis and Development of a Scientific Fact*. University of Chicago Press. London 1979.

Fåhraeus R. *The Art of Medicine – An Overview* (in Swedish). Bonniers. Stockholm 1944.

Hemlin S, Allwood CM, Martin BR (eds.). *Creative Knowledge Environments. The Influences on Creativity in Research and Innovation*. Edward Elgar Publishing. Cheltenham Glos 2004.

Hessen B. The Social and Economic Roots of Newton's Principia. In Bukharin NI, et al. (eds.). *Science at the Crossroads: Papers from the Second International Congress of the History of Science and Technology, 1931*. Frank Cass & Co. (reprint). London 1971.

Kuhn T. *The Structure of Scientific Revolutions*. University of Chicago Press. Chicago 1970.

Kuhn T. Interview in Borradori G., Crocitto R. *The American Philosopher: Conversations with Quine, […], Kuhn*. University of Chicago Press. Chicago 1994.
Lakatos I. 'History of Science and Its Rational Reconstructions'. In Buck RC, Cohen RS (eds.). *Boston Studies in the Philosophy of Science* vol. VIII. Dordrecht 1971.
Latour B, Woolgar S. *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press. Princeton 1986.
Lyons AS, Petrucelli RJ. *Medicine – An Illustrated History*. Abradale Press. New York 1987.
Mannheim K. *Ideology and Utopia*. Routledge & Kegan Paul. London 1939.
Marshall B. *Helicobacter Pioneers. Firsthand Accounts from the Scientists Who Discovered Helicobacters, 1892-1982*. Blackwell Science. Victoria Australia 2002.
Marshall BJ, Armstrong JA, McGechie DB, Glancy RJ. Attempt to fulfill Koch's postulate for pyloric Campylobacter. *Medical Journal of Australia* 1985;142: 436-9.
Merton RK. *Social Theory and Social Structures*. The Free Press. New York 1968.
Moynihan R, Heath I, Henry D. Selling Sickness: the pharmaceutical industry and disease mongering. *British Medical Journal* 2002; 324: 886-91.
Moynihan R. The Making of a Disease: Female Sexual Dysfunction. *British Medical Journal* 2003; 326: 45-7.
Porter R. *The Greatest Benefit to Mankind – A Medical History of Humanity from Antiquity to the Present*. Harper Collins Publishers Ltd. London 1997.
Roberts RM. *Serendipity. Accidental Discoveries in Science*. John Wiley & Sons. New York 1989.
Scheler M. *On Feeling, Knowing, and Valuing: Selected Writings* (ed. H Bershady). Chicago University Press. Chicago 1992.
Shyrock RH. *The Development of Modern Medicine*. University of Wisconsin Press. London 1979.
Singer C, Underwood EA. *A Short History of Medicine*. Clarendon Press. Oxford 1962.
Toulmin S, Goodfield J. *The Fabric of the Heavens*. Penguin Books. Harmondsworth. London 1961.
Wright P, Treacher A. *The Problem of Medical Knowledge. Examining the Social Construction of Medicine*. Edinburgh University Press. Southhampton 1982.
Wulff H, Pedersen SA, Rosenberg R. *Philosophy of Medicine – an Introduction*. Blackwell Scientific Press. Cambridge 1990.
Young A. *The Harmony of Illusions: Inventing Post-Traumatic Stress Disorder*. Princeton University Press. Princeton 1997.
Ziman J. *Real Science. What it is, and what it means*. Cambridge University Press. Cambridge 2000.

3. What Is a Scientific Fact?

‘Facts are beyond dispute’, it is often said. But if this were the whole truth of the matter, it would be nonsensical to make the nowadays common contrast between scientific facts and other facts. Mostly in everyday life, we have to speak naively as if there were no epistemological problems. Otherwise we would not be able to act. In most everyday situations it is adequate to talk of facts as being simply obtaining states of affairs and as such being beyond dispute. If the facts in question are facts of nature, then they should obtain independently of all human consciousness. If they are social facts such as currencies, languages, and laws, they cannot possibly exist independently of all human consciousness, but they can nonetheless exist quite independently of the mental states of the scientists who, at a certain moment, are investigating them.

Some natural and some social facts change very rapidly, sometimes due to the intervention of researchers, but this does not mean that in such a flux there are no facts. It only means that either it is impossible to acquire any knowledge at all, or that the acquired knowledge is knowledge only about past facts, i.e., knowledge about the world as it was before the researchers came around.

The need to make a contrast between scientific facts and other facts has probably arisen because science has in many ways corrected common sense. Here are two of the most conspicuous examples. First, during non-cloudy days, the sun is perceived as moving up on one side on the heaven and going down in another, but science tells us that, in fact (!), these perceptions are illusory. Our own movement, which is due to the earth’s rotating, is in perception presented as a movement of the perceived object. Second, often when something burns, smoke and fire seem to *leave* the burning thing and nothing seems to enter it, but science tells us that the essence of burning is to be found in a process where oxygen *enters* the burning thing.

To say that something is a scientific fact is to bring in epistemology and show awareness of the epistemic brittleness of common sense. Advertisers often do it. Many products are sold (sometimes correctly and sometimes

wrongly) in the name of science. As consumers, we are often told that products have been scientifically tested, and that the promised effects have been scientifically proven to be there.

The fallibility problem stated is, however, no longer a problem only for common sense. It has shown itself to go deeper. What started as a process where science corrected ordinary perception and common sense has later turned into a process where new scientific results correct old scientific results. Today, part of yesterday's science is as wrong as the belief that the sun moves around the earth. This self-correcting feature of science became visible already in the second phase of the scientific revolution, when Kepler showed that the planets do not orbit in circles but in ellipses (see Chapter 4.6). It became obvious in the twentieth century when Newtonian physics was found doubly false – in one way by relativity theory and in another way by quantum mechanics. Similar remarks apply to the medical sciences; some cases will be mentioned later in this chapter. During the last fifty years, the fallibility of medical knowledge has become evident even to laymen. Many 'scientifically established facts' about various treatments have had to be publicly withdrawn and exchanged for new ones, which also have been discovered by science.

Since several decades, it is regarded as a commonplace in many academic circles that the fact that science-corrects-science proves it foolish and simple-minded to talk literally about facts. Why? Because, it is argued, if so many old 'scientific facts' have turned out to be no real facts, it seems ludicrous to think that the same fate will not sooner or later befall also our present-day scientific facts. However, even if this epistemological nihilism may continue to flourish in academic seminars, in serious life situations it always quickly disappears. When they become sick, even hard-headed social constructivists ask physicians for the facts about their diseases. In everyday life, no one is able to stick to an outright social constructivism and the concomitant epistemological nihilism. At the end of this chapter, we will show that social constructivism is as philosophically wrong and superfluous as it is impossible to live by. Here, we just want to point out that it is incoherent to claim that a fact (that science-corrects-science) proves that there are no facts.

That all empirical-scientific theories are fallible – even present and future ones – mean that we can never be absolutely sure that the contents

of our scientific theories completely correspond to reality. But this view implies a rejection neither of concepts such as ‘knowledge’ and ‘scientific knowledge’ nor of the concepts of ‘truth’ and ‘facts’. However, before we can discuss from a fallibilist perspective the question ‘what is a scientific fact?’, the true-falsity dimension of statements and theories has to be isolated from some other dimensions that pertain to human actions and speech acts. When discussing science, three pairs of opposites – to be elucidated in the sections below – have to be kept conceptually distinct:

- *true* versus *false* hypotheses (and *correct* versus *incorrect* data)
- *honest* versus *deceitful* presentations of data and hypotheses (whether true or false)
- *autonomous* versus *ideological* research processes (whether leading to truths or falsities).

An autonomous research process is a process where no authorities tell the researchers what the result has to look like. It does not mean that the researchers in question are free to research about anything they want; even if this is also sometimes a fruitful thing to allow.

Each of the three oppositions allow for degrees between them; even the first one, as we will explain below in Chapter 3.5. That there is no gap between honest and deceitful research on the one hand, or between autonomous and ideological research on the other, is more obvious. In both cases, the opposites fade into each other. Nonetheless we must often for pragmatic reasons use the dichotomous conceptualizations. This is no more odd than the fact that we need the concepts ‘red’, ‘orange’, and ‘yellow’ in spite of the fact that the spectrum is continuous. One intermediate between autonomous and ideological research looks as follows. A group of scientists receives money from a fund with very definite general views around the issue to be investigated. These scientists are completely autonomous in their ensuing research and in the publication of their results. However, they can be pretty sure that they will not in the future receive any money for other research projects if they don’t present results that are somewhat in conformance with the views of those who take the funding decisions.

It is important for everyone concerned with or involved in science to know that the following possibilities exist (and have been realized in the history of science):

(a) data and hypotheses that are the result of autonomous research processes need not be honestly presented (since researchers may lie and cheat for very private reasons)
(b) honestly presented results of autonomous research processes are not necessarily true (since research is fallible)
(c) results of ideological research processes are not necessarily false (since truth versus falsity brings in another dimension than that of being autonomous versus being put under ideological pressure)
(d) hypotheses that are false can nonetheless have some truthlikeness (since the scale between truth and falsity is continuous).

Science should aim at true results and autonomous research processes. Very exceptional situations disregarded, the results should be honestly presented, too. Sadly, this is not always the case.

3.1 Deceit and ideological pressure

Deception and fraud means deliberately presenting something as a fact despite actually knowing that it is not. To the uninitiated, it might appear superfluous to require that scientists should abstain from all forms of misconduct, but it is not. Misconduct surreptitiously occurs in science, and medical research is by no means an exception.

The borderline between deliberate deceiving and honest but bad science is not clear-cut. If a researcher (i) opts for a special statistical test only because he thinks it will give results that support his preferred hypothesis, then his choice is misconduct. If a researcher (ii) opts for the same test because he falsely thinks it is the only possible one, then it is only bad research. If a researcher (iii) opts for it because it is the only test he has efficiency in, but being well aware of other possibilities, then he is in the gray zone between cases of clear misconduct and cases of bad science.

In many kinds of experiments, it is hard to assess the outcome correctly since it is hard to judge if some data are mere ‘noise’ or not. Is a single extreme value, a so-called ‘outlier’, a true measurement value or an artifact

created by methodology and/or the technical apparatus used? Outliers do not only make statistical analyses difficult. Radiological pictures can contain several odd spots. How to answer the following question: if a radiologist comes to the considered conclusion that a certain spot is only a by-product of the radiological technique used, should he then be allowed to retouch the picture before he publishes it?

Next, a real case. In February 1922, the Canadian physician Frederick Banting (1891-1941) and a medical student, Charles Best (1899-1979), presented their first work on how to reduce sugar in the blood (and excretion in the urine). By adding doses of extract from degenerated pancreatic tissue to diabetic animals, they had managed to reduce the sugar in the blood of the latter. The article 'The Internal Secretion of the Pancreas' was published in the *Journal of Laboratory and Clinical Medicine* and the authors concluded:

> In the course of our experiments we have administrated over seventy-five doses of extract from degenerated pancreatic tissue to ten different diabetic animals. Since the extract has always produced a reduction of the percentage of sugar of the blood and the sugar excreted in the urine, we feel justified in stating that this extract contains the internal secretion of the pancreas.

Studies of the results and protocols of Banting and Best from 1921 have shown that the extracts did not always, as they maintained, produce a reduction of the blood sugar. Of the 75 injections they conducted on nine dogs, only 42 were favorable; 22 injections were unfavorable, and 11 were inconclusive.

Was this work scientific misconduct or merely bad science with an accidentally true result? In 1923, Banting and the head of the laboratory, John Macleod (1876-1935), received the Nobel Prize for the discovery of insulin.

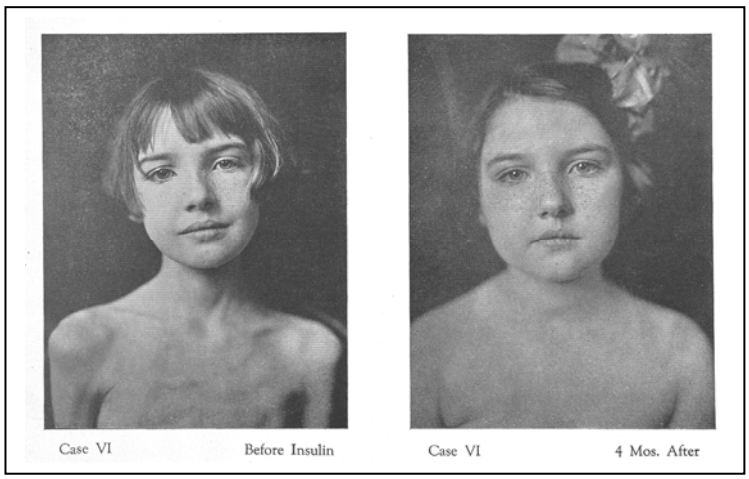


Figure 1: *Old pictures of a diabetes patient before and after four months of insulin treatment.*

In ‘ideological research processes’, as we are using this term, the researchers are put under strong social pressure in order to deliver the results the authorities in question wish for. But even in autonomous research processes there may be strong wishes among the researchers to obtain results that promote the short term interests of some external larger community; be the latter a nation, an ideological movement, a political party, or a company. This must be reckoned as being acceptable as long as ordinary scientific skepticism and testing procedures are adhered to. To wish that a certain view is true is as natural for researchers as it is for other people. There is no need to require that researchers should become emotionless truth-seeking machines, but it is necessary to require that researchers be honest, take criticism into account, and be more careful than other people in evaluating the results of their investigations. This being noted, how to evaluate the Blondlot case?

When the established French physicist René Blondlot (1849-1930) in 1903 presented his presumed findings of N-rays, many physicists immediately thought that the existence of such rays was an established scientific fact. This ‘discovery’ was for a time considered comparable to Wilhelm Röntgen’s (1845-1923) discovery of the X-rays some years earlier.

Figure 2: *Röntgen's first X-ray picture of a hand.*

Blondlot was honored by the French Academy before it was realized that his results could not be repeated. Many experiments were made, many papers concerning N-rays were published, and several French researchers became occupied with N-rays before suspicions about their existence arose. It was an American scientist, with no ties to the French research milieu, who demonstrated that Blondlot's N-rays had to be regarded as a result of hope and wishful thinking. The French had a national interest in exhibiting successful research – especially towards Germany. This is the time in-between the Franco-Prussian War of 1870-71 and the First World War. Germany could boast with Koch and Röntgen, and France had Pasteur and – they probably wished – also Blondlot. The national competition between the countries might have made French scientists too eager to support Blondlot's presumed discovery.

The Blondlot case took place a long time ago, and some may think that such things cannot happen anymore. But in 2002 it became obvious that a researcher at the Bell Laboratory in the US, Hendrik Schön, had falsely and deceitfully claimed to have discovered and isolated a biological single-

molecule transistor. This would have been a great breakthrough; allowing new prospects and expectations in the whole IT-sphere. During three years, Schön managed to write some 90 scientific papers, 15 of which were published in the very prestigious journals *Nature* and *Science*. In 2001, Schön received the 'Outstanding Young Investigator' award and $3000 in prize money from the Materials Research Society, and many thought that it was only a question of time before he would receive the Nobel Prize. However, some of his colleagues became suspicious when they noted that he used exactly the same graph to illustrate the outcome of two different experiments. The results were, so to speak, too perfect. Schön was then soon by his colleagues found guilty of having substituted data and presented data with unrealistic precision, and for not having considered that some results contradicted known physics; and a few months later he was fired. Even though the basic guilt and misconduct rests upon Schön, it is not reasonable to blame him only. Some of the reviewers of the famous journals – which are meant to implement the scientific principle of organized skepticism – seem in this case to be a bit blameworthy, too.

Papers of Schön in which misconduct was found
"Ambipolar pentacene…, *Science* (11 February 2000)
"A superconducting field-effect switch," *Science* (28 April 2000)
"An organic solid state injection laser," *Science* (28 july 2000)
"A light-emitting field-effect transistor," *Science* (3 November 2000)
"Superconductivity at 52 K in …C60" *Nature* (30 November 2000)
"Perylene: A promising…," *Appl. Phys. Lett.* (4 December 2000)
"Ambipolar organic devices…," *Syntetic Metals* (2001)
"Gate-induced superconductivity…," *Nature* (8 March 2001)
"Solution processed Cds…," *Thin Solid Films* (2 April 2001)
"High-temperature superconductivity in lattice-expanded C60" *Science* (28 September 2001)
"Ballistic hole transport in pentacene with a mean free path exceeding 30μm," *J. Appl. Phys.* (1 October 2001)
"Self-assembled monolayer organic…," *Nature* (18 October 2001)
"Superconductivity in CaCuO2…" *Nature* (22 November 2001)
"Field-effect modulation…" *Science* (7 December 2001)
"Fast organic electronic circuit based on ambipolar pentacene…" *Appl. Phys. Lett.* (10 December 2001)
"Nanoscale organic transistors…," *Appl. Phys. Lett.* (4 February 2002)

A large group of scientists within the research field in question were probably at the time suffering from wishful thinking. As in the case of Blondlot, several other research groups were preoccupied with trying to reproduce and improve upon the deceitful researcher's experiments. What once nationalism could do to stain research, other factors may do today. Both general ideological opinions and the stock market can be very sensitive to proclaimed scientific results. Often, when a pharmaceutical company can show that a randomized clinical trial is favorable to a new promising product, the stock market reacts promptly.

Trivially, science, scientists, and scientific results can only exist in particular societies; less trivially but quite naturally, often science, scientists, and scientific results are in a bad way influenced by the surrounding society. Science and political ideologies often overlap. One classic example is the race biological research in Nazi Germany. But this case is worth some more words than it normally is afforded. Race biology is not a specific Nazi phenomenon. It has a relatively long history prior to and even after its German phase 1933-1945, and it has been a university discipline in several countries. The truth is rather that the Nazi ideology took its departure from certain parts of the then current 'state of the art' in race biology. Modern genetics is, by the way, partly a descendant from race biology. However, during the Nazi-period race biology was used to legitimize racism, and was assumed to have proven that some ethnic minorities (in particular, the Jews and the Gypsies) were genetically defective, and that some sexual behaviors (e.g., homosexuality) and some political views (e.g., communism) were caused by genetic defects in the individuals in question.

The Nazis maintained that natural selection is a natural norm that the social and family policies of the Weimar Republic had made malfunction. The Nazis' sterilization program, euthanasia program, and, at the end, their systematic extermination of presumed genetically inferior groups were regarded as a way of helping nature along. Hitler was seen as the doctor of the German people ("Arzt der Deutsches Volkes") who, unsentimentally, tried to cut off diseased limbs from the body of the people.

Both before and during the Nazi period, several German physicists talked about a special 'German physics' or 'Aryan physics'. In the lead were two Nobel Prize winners, Philip Lenard (1862-1947) and Johannes

Stark (1874-1957); both were dispelled from their posts after the War. Despite being outstanding physicists, they regarded relativity theory and quantum mechanics as false theories. They looked upon Einstein's relativity theory as a 'Jewish bluff', and upon the German Werner Heisenberg and his group as 'cultural half-Jews'. Happily enough, this resistance was one reason why the Nazis did not succeed in producing an atomic bomb. As nations benefit more from intelligence services and spies who tell their presidents and prime ministers the truth, than from intelligence services and spies who tell the leaders what the latter want to hear, nations do probably in the long run benefit more from research communities and scientists that seek the truth than from research communities and scientists that merely reproduce existing theories, ideologies, and wishes. Here comes a case where this is obvious, the Lysenko affair.

Trofim Lysenko (1898–1976) was a Russian agronomist suspicious of Darwinism. In particular, he did not believe in Mendelian inheritance and the complementing chromosome theory of heredity put forward by T. H. Morgan (1866-1945). He thought in a Lamarckian manner that grains and plants could acquire new properties by environmental influences, e.g., by subjecting grain to extreme temperatures or injections, and that these changes then could become hereditary. When Lysenko entered the scene at the end of the twenties, the political-economic situation was as follows.

The agriculture of the Soviet Union was since long in a crisis. Immediately after the revolution this was due to extremely hard crop taxation-deliveries, and at the end of the twenties the crisis was due to the collectivization of the farms. This process contained the deportation and eventual deaths in camps of hundreds of thousands of formerly relatively well off peasants; and in the years 1932-33 a famine in Ukraine killed millions. Most agronomists were educated before the revolution and felt no enthusiasm for the new regime and its measures; in particular for the collectivization policies. Like most theoretical biologists of this day, the Soviet ones were not interested in agriculture but in the new genetics that was emerging out of Morgan's studies. Since this research only much later had implications for agriculture, it was for several decades easy to castigate theoretical biologists for spending their time with fruit flies while agriculture was in need of fast radical improvements.

Lysenko claimed to have invented a new technique of vernalization (subjecting seeds to low temperatures in order later to hasten plant growth), and he promised to triple or quadruple yields using his technique. But he also made claims that his views were more in conformance with Marxism than those of his opponents. He successfully attracted official support not only from the Communist Party and Stalin, but also for a time from the internationally well respected Soviet botanist and geneticist Nikolai Vavilov (1887-1943), who even wanted Lysenko to explain his ideas on international biological conferences.

Lysenko was later put in charge of the Academy of Agricultural Sciences of the Soviet Union, and when Stalinism reached its peak in the mid and late 1930s, Lysenko was made responsible for ending the propagation of ‘harmful’ ideas among Soviet biologists. Lysenko is in this position responsible for the expulsion, imprisonment, and death of hundreds of scientists – and the demise of genetics. Among the scientists killed were Vavilov. The period between the mid-thirties and mid-sixties is known as Lysenkoism. Only in 1965 was Lysenko removed from the leadership of the Soviet Institute of Genetics.

Due to the Lysenkoist part of Stalinism, modern biological and genetic research was in the Soviet Union neglected for almost 30 years. Probably, Lysenko believed in the basics of what he preached. He later admitted that his ‘vernalizations’ had not been successful, but never that Morgan was right. Lysenkoism was a delicate interplay between researchers with wrong ideas and politicians that were too sure to have truth on their side to admit real criticism. Even in our next ideological case, the central figure might have been honest.

Whatever the term race has been used to refer to, no one has denied that males and females from different so-called races can produce fertile and healthy offspring. One may today think that this fact should have made it impossible to turn ‘race’ into a more basic taxonomic concept than ‘subspecies’ in zoology (and ‘variety’ in botany). When met in nature, subspecies of animals are due to accidental geographical isolation of various populations; when met in domesticated animals, it is due to planned inbreeding. However, in the nineteenth century ‘race’ had become a truly central concept in anthropology, where it relied on various

speculative assumptions about the origin of mankind and what a future interracial development might lead to.

Samuel G. Morton (1799-1851) was an American doctor who studied the size of human skulls from different ‘races’. He was living in the time of the American slavery, and he was confident that whites are naturally superior. Since he assumed that brain size was directly related to intelligence, he tried to rank the intelligence of ‘races’ by measuring the brain cavities of human skulls from different ‘races’. To this purpose, he collected hundreds of human skulls, whose volumes he measured by filling them with lead pellets that he later dumped into a glass measuring cup. He ended up with tables containing results from more than 400 skull measurements. According to his data, Europeans (with the English at the top) have the largest skulls. As number two comes the Chinese, as three the Southeast Asians and Polynesians, as four the American Indians, and on the fifth and last place comes Africans and Australian aborigines.

In 1977 the biologist Stephen Jay Gould re-analyzed Morton’s data. He found that Morton had been overlooking inconvenient exceptions and that brain size correlates more closely with body size than with any ‘race’. The larger the body is, the more voluminous the brain, regardless of race. Gould found that Morton had got his results by studying smaller individuals (in particular women) in the races he presumed was inferior. Once Gould eliminated body size as a factor, he found that all races have roughly the same brain size. Due to the openness of Morton’s data, Gould does not suspect conscious fraud on Morton’s part.

Even the prominent French surgeon and neuroanatomist Paul Broca (1824-1880), whose work is today remembered mainly through the names ‘Broca’s area’ and ‘Broca’s aphasia’, made craniometric studies. He came to the conclusion that, on average, the brain is larger among middle-aged people than among elderly, larger among males than females, larger among prominent men than average Joes, and, finally, larger among ‘superior’ races than ‘inferior’ races. Again, it was falsely taken for granted that there is a simple correlation between brain size and intelligence. Today it may look as a laughable assumption, since it is so obvious that computer capacity does not correlate with computer size, but this is an anachronistic laughter. Human brains are larger than the brains of other animals.

Later, a related ideologically loaded question came to the research fore: is intelligence the result of heredity or is it due to conditions in the social milieu during childhood? Here there is a famous case of real deceit. Be the hypothesis false or not.

The fraudulent man, Cyril Burt (1883-1971), was in 1946 knighted for his scientific work, which had influenced the organization of both schools and prisons in England. One of Burt's most famous studies (from the mid of 1940s to 1960s) is concerned with identical (monozygotic) twins. Initially he reported to have studied 21 pairs of twins reared apart, and found a correlation coefficient for IQ's of 0.771. Later on he repeated the findings adding several twins to the study – the last report refers to 53 pairs of twins, resulting two times in the same correlation coefficient of 0.771. He concluded that genetic factors are more important than environmental factors in determining intelligence.

In the 1970s, after the death of Burt, some other researchers noted that the correlation coefficients of the three different samples of twins were almost the same down to the third decimal, and they found such a match unbelievable. Investigations of Burt's scientific work showed that, probably, he had invented many data and even referred to fictional co-workers. Five years after Cyril Burt's death, he was in a newspaper officially accused of having published a fraudulent series of separated twin studies, and soon also The British Psychological Society found him guilty of fraud. The organized skepticism of the scientific community functioned, but this time it functioned too slowly.

3.2 Perceptual structuring

Whether some empirical data are honestly presented or not can for many reasons be hard to determine. But one particular difficulty stems from the fact that our perceptions are complex entities structured by many factors. What at first looks like a case of obvious fraud may turn out to rest on an uncommon kind of perceptual structuring. Let us explain with the help of the philosopher of science N. R. Hanson (1924-1967). He claims (and we agree) that to invent a new hypothesis is often the same as to become able to perceive something in a new way. To create a new idea can be like observing numerous strokes, acquiring a feeling of a form, and eventually perceiving all the original strokes as constituting a definite pattern. Look at

Figure 3 below. In science, empirical data correspond to the strokes, and the hypothesis to the whole picture or pattern of such pictures.

Figure 3: *Do you see the horse and the cowboy, or do you see only black and white areas?*

According to Hanson, we structure our observations differently depending on what theories and concepts we have acquired; in order to construe new theories and concepts we have sometimes also to have new kinds of perceptions. Put bluntly, a layman does simply not see exactly the same things in a laboratory as a researcher does. One might say that to the layman the laboratory appears as a foreign language that he does not understand. He sees only boxes, buttons and instrument pointers (compare: non-understandable words), where the researcher sees functional wholes and experiment results (compare: meaningful words). To perceive an X-ray tube *as an X-ray tube* is very different from seeing the same thing as merely a glass-and-metal-object. To the uninitiated an X-ray of the stomach appears as composed of a number of more or less dark and light areas without any coherence, while a radiologist directly can see, say, the diagnosis ileus (absent intestinal passage). Similarly, psychologists, psychiatrists, and general practitioners perceive certain human behaviors differently from laymen. Let us repeat the glasses metaphor. During their

education, researchers usually acquire a pair of scientific glasses through which they then observe the world; these glasses are difficult to remove.

Figure 4: *Do you see the antelope or the pelican or maybe both?*

Look at Figure 4 above. Probably, you can easily switch between seeing an antelope and a pelican. Before experimental psychology publicized that kind of figure, most people probably thought that a picture or drawing could represent one thing only. If one person saw an antelope (had antelope-glasses) and another saw a pelican (had pelican-glasses) they might have quarreled about what was really pictured. The solution here of course is to realize that pictures can be as equivocal as words; ‘blade’, for instance, can mean both leaf and part of a knife. But this solution cannot be applied to mind-independent reality. No real animal can simultaneously be two kinds of animals. Nonetheless, one person may perceive one kind of animal where another person perceives another kind. In order to discuss such a situation rationally, both persons have to become aware of the fact their perceptions are not infallible and can be, partly or wholly, illusory. In the best of discussion cases, both persons learn to switch and then start to discuss what might be the true view or the veridical perception.

In the history of science, there are many cases where people who should have questioned their perceptions have not been able to do so. It is important to realize that there can be perceptual disagreement in science, since this disagreement is not only of pure empirical nature and, therefore, cannot be solved only by experimental methods. Theoretical issues are also important, and sociological factors may influence the final decision. One cannot take it for granted that one’s opponents are cheating, or are stupid, only because one cannot at once perceive what they claim to perceive.

It is important to remember this point for the later discussion of placebo and nocebo phenomena (Chapter 7). Such phenomena might represent a

pelican perspective in a world where clinicians, until some decades ago, have only been interested in observing antelopes.

Back to Hanson; here comes a quotation:

> Consider two microbiologists. They look at a prepared slide; when asked what they see, they give different answers. One sees in the cell before him a cluster of foreign matter: it is an artifact, a coagulum resulting from inadequate staining techniques. This clot has no more to do with the cell, *in vivo*, than the scars left on it by the archeologists spade have to do with the original shape of some Grecian urn. The other biologist identifies the clot as a cell organ, a 'Golgi body'. As for techniques, he argues: 'The standard way of detecting a cell organ is by fixing and staining. Why single out this technique as producing artifacts, while others disclose genuine organs?'
>
> The controversy continues. It involves the whole theory of microscopical technique; nor is it an obviously experimental issue. Yet it affects what scientists say they see. Perhaps there is a sense in which two such observers do not see the same thing, do not begin from the same data, though their eyesight is normal and they are visually aware of the same object. (Hanson, p. 4)

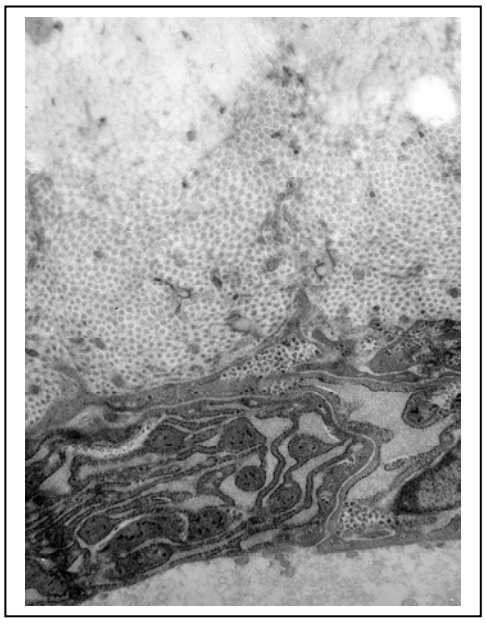


Figure 5: *Electron micrograph of a Golgi apparatus in an osteoblast.*

The point made by Hanson (and several other philosophers such as Fleck and Kuhn) is that the answers to scientific questions are partly preformed by the paradigm in which the questions are asked. A typical example of this assumption is the interpretation of the function of the vein valves made by William Harvey when he studied in Padua in 1601. His teacher, Fabricius (1537-1619, who had discovered the vein valves, understood the function of the valves as to moderate the speed of the bloodstream, which according to Galen moved in the direction from the heart to the organs and extremities (centrifugal). This interpretation was part of Harvey's departure when he eventually developed his theory of the movement of the heart and blood. Was it possible to interpret the function of these valves in another way, according to which the bloodstream has the opposite (centripetal) direction? In a world of antelopes, Harvey managed to see a pelican.

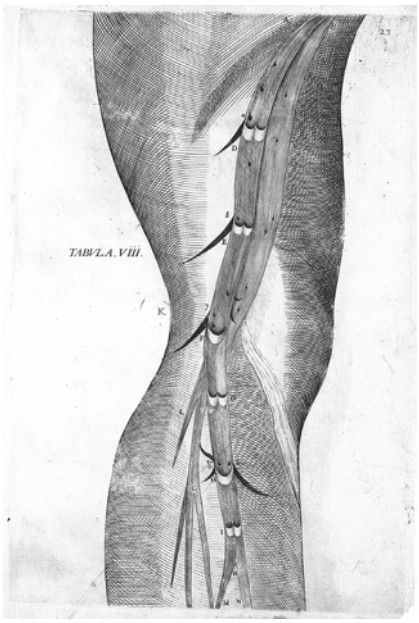


Figure 6:
Harvey interpreted the function of the vein-valves in the opposite way compared to Fabricius, who is acknowledged for the discovery of the vein-valves.

But even researchers and clinicians that belong to the same paradigm may have problems when they compare their observations. Even persons who embrace the same paradigm apply the same methodology, and study the same kind of phenomena may differ in what they claim to have observed. Think of different physicians who palpate the same patient's stomach, different physiologists who scrutinize a certain EKG, and different radiologists who observe the same X-ray-picture. Very often they report somewhat different results. In such cases there is *inter-observer variance*. Sometimes these differences are quite significant (compare a

novice to an expert) and sometimes less (compare two experts). The average inter-observer variance is estimated to be about fifteen percent. Also, there is even variance over time for one and the same observer; this phenomenon is called *intra-observer variance*.

Inter-observer variance and intra-observer variance have another character than the difference that occurs between observers belonging to different paradigms, but all these differences show that empirical data are never completely unproblematic. Nonetheless, perception is in science as in everyday life the primordial connection with the world. The fact that empirical data contains interpretative problems does not imply that all interpretations are on a par.

3.3 Collectively theory-laden observations

We will next present (a) the emergence of the Wasserman test as described by the Polish microbiologist and philosopher of science Ludwik Fleck (1896-1961) and (b) the emergence of the antiseptic procedures that are due to the Austrian-Hungarian Ignaz Semmelweis (1818-1865). The first case is meant to illustrate how many theoretical presuppositions there are behind a modern diagnostic test, and the second case is meant to illustrate that the phenomenon of perceptual structuring does not exclude the existence of empirical data that both sides in a scientific conflict can agree upon. Such neutral data, however, are seldom in themselves able to settle the conflict.

Fleck was one of the first who claimed that the concept ‘scientific fact’ is not as simple as it may seem. There is a social constructive moment to it that has been neglected. In 1935 he published the book *The Genesis and Development of a Scientific Fact* in which he analyzed and discussed how to understand what a scientific fact is. Fleck influenced Thomas Kuhn but, one might say, he became himself a famous figure only after, and partly due to the fact that, Kuhn’s views had already become widespread. Fleck uses a medical-historical case in order to make his points, the development of the first diagnostic test for detecting syphilis, the so-called Wassermann test.

This test was envisaged long before the immune system was discovered, but after vaccination, spontaneous immunological reactions, and the

existence of antibodies had become well known. The point of the Wassermann test is that it can detect the presence of the bacterium that is the central cause of syphilis, the spiral-shaped microorganism *Treponema pallidium*. The test represents one of the earliest applications of an immunological reaction that is termed ‘complement fixation’. In the test, the patient’s blood serum is first heated (in order to destroy its so-called ‘complement’), and then a known amount of a complement and antigen (the bacterial phospholipid) are added to the patient’s serum. The natural action of the new complement is to bind to the red blood cells and cause them to burst (undergo lysis). Visually, this is evident as a clearing of the red-colored suspension. However, if the added antigen has bound to antibodies that are present in the suspension, the complement becomes associated with the antigen-antibody complex. In technical terms, the complement is described as being ‘fixed’. Thus, if lysis of the red blood cells does not occur, then antibodies to *Treponema pallidum* are present in the patient’s serum, and a positive diagnosis for syphilis is very probable.

According to Fleck, it was Wassermann’s interest in blood that enabled him and his co-workers to develop a reliable test in practice. Fleck stresses that Wassermann’s fortunate interest in blood might be explained by the history of ideas. According to Galen, it is the composition of the blood and body fluids that determine whether one is suffering from a disease or is a healthy person. Fleck thinks that if Wassermann had not been aware of the old Galenic paradigm and its focus on the blood, he had probably never been able to develop a reliable Wassermann test.

At first, Wassermann thought that he was isolating specific antigens. This was not the case, but nevertheless he obtained (according to Fleck’s presentation) the remarkable results below:

		The true values (according to Wassermann):	
		positive	negative
Wassermann test reactions:	positive	64	5
	negative	0	21

Using the modern notions of sensitivity and specificity, we obtain for the Wassermann test a sensitivity of 100% (all sick come out as sick) and specificity of 81% (81% of the healthy come out as healthy):

- Sensitivity = true positives / (true positives + false negatives) = 64 / (64 + 0) = 100%
- Specificity = true negatives / (true negatives + false positives) = 21 / (21 + 5) = 81%

Among the 21 negative controls that Wassermann used, 7 were taken from brain tissue; at this time it was falsely supposed that it is not possible for the brain to be infected by syphilis. The contemporary concept of syphilis did not include the third stage, generalized syphilis. Against this backdrop, it is interesting to note that if these 7 controls were positive, in the last row of the table above we would have 14 true negatives and 7 false negatives instead of 21 true negatives; and the following – not so good but good – sensitivity and specificity:

- Sensitivity = true positives / (true positives + false negatives) = 64 / (64 + 7) = 90%
- Specificity = true negatives / (true negatives + false positives) = 14 / (14 + 5) = 74%

It was, however, impossible to reproduce these optimistic figures, and accordingly focus was redirected towards the isolations of antibodies. But neither did the attempt to conduct immunization by means of dead syphilis bacteria (antigens) give rise to a specific antibody reaction. Alcohol muscle extracts from a bull's heart also happen to imitate the specific properties of the 'antigen'. A specific antigen-antibody reaction is not the underlying mechanism of the Wassermann-reaction. It is as a diagnostic of syphilis actually a so-called complement-fixation test, and it is possible to use both blood serum and cerebrospinal fluid. The active substance is a lipid called cardiolipin, which has nothing to do with syphilis. However, this substance reacts with so-called *réagines*, which commonly occur in the blood of syphilis patients, but also in relation to other diseases such as malaria and leprosy. Nevertheless the Wassermann reaction proved a reliable test when

it was launched in 1906, and was in use (slightly modified) almost until the end of the century.

As previously stated, results that are too perfect give rise to suspicions that the research project has not been conducted as expected – in the worse case you might suspect fraud or misconduct. According to Fleck, the result of the isolation of the supposed antigens was so perfect that one can suspect that data had been 'cooked' or arranged. But, still according to Fleck, the point is that the scientific community (or thought collective) stylizes the observations. The observations are perceptually cultivated (probably both consciously and unconsciously) so they are brought in accordance with the current thought style (paradigm). We ought to perceive Wassermann's expectations against the backdrops of his experiences in immune diagnostic.

To Fleck observation is a social activity that is always theory laden; otherwise we would not be able to observe anything coherent at all. According to Fleck, we are prepared to focus and to constrain our perception; thought style determines what we include and exclude in our observation. He went so far as to maintain that it is only possible to talk about real observations when the answer is already prepared and imbedded in the question. Even scientific facts are thus according to Fleck (as well as Kuhn) necessarily deeply rooted in the worldview of the thought collective.

In order to illustrate further how a paradigm or thought collective comes to the conclusion that a certain phenomenon should be regarded as a scientific fact, we will present the famous example of Ignaz Semmelweis, the problem of puerperal fever, and the reactions of his colleagues.

In 1846 at the Allgemeine Krankenhaus in Vienna, where Semmelweis worked, there were two clinics. In the first 459 women in labor out of 4,010 died of puerperal fever, i.e., 11.4%. During the same period in the second clinic – where midwives were responsible for the care – only 105 out of 3,754 (2.7%) women died giving birth. A change in the mortality rate in the first clinic was first observed in 1840 when this clinic was made available for medical students. From 1841 to 1846 the mortality rates in the two clinics were as follow:

Mortality Rate

	First division (physicians/students)			Second division (midwives)		
	Delivering	Died	%	Delivering	Died	%
1841	3,036	237	7.8	2,442	86	3.5
1842	3,287	518	15.8	2,659	202	7.5
1843	3,060	274	10.0	2,739	164	5.4
1844	3,157	260	8.2	2,956	68	3.3
1845	3,492	241	6.9	3,241	66	2.0
1846	4,010	459	11.4	3,754	105	2.8
Totally	20,042	1,989	9.92	17,791	691	3.8

Semmelweis had noted that the pathological findings (inflammation of the peritoneum and viscera) in the cadavers of the mothers were similar. The crucial idea came to Semmelweis's mind when he read the post mortem report of his colleague, Kolletschka, who was a professor of forensic medicine. The latter had cut his finger during an autopsy, the wound became infected, and Kolletschka died of sepsis. In the report Semmelweis read that the findings were similar to those found in women that had died in puerperal fever: abscesses, peritonitis, phlebitis, etc. If Kolletschka's wound was affected by living organic material (cadaveric matters) that caused his death, then it was possible, so Semmelweis thought, that cadaveric matters also affected the women (and many of the newborn babies) who died in puerperal fever. The cadaverous particles might have come from the hands of the physicians and the medical students (belonging to the first division). They started the day by conducting autopsies, after which they entered the obstetric division to examine the vulnerable and torn genitalia of the women in labor.

Semmelweis proposed that all physicians and medical students should wash their hands in a chlorine solution (chlorina liquida) before examining patients. His chief, Professor Klein, was not convinced of the importance of the procedure, but eventually he allowed it. It was Klein who had introduced the practice of early autopsy every morning, and ordered that medical students should attend the first division in order to train their skills. Interestingly, Klein's predecessor had recommended that the

medical students should practice and train on a leather model, not on actual women.

At the end of May 1847 the proposed hand washing procedure was introduced; and then the mortality rate decreased as follows:

Mortality rate at First division during 1847

	Delivery	Died	%
January	311	10	3.21
February	912	6	1.92
Mars	305	11	3.6
April	312	57	18.3
May	294	36	12.2
Totally	2,134	120	5.62

(Chlorina liquida procedure introduced)

	Delivery	Died	%
June	268	6	2.38
July	250	3	1.2
August	264	5	1.89
September	262	12	5.23
October	278	11	3.95
November	246	11	4.47
December	273	8	2.93
Totally	1,841	56	3.04

The corresponding mortality rate at the Second division, where only midwives worked, was 0.09%. In 1848, the statistics for the First division looks as follows:

Mortality rate at First division during 1848

	Delivery	Died	%
January	283	10	3.53
February	291	2	0.68
Mars	276	0	0.0
April	305	2	0.65
May	313	3	0.99
June	264	3	1.13
July	269	1	0.37
August	261	0	0.0
September	312	3	0.96
October	299	7	2.34
November	310	9	2.9
December	373	5	1.34
Totally	3,556	45	1.27

In the Second division this year, the average was 1.33%.

In March 1849, Semmelweis' colleague, but opponent, Carl Braun took over the responsibility, and then the mortality rate increased intermittently, and in 1854 it was 9.1%, i.e., almost the same as when Semmelweis started to intervene.

Viewed wearing the current paradigmatic glasses, the effect of Semmelweis' proposal seems to be so self-evident that even without any modern statistical analysis it should have been immediately accepted. But this was not the case. To the contrary, Semmelweis' data was received with skepticism and rejected by several doctors, especially in Vienna. The fact that the responsible professor Klein and Semmelweis' colleagues did not accept his analysis may be due to their personal involvement. Accepting the results would imply that the physicians (including Semmelweis himself) and the medical students had caused the deaths of thousands of patients (mothers and their babies). Indirectly, it could be seen as an accusation of guilt, and this might be hard to accept. After some time Semmelweis left Vienna and went back to Budapest. Initially, Semmelweis presented his results only orally. It was not until 1861 that he presented his results in a book, *Die Aetiologie, der Begriff und die Prophylaxis des Kindbettfiebers*. At first, even the international reaction was critical and reluctant. Four years later, i.e., before his results eventually were accepted, he died suffering the same cruel fate as Kolletschka.

The main reason for the resistance towards Semmelweis' results was the fact that he used pre-scientific concepts (such as 'cadaveric matters' and 'rotten particles from living human beings') and conformed to the less popular contact theory. The French Academy, the Royal Society in London, the prominent German physician Rudolf Virchow, and several other leading medical circles did not find Semmelweis' hypothesis at all convincing. Semmelweis' hypothesis was rejected because his opponents comprehended fever diseases, including puerperal fever, as caused, e.g., by miasma. According to the miasma theory, it is changes in the air that make people ill, not direct contact. The consequence of this understanding was that the hospital ward was air-conditioned or the patients were moved from one ward to another.

The resistance was also motivated in some other ways. Infectious diseases like tuberculosis were for a long time understood as the result of

heredity, and even arguments against possible mono-causality was provided.

Almost twenty years passed before the effect of Semmelweis' preventive intervention was theoretically explained and practically accepted. The final acceptance was due to Pasteur's and Koch's development of microbiology and the microbiological paradigm. Under the influence of this paradigm it was possible to interpret, understand and explain Semmelweis' apparently pre-scientific concept 'cadaveric matters' and 'rotten particles". The empirical anomalies, which for theoretical reasons were not taken seriously in the old paradigm, were in the new one explained in terms of microorganisms (bacteria). The Semmelweis story shows how the theoretical framework can influence what becomes accepted as a scientific fact. Note though that this kind of theory-ladenness is quite consistent with the fact that both sides accept the numbers in the mortality rates.

The Semmelweis case shows the rigidity, resistance, and conservative function of a paradigm. This is part of the nature of a paradigm, and may not always be a bad thing. Empirical data from one single trial, how significantly they might appear, is not sufficient to cause a scientific revolution.

3.4 Positivism: classical and modern

As made clear in the preceding sections, observation is a theory-dependent activity. Almost all contemporary philosophers of science seem at least to agree on this. But who is then the enemy? Answer: positivism. In the following pages we shall briefly present some aspects of both classical and modern versions of positivism, in particular logical positivism.

Francis Bacon (1561-1626) is perhaps the first positivist. Apart from this label, he has been called 'the philosopher of the industrial revolution' since he not only (like several others) polemicized against the medieval belief in authorities and stressed the importance of *systematic* empirical investigations, he also propagated – and even correctly prophesized – that empirical science can lead to undreamt of discoveries and inventions. His most influential work is called *Novum Organum Scientiarum*, i.e., 'The New Instrument of the Sciences', the systematic empirical method.

Central to all forms of positivism is the assumption that it is possible by means of observation to obtain certain knowledge. This assumption means

that somewhere and in some way there are empirical data that are indisputable and not theory impregnated. It does not mean that everyday observations are certain and unproblematic. To the contrary, Bacon stresses that everyday observation are often misleading. He warns against what he calls ‘idols’ or illusions. Idols distort one’s observations; people with idols cannot see the world clearly. According to Bacon, four types of idols have to be taken into consideration:

1) The idols of the tribe (*idola tribus*), which are the result of our need for wishful thinking; as human beings we are by our nature inclined to see what we want to see independently of whether it is actually there or not. Among other things, these idols make us prone to provide nature with the properties of human beings and look for purposes in nature although there are none.
2) The idols of the den or the cave (*idola specus*), which are each individual person’s specific prejudices; these may be both congenital and acquired.
3) The idols of the market place (*idola fori*), which are the prevailing delusions that language brings with it.
4) The idols of the theater (*idola theatri*), which are the false doctrines of the old philosophical systems. It was especially these that Bacon wanted the scientists to become aware of.

Bacon’s positivist point is that one should try to get rid of these idols, since if one succeeds one can see the world as it really is. If one is free from idols one will be able to make theory-independent observations, and it is on this kind of observations that science should be based.

Another classical positivist is the French philosopher and sociologist Auguste Comte (1798-1857). He introduced the very label ‘positivism’, and his most famous work is called *Course in the philosophy of positivism*. Comte presents an overarching theory about the emergence of scientific thinking and how it ought to continue to develop; such theories about historical developments are typical of this time. He divides human thinking into three stages: the theological stage, the metaphysical stage, and (last and highest) the positive stage. In the first stage unobservable gods are posited; in the second unobservable material entities and forces are posited.

Only in the newly begun third stage thought rests content with the observable.

Comte loads the terms ‘positive’ and ‘positivism’ with several meanings. The positive is what is real in opposition to what is fictive or imaginative; and research ought of course to concern itself with what is real. Positive is also everything that is useful in opposition to what is harmful. According to Comte, the fundamental goal of thinking should be to improve our individual and collective life-conditions; thinking is there not only for satisfying our curiosity. Third, what is positive is certain in contrast to the uncertainty of, say, theological speculations about angels and metaphysical speculations about unobservable forces. Furthermore, positivism should be understood as the precise in contrast to the vague. Finally, positive is of course also understood as the normatively positive in opposition to the normatively negative.

Positivism and its strong stress on science as being also a social liberator should be understood against the contemporary social setting of the universities. Not only in Bacons’ time, but even in Comte’s and for a long time after, the church had quite an influence over the universities. Even if the universities were not formally part of the church, many clergy positions were automatically also administrative university positions. From a socio-political point of view, classical positivism functioned as a research ideology that tried to make research independent of theology. Modern positivism had the same function, but it wanted to liberate science also from philosophy and politics. Apart from this, it wanted to make the study of societies and cultural artifacts more scientific.

The main founder of modern positivism is the Austrian scientist and philosopher Ernst Mach (1838-1916). He argues that everything, including common things like tables and chairs, stones and plants, are merely complicated complexes of sensations (this ontological view is often called phenomenalism). Accordingly, since there are no material things, there is no reason to discuss whether in addition to macroscopic material things there are also microscopic material things such as atoms and electrons. At the turn of the nineteenth century, a controversy raged within physics as to whether or not atoms and electrons were fictional objects.

According to Mach, there is nothing else for science to study than regularities among sensations. This goes for the natural sciences, the

medical sciences, the social sciences, and the humanities. This means, first, that there are no ontological gaps between different research disciplines, and this is taken to imply, second, that there are no basic methodological gaps either. According to logical positivists, all branches of science ought to study only what is observable strictly in accordance with the hypothetico-deductive method. The logical structure of this method will be described in Chapter 4.4.

Mach gave rise to the idea of a ‘unified science’, an idea that was developed particularly within the so-called logical positivism (sometimes called logical empiricism). Examples of well-known logical positivists are Moritz Schlick (1882-1936), Rudolf Carnap (1891-1970), and Otto Neurath (1882-1945). Independently of this group, who worked in Vienna and is often referred to as the Vienna Circle, other positivistically oriented schools at other Western places emerged simultaneously. For instance, in Berlin there was a group around Hans Reichenbach (1891-1953) and Carl G. Hempel (1905-1997), and in the United States there were related movements such as instrumentalism, operationalism and pragmatism. Australia and Scandinavia were heavily influenced. Common to all kinds of positivism is a criticism of metaphysics. But logical positivists gave this criticism a special twist. They argued that traditional philosophical problems and speculations are not only sterile and unsolvable but literally semantically meaningless.

Since several logical positivists fled from central Europe to the US because of Nazism, after the Second World War there was quite an interaction between all these ‘isms’. Modern positivism, instrumentalism, operationalism, and pragmatism can be considered parts of a much larger and quite heterogeneous philosophical movement called analytic philosophy. Parts of this movement, let it be said, have delved deep into metaphysics.

The logical positivists wanted to erase borders and gaps between different scientific disciplines but create a gulf between scientific work and other intellectual activities. This combined effort to unify all sciences but create a gap between these sciences and everything else had one very conspicuous result, the positivists’ *principle of verification*. Taking the purely formal sciences of logic and mathematics aside, the principle says, in a vague formulation:

- Theories and hypotheses that by means of empirical data are able to be deemed either true or probably true (verifiable) are meaningful and scientific; those who are non-verifiable, e.g., ‘there are electrons’ and ‘god exists’, are meaningless and unscientific.

The verification principle was meant to function as a robust criterion for what to allow inside universities and what to block at the entrances. What was verifiable should be allowed, what was not should be forbidden. The problem was that it was hard to find a concise formulation that even all of the positivists could agree upon.

Another characteristic trait of much positivism is its denial of the classical conception of causation, according to which a cause with necessity brings forth an effect. To think in terms of such a natural necessity was considered metaphysical speculations and as such being nonsense in the strongest possible way, i.e., being semantically meaningless. Here the logical positivists referred to David Hume (1711-1776), another forefather of modern positivism, and his reduction of causality to correlation. According to the positivists, the search for laws of nature and causal relations should be replaced by a search only for correlations between empirical data. For instance, they found it nonsensical to talk about a gravitational *force* between the moon and the sea that causes the tide to ebb and flow; there are only correlations between the positions of the moon and the tide phenomena. This analysis of causality fits well with the positivist dream of a unified science. If all sciences whatsoever studies only correlations, then there is no need to discuss whether different scientific disciplines have to posit different kinds of causes and mechanisms.

Words can take on many meanings, and so has the word ‘positivism’. In various quarrels in various disciplines in the second half of the twentieth century, the term has sometimes been used to stress things that are not especially intimately connected with the kernel of positivism that we have stressed, but which were part of the logical positivists’ interest in the formal-logical structure of theories and explanations. Therefore, one can find (especially during the seventies) papers that classify *all* search for axiomatizations and/or quantifications as being positivist research. This is

bad terminology. Axiomatization and quantification can go well together with non-positivist views in the philosophy of science.

Even though modern positivists have not talked in terms of Baconian idols, they have talked in a similar way. They have said that scientists in their research (not in their private lives) should try to be *disinterested*, i.e., try to be free from authorities, emotions, and all non-scientific values such as political, ethnical, and religious priorities. Only with such an attitude, they said, is it possible truly to see empirical data as they are in themselves; only the disinterested researcher can make objective observations and see the infallible core of knowledge.

Let us make the central contrast between our views (to be spelled out in the next section) and positivism clear: positivism asks the researcher to *take away* idols in order to receive reliable empirical data, but we would like researchers to *create* whatever means they can in order to make (inevitably fallible) observations of the world.

The positivist ideal of disinterestedness in research has been important from a socio-political point of view. Probably, it helped scientific truth-seeking to liberate itself more definitely from religion and politics. Nonetheless, it is impossible as a general ideal. Today, much research merge pure and applied science so intimately that the researchers have to be deeply interested in the fate of the technological and inventive side of their research project in order to function well as truth-seekers on the project's discovering side. Also, and especially in medical research, ethics has become such an important issue that researchers are simply required to have some ethical interest. The end of the Second World War made both these kinds of interestedness obvious. Many famous theoretical physicists felt a need to help in the construction of the atom bomb, and the medical case of the post-war Nuremberg trial's showed the need to bring science and ethics into contact with each other (see Chapter 10).

3.5 The fallibilistic revolution

At the end of the nineteenth century, it was still possible to regard the post-Newtonian theories of physics as merely adding new bits of knowledge to Newton's miraculous theory. James Clerk Maxwell (1831-1879) had managed to combine the laws of electricity and magnetism into a single theory of electromagnetic waves, and such waves are of another kind than

the material particles that Newtonian mechanics deal with. Einstein changed it all. His special theory of relativity (1905) implies, as it is often and correctly said, that Newton's theory can only give approximately correct predictions for particles with velocities much smaller than that of light. But this is not the whole truth. Strictly speaking, the theoretical predictions from Newton's theory and Einstein's theory *never* give for any velocities exactly the same values; although the larger the velocity is, the larger the difference becomes. In other words, the theories logically contradict each other and, therefore, both cannot be strictly true. The contradiction arises because the theories contain different formulas for how to transform the quantitative values (of magnitudes such as mass and velocity) found in one inertial reference system into the values they obtain in another such reference system. Newtonian mechanics has so-called Galilei transformations, and relativity theory has Lorentz transformations, and these transformations give different values for all velocities, not only for high velocities.

Since Newton's theory, which once stunned the whole educated world, had turned out to be partly false, it became much easier to think that all scientific theories are fallible. Especially since quantum mechanics some decades later repeated the lesson. For a long time, all versions of quantum mechanics contradicted both Newtonian mechanics and relativity theory. At the microscopic and macroscopic levels it gives approximately the same predictions as Newtonian mechanics, but at the micro-micro level some predictions differ dramatically. The original quantum mechanics contains, like Newtonian mechanics, Galilei-transformations and contradicts relativity theory.

Even if seldom spelled out aloud, the *epistemological view* that physical theories are fallible (which is not the same as putting forward a specific *methodology*) slowly entered physics. Among philosophers of science, only a few, in particular Karl Popper and Mario Bunge (b. 1919), drew the general conclusion that scientific knowledge is fallible and began to defend this view explicitly. Most philosophers interested in the natural sciences discarded epistemological realism (the view that we have at least partial knowledge of a mind-independent world). They became positivists and/or instrumentalists saying that all physical theories – classical physics, relativity theories and quantum mechanics included – should be regarded

as being only instruments for predictions about observable events; not as saying anything about substances, properties, relations, and processes in the world.

Today, at the beginning of the twenty-first century, the fallibilist view of science seems to have become the natural view among all researchers who think that scientific theories can describe structures in the world. Positivism and instrumentalism, on the other hand, have been substituted by social constructivism, i.e., the view that all structured entities that we can get hold of at bottom are like the entities we in our non-philosophical everyday lives call fictional, i.e., entities that like novel characters only exist in and through our language acts. Molecules are existentially put on a par with Hamlet. Scientific geniuses are doing the same kind of work as Shakespeare did. Most social constructivists say: 'Yes, there *might* be something out there in an external language-independent world, but even if there is, we can nonetheless not possibly know anything about it; so, let's forget it.' Put in Baconian terms: we cannot know anything else than our own idols – so, let's stick to them.

Fallibilism is the view that no empirical knowledge, not even scientific such knowledge, is absolutely certain or infallible, but in contradistinction to epistemological skepticism it is affirmative and claims that it is incredible to think that we have no knowledge at all. It presupposes the view that there is a mind-independent world, i.e., it presupposes 'ontological realism'. From its perspective, it is amazing what an influence the quest for certainty has had in science and philosophy. The epistemological dualism 'either certain knowledge or complete skepticism' echoes through the centuries. In philosophy, fallibilism was first stressed and baptized by the chemist and philosopher Charles S. Peirce (1839-1914). Often, Peirce is called a pragmatist, even a father of pragmatism, but his conception of truth differs radically from that of pragmatists such as William James (1842-1910) and John Dewey (1859-1952), not to speak of the most famous contemporary pragmatist, Richard Rorty (1931-2007). According to James and Dewey, truth is – schematically – what is practically useful, but Rorty wants to drop the notion of truth altogether. Peirce, in contrast, thinks that truth means correspondence to reality; but he also thinks that what is true can only show itself as a future consensus in the scientific community. He does not speak of consensus *instead* of

correspondence (as social constructivists have it), but of consensus *around* correspondence. He deserves to be called a ‘pragmatic realist’.

Popper and Bunge found their way to fallibilism, seemingly independently of Peirce, by reflecting on the development of physics that we have described above. It might be argued that even mathematics and logic are fallible disciplines, but we will not touch upon this philosophical issue. Nor will we bother about whether there is a couple of abstract philosophical statements such as ‘something exists’ or ‘I think, therefore I exist’ that may be regarded as supplying infallible knowledge.

Below, we will stress the Popperian concept of ‘truthlikeness’ (Bunge: ‘partial truth’). Such a concept is implicitly present in Peirce’s view that the scientific community is moving towards truths. (Let it be noted, though, that in saying this we skip over a subtle difference between Popper and Bunge on the one hand and Peirce on the other. The former find no problem in speaking about completely ‘mind-independently existing entities’. Peirce, however, sometimes seems to find such a notion semantically meaningless, but he does nonetheless allow himself to believe in the existence of real entities and define what is real as “anything that is not affected by men’s cognitions about it (Peirce, p. 299)”.)

Fallibilism is linked to openness to criticism. If science were infallible, then there would be methodological rules to make sure that the results are true, and scientists would be immune to criticism. But if science is regarded as fallible, the results can never be regarded as completely immune to criticism. However, not only believers in the infallibility of science make themselves untouchable by criticism, the same goes for social constructivists. If there is no knowledge at all, then of course the results of one’s ‘investigations’, i.e., one’s constructions, cannot be criticized for being false. Sometimes in some respect extremes meet. Here, in their refusal of taking criticism seriously, scientistic epistemological infallibility and constructivist epistemological nihilism go hand in hand. Fallibilism makes it possible to adhere simultaneously to the views that:

(a) science aims at truths
(b) science captures partial truths
(c) science accepts theories partly because of the way these conform to dominant views in the surrounding society.

Peirce and Bunge admit this social dimension of science, but do not comment upon it in the way sociologists of knowledge do. For some peculiar reason, despite being a fallibilist, Popper thinks that all sociology of knowledge implies epistemological relativism and, therefore, should be let down.

Outside the philosophy of science, Karl Popper is mostly known for his defense of democracy in *The Open Society and Its Enemies*. Within the philosophy of science, he is best known for his *falsifiability criterion.* Popper was, in the midst of Vienna, an early critic of logical positivism. He claimed that metaphysical speculation is by no means sheer semantic nonsense, and often even an important precursor to science. Nonetheless, just like the positivists, he thought there is a gap between science and metaphysics, and that science has to free itself from metaphysics, which, he stresses, includes pseudo-science. He even tried to find a criterion by means of which metaphysics could in a simple way be kept outside universities. He claimed that metaphysics is not at all, as the logical positivists had it, impossible to verify. To the contrary, he said, the problem with metaphysical views is that it is all *too easy* to find empirical support for them. For instance, some religious people can see the hands of god everywhere. Instead, what makes a view scientific is that it is falsifiable, i.e., that it can be shown to be false.

On Popper's view, true scientists, but no metaphysicians, are able to answer the question 'What empirical data would make you regard your theory/belief as being false?' We will later show in what way Popper overstates his case (Chapter 4.4), but this is of no consequence here. His general ontological and epistemological realism can be dissociated from his falsifiability criterion and his concrete methodological rules. In particular, this criterion and these rules can be cut loose from a notion that is crucial to fallibilism and which Popper verbalizes using three expressions: 'truthlikeness', 'verisimilitude', and 'approximation to truth'. This important notion (which has nothing to do with the probability

calculus) is unfortunately neglected outside circles of Popper experts and, as we have said, only implicit in Peirce. The core of Popper's fallibilist epistemological realism can be captured by the following thesis and proposal:

- Thesis: Every conceptualization and theory almost certainly contains some mismatch between theory and reality.
- Proposal: Seek truth but expect to find *truthlikeness*.

Popper's epistemological realism combines fallibilism with the traditional idea that truth seeking has to be the regulative idea of science; epistemological realism presupposes ontological realism. The key to Popper's mix is the notion of truthlikeness, roughly that a statement can be more or less true (which is not the same as 'probably being true'). The intuition behind this notion is easily captured. Compare the three assertions in each of the columns below:

1	The sun is shining from a completely blue sky	There are four main blood groups plus the Rh factor
2	It is somewhat cloudy	There are four main blood groups
3	It is raining	All blood has the same chemical composition

In both columns it holds true that if the first assertion is true, then the second assertion has a higher degree of truthlikeness and approximates truth better than the third one. This is *not* to say that the second one is epistemologically 'more likely to be wholly true' than the third one. Compare the following two pairs of sentences, 'X' represents assertions such as 'there are four main blood groups':

Ia) *probably*, X is true
Ib) *probably*, X has a high degree of truthlikeness

IIa) X is true
IIb) X has a high degree of truthlikeness

The sentences Ia and Ib hint at coherence relations between an assertion X and its evidence (see also Chapter 4.7), whereas the sentences IIa and IIb express relations between the assertion X and facts (truthmakers) in the world. The former sentences express evidential epistemological relations, the latter express semantic-ontological relations, i.e., they say something about the relationship between an assertion and the world. Note that in itself a sentence such as ‘there are four main blood groups’ has *both* evidential relations of conformance to other sentences and observations *and* a relation of correspondence to the world. Constructivists note only the former kind of relations (and reduce ‘truth’ to coherence), old-fashioned realists only the latter, but reflective fallibilists see both.

The idea of truthlikeness belongs to a correspondence theory of truth. Such theories say that the truth of an assertion (truthbearer) rests upon a relation (correspondence) that the assertion has to facts (truthmakers). There can be no degrees of ‘falsitylikeness’ since there are no non-existent facts to which an assertion can be related, but one may use the expression ‘being falsitylike’ as a metaphor for having a low degree of truthlikeness.

At the end of a line of all possible progressively better and better approximations to truth, there is of course truth. To introduce degrees of truthlikeness as a complement to the simple opposition between true and false is a bit – but only a bit – like switching from talking only about tall and short people to talking about the numerical or relative lengths of the same people. The difference is this. Length corresponds both to comparative (‘is longer than’) and numerical (‘is 10 cm long’) concepts of length, but there are no such concepts for verisimilitudes. All lengths can be linearly ordered (and thus be represented by a comparative concept), and a general numerical distance measure can be constructed for them (which gives us a quantitative concept). Popper thought that such concepts and measures of degrees of truthlikeness could be constructed, but like many others we think that the ensuing discussion shows that this is impossible (Keuth, Chapter 7). That is, we have only a qualitative or semi-comparative concept of truthlikeness. Some philosophers think that such a concept of truthlikeness can be of no use (Keuth, Chapter 7), but this is too rash a conclusion.

To demonstrate that even a semi-comparative concept of truthlikeness can be useful and important, we will use an analogy. We have no real

comparative concept for geometrical shapes, to say nothing of a quantitative concept and measure. Nonetheless, we continue to use our qualitative concept of shape; we talk about shapes, point to shapes, and speak informally about similarities with respect to shape. Sometimes we make crude estimates of similarity with respect to shapes and are able on this basis to order a small number of shapes linearly (shape A is more like B than C, and A is more like shape C than D, etc.); we might be said to have a semi-comparative concept. In our opinion, such estimates and orderings of a small number of cases are also sufficient to ground talk of degrees of truthlikeness.

In the same way that a meter scale cannot be used before it has been connected to something external to it, a standard meter, so the concept of truthlikeness of theories cannot be used until one has judged, for each domain in which one is working, some theory to be the most truthlike one. In this judgment, evidential relations, left out of account in the definition of truthlikeness, stage a comeback. As we have said, truthlikeness informally measures the degree of a theory's correspondence with facts, not the degree of its conformance to evidence; 'truthlikeness' is a notion distinct from 'known truthlikeness'. Nonetheless, in order to judge how close a theory comes to the facts, degrees of evidence must somewhere come into play. Note that such evidential judgments are commonplace decisions; they are made every time some course book in some discipline is chosen to tell students some facts.

The notion of truthlikeness is important for the following reason. The history of science tells us that it is no longer possible to believe that science progresses by simply adding one bit of truth to another. Now and then whole theory edifices have to be revised, and new conceptualizations introduced; this sort of development will probably continue for a long time, perhaps forever. If, in this predicament, one has recourse only to the polar opposition between true and false, and is asked whether one believes that there are any true theories, be it in the history of science, in today's science, or in the science of tomorrow, then one has to answer 'There are none'. If, however, one has recourse to the notion of truthlikeness, then one can answer as follows:

There are so far no absolutely true empirical theories, but, on the other hand, there are not many absolutely false theories either. Most theories in

the history of science have some degree of truthlikeness, even if only to a very low degree. Today, however, some theories have what is probably a very high degree of truthlikeness. Why? Because many modern inventions and modern standardized therapies which are based on scientific theories have proven extremely effective. It seems highly unlikely that all such inventions in technology and medicine are based on theories with very low degrees of truthlikeness, to say nothing of the thought that these theories are mere social fictions. Think, for instance, of traveling to the moon, images from Pluto, computers, the internet, the GPS system, magnetic resonance imaging, physiologic contraception, artificial insemination, and organ transplantation. Can they possibly be based on mere figments of the imagination?

It is now time to add a quotation from Popper in order to show how he himself summarizes his views on truthlikeness:

> I have in these last sections merely sketched a programme […] so as to obtain a concept of *verisimilitude* which allows us to speak, without fear of talking nonsense, of *theories which are better or worse approximations to truth*. I do not, of course, suggest that there can be a criterion for the applicability of this notion, any more than there is one for the notion of truth. But some of us (for example Einstein himself) sometimes wish to say such things as that we have reason to conjecture that Einstein's theory of gravity is *not true*, but that it is a *better approximation to truth* than Newton's. To be able to say such things with a good conscience seems to me a major desideratum of the methodology of the natural sciences (Popper 1972, p. 335).

Just as in ethics there are people who only think in terms of white or black, and who always want to avoid nuance and complication, so in science there are people who simply like to think only in terms of true or false/fictional. Not many decades ago scientists thought of their research only in terms of being certainly true; today, having familiarized themselves with the history of science, many think of it only in terms of being certainly false/fictional.

Popper's remark about criteria is more important than it might seem. Among other things, it has repercussions on how to view a phenomenon that Kuhn and Feyerabend have given the misleading name 'the incommensurability of basic theories'. It sounds as if it is claimed that basic theories in a scientific discipline are completely incomparable. But this is not the claim. Rather, 'incommensurability' here means un-translatability. As translators of plays, novels, and poems are well aware of, there can be parts of a text in one language that are impossible to give an exact translation in the other language; some concepts used in the first language have no exact counterpart in the other. And the same is often true of basic physical theories. For instance, where Newtonian mechanics has one single concept 'mass', special relativity has two, 'rest mass' and 'relativistic mass'; and the following holds true. If the Newtonian concept 'mass' has at all a counterpart in relativity theory, it must be 'rest mass', but these concepts are nonetheless not synonymous. Synonymous concepts can be contrasted with the same other concepts, but only 'rest mass' can be contrasted with 'relativistic mass'; 'mass' cannot. This un-translatability does not, however, imply general incomparability and epistemological relativism. Any physicist can compare the theories and realize that both cannot be wholly true. As translators are bilinguals, physicists may become bi-theoreticals. And as translators – without using any criterion manual – can discuss what is the best translation of an 'un-translatable' poem, physicists can – without using any criterion manual – discuss what is the most truthlike theory of two incommensurable theories.

Applying the notion of truthlikeness to the history and future of science allows us to think of scientific achievements the way engineers think of technological achievements. If a machine functions badly, engineers should try to improve it or invent a new and better machine; if a scientific theory has many theoretical problems and empirical anomalies, scientists should try to modify it or create a new and more truthlike theory. As in engineering it is natural and common to invent imperfect devices, in science it is natural and common to create theories that turn out not to be true. In both cases, however, there is an obligation to seek to improve things, i.e., improve problematic machines and problematic theories, respectively. Also, and for everybody, it is of course better to use existing

technological devices than to wait for tomorrow's, and it is better to trust existing truthlike theories than to wait for the science of tomorrow.

False assertions and fictional assertions are in one respect different and in another similar. They are different in that it is possible to tell a lie using a false assertion but not using a fictional one. When we lie we present as true an assertion that is false, but fictional assertions are beyond the ordinary true-false dimension. The two are similar in that neither refers to anything in reality that corresponds exactly to the assertion in question. A false empirical assertion lacks a truthmaker, and a fictional assertion cannot possibly have one. Therefore, it is easy to confuse the view that all theories are false with the view that all theories are about fictions. Nonetheless, it is astonishing how easily social constructivists move from speaking about false theories in the history of science to speaking about theories as being merely social constructions, i.e, as being about what is normally called complete fictions. Why don't they believe that stories can contain a mix of true statements and fictional statements?

If one assertion is more truthlike than another, then it is by definition also less false. However, this 'falsity content' (to take an expression from Popper) can easily be turned into a 'fictionality content', whereupon the more truthlike assertion can also be said to be a less fictional assertion. When we are reading about, say, Sherlock Holmes, we have no difficulty placing this fictional character in a real setting, London between 1881 and 1904. In many fictional discourses not everything is fictional, and we often have no difficulty apprehending such mixtures of real and fictional reference. Something similar is true when one reads about the history of science. For example, when one reads about the false hypothesis that there is a planet Vulcan between Mercury and the Sun, which would explain some anomalies that Newtonian mechanics were confronted with, there is no problem in taking Vulcan to be a fictional entity postulated as existing in the real solar system in about the same way as we take Holmes to be a fictional character in a real London. When one reads about the false hypothesis that there is a chemical substance, phlogiston, which *exits* burning material (where in truth, as we now know, oxygen *enters* burning material), then there is no problem in taking phlogiston to be a fictional substance in the world of real burnings. When one reads about Galen's view that the arterial system contains pneuma or spiritus, then there is no

problem in taking this pneuma to be fictional, but the arterial system to be real.

Those who write about the history of science often make the reader look upon statements which were once false assertions as being assertions about fictions. In retrospect, we should look upon superseded theories as *un*intentionally containing a mix of reality and fiction in the way reality and fiction can be intentionally mixed in novels. This is to give fictions their due place in science.

Apart from all other curiosities, social constructivism is self-reflectively inconsistent. Social constructs are created, but if everything is a construction, then nothing can construct. Unfortunately, social constructivists shun this kind of self-reflection.

The fact that Popper's fallibilistic epistemological realism is far more reasonable than all forms of positivism and social constructivism does not imply that it is in no need of improvements. We will stress a semantic observation that underpins epistemological realism; we will present it by means of a detour.

When we look at things such as stones, trees, and walls, we cannot see what is on the other side. But things like water and glass are such that we can look through them to the other side. In the case of glasses, microscopes, and telescopes, this feature is extremely useful. By *looking through* such lenses, we are able to have a better *look at* something else. This phenomenon of 'being-aware-of-*x*-through-*y*' is not restricted to the visual sense. It can be found in the tactile realm as well. You can grip a tool and feel the tool against your palm, but when you are very good at using such a tool, this feeling disappears. You are instead primarily aware of whatever it is that the tool is affecting or is affected by. For instance, when you are painting a wall with a brush, you are only (if at all) indirectly aware of your grip of the brush, and are instead aware only of the touching of the wall. You are *feeling through* the brush and *feeling (at)* the wall. What glasses are for people with bad sight, the white cane is for blind people.

Speech acts, listening acts, writing acts, and reading acts – in short, language acts – are, just like glasses and white canes, tools for improving everyday life. They can be used to convey and receive information, to give and take orders, to express emotions, and to do many other things. Even

though language acts do not have the same robust material character that tools have, they nonetheless display the same feature of being able to be both ‘looked at’ and ‘looked through’. When you look at linguistic entities, you are directly aware of them as linguistic entities, but when you look through them you are at most indirectly aware of them. When, for example, you are conveying or receiving information in a language in which you are able to make and understand language acts spontaneously, you are neither looking *at* the terms, concepts, statements, and propositions in question, nor are you looking at grammar and dialects. Rather, you are looking through these linguistic entities in order to see the information (facts, reality, objects) in question. When, then, are we looking *at* linguistic entities? We look at them, for example, when we are reading dictionaries and are examining terminologies. If I say ‘Look, the cat has fallen asleep’, I want someone to look through the term ‘cat’ and my assertion in order to receive information about a state of affairs in the world. But if I say ‘In WordNet, the noun ‘cat’ has 8 senses’, then I want someone to look at the term ‘cat’.

Our distinction between looking *at* and looking *through* is similar to the traditional distinction in semantics between the *use* and *mention* of linguistic entities, and it applies both to factual talk and to reading novels. In fictional discourse, terms are *used* as much as they are in talk about real things, but they are used in a very special way. Fictional discourse is *about* fictional characters; it is not about terms and concepts. In fact, we are standardly using the same terms and concepts both in fictional and factual discourse.

When you are not using lenses, you can look at them and investigate them as material objects of their own in the world. For instance, you can try to find out what their physical properties and internal structures are like. In the world of practice, we investigate tools this way only when they are not functioning properly and are in need of repairing. Something similar holds true of terms and concepts. We normally bother to look *at* terms and concepts in dictionaries only when our language acts are not functioning well – think for instance of learning a new language.

Furthermore, we are able to switch quickly between looking through and looking at things. Car drivers should look through, not at, the windshield, but when driving they should also have the ability to take a

very quick look *at* it in order to see whether, for instance, it has been damaged by a stone. Something similar is true of people using a foreign-language dictionary. They should be able to take a look at a certain foreign term and then immediately start to look through it by using it. Let us summarize:

1. In the same way that we can both look at and look through many material things, we can both look at and look through many linguistic entities.
2. In the same way that we can quickly switch between looking at and looking through glass, we can quickly switch between looking at and looking through linguistic entities.

And let us then continue the analogy by adding still another similarity:

3. In the same way that consciously invented material devices for 'being-aware-of-*x*-through-*y*', such as microscopes and telescopes, have provided new information about the world, consciously invented linguistic devices for 'being-aware-of-*x*-through-*y*', such as scientific conceptual systems, have provided new information about the world.

By means of the *invention* of new concepts, we can sometimes *discover* hitherto completely unnoticed facts. Often, we (rightly) regard discoveries and inventions as wholly distinct affairs. Some things, such as stones, can only be discovered, not invented; others, such as bicycles, seem only to be inventions. One person might invent and build a new kind of bicycle, and another person may later discover it; but the first person cannot both invent and discover it. These differences between inventing and discovering notwithstanding, devices for 'being-aware-of-*x*-through-*y*' present an intimate connection between invention and discovery. By means of new 'being-aware-of-*x*-through-*y*' inventions, we can discover *x*. There are many *x*'s that we can discover only in this way.

The third point above should partly be understood in terms of the notion of truthlikeness: if an existing conceptual system is confronted by a conflicting conceptual system which has a higher degree of truthlikeness, the latter should supersede the former. But the notion of truthlikeness should also be understood by means of the distinction between looking at

and looking through. We introduced the idea of truthlikeness with the three assertions ‘The sun is shining from a completely blue sky’, ‘It is somewhat cloudy’, ‘It is raining’, and we said that, *given that the first assertion is true*, the second one seems intuitively to be more truthlike than the third. A standard objection to such a thesis is that this sort of comparison can show us nothing relevant for a correspondence theory of truth, since what we are comparing are merely linguistic entities, namely assertions, and the result can only show conformances between assertions. However, this objection overlooks the distinction between looking at and looking through. Looking at the assertions allows us to see only conformances between the assertions as such, but when we have learned to switch from looking at them to looking through them – at reality – then we can coherently claim that the second corresponds better to reality (is more truthlike) than the third.

In the same way that our choice of kind of lens may determine what we are able to see, so our choice of concepts determines what we can grasp. Such a determination is compatible with the view that we can acquire knowledge about the world: it does not render truth a wholly social construction. When, through a concept, we look at and grasp some thing and/or features in the world, this concept often does for us at least three different things:

(i) it *selects* an aspect of the world (for instance, physical, biological, or social)
(ii) it *selects* a granularity level (for instance, microscopic or macroscopic)
(iii) it *creates* boundaries where there are no pre-given natural boundaries.

Nonetheless,

(iv) the concept *does not create* this aspect, this granularity level, or what is bounded.

Think of the concept ‘heart’. It selects a biological aspect of the human body, it selects a macroscopic granularity level, and it creates a boundary line between the heart and its surroundings, which does not track physical discontinuities at all points, as for example where the heart meets the aorta and the veins. But, nonetheless, our invention of the concept ‘heart’ does

not *create* our hearts, and there were hearts many millions of years before there were concepts.

Both perceptions and linguistic acts (talking, listening, writing, and reading) are intentional phenomena, i.e., they are *directed at* something which they are about. Like all intentional phenomena, they are marked by a tripartition between (intentional) *act* or state, (intentional) *content*, and (intentional) *object*. Assume that you are reading a physician's report about your heart, which tells you that your heart has some specific features. At a particular moment, there is then your reading *act* along with what you are reading about, the intentional *object*, i.e., your heart and its properties. But since your heart exists outside of your reading act, there must be something within the act itself in virtue of which you are directed towards your heart and its properties. This something is called the *content*; in assertions, it consists of propositions. When an assertion is completely false there is no corresponding intentional object; when it is completely true there is a corresponding intentional object; and when it is partly true there is only a partly corresponding intentional object.

A move made by many idealists in the history of philosophy is to argue that there *never* are any intentional objects that are distinct from the intentional contents of our acts of thinking and perceiving. Modern social constructivists make the same kind of move. But since they think that there is no thinking without a language and no perception not structured by language, they think that all there is are language acts and language content. It deserves the label 'linguistic idealism'.

Social constructivists often ask: 'From what position are you talking?' In order to answer this question, we will bring in Thomas Nagel (b. 1937). We regard ourselves as speaking from the kind of naturalist rationalist position that he has tried to work out in *The View from Nowhere* and *The Last Word*. Below are two quotations. The first is from the introduction to the latter book, and the second is its ending paragraph.

> The relativistic qualifier—"for me" or "for us"—has become almost a reflex, and with some vaguely philosophical support, it is often generalized into an interpretation of most deep disagreements of belief or method as due to different frames of reference, forms of thought or practice, or forms of life, between

> which there is no objective way of judging but only a contest for power. (The idea that everything is "constructed" belongs to the same family.) Since all justifications come to an end with what the people who accept them find acceptable and not in need of further justification, no conclusion, it is thought, can claim validity beyond the community whose acceptance validates it.
>
> The idea of reason, by contrast, refers to nonlocal and nonrelative methods of justification—methods that distinguish universally legitimate from illegitimate inferences and that aim at reaching the truth in a nonrelative sense. Those methods may fail, but that is their aim, and rational justification, even if they come to an end somewhere, cannot end with the qualifier "for me" if they are to make that claim (Nagel 1997, p. 4-5).

> Once we enter the world for our temporary stay in it, there is no alternative but to try to decide what to believe and how to live, and the only way to do that is by trying to decide what is the case and what is right. Even if we distance ourselves from some of our thoughts and impulses, and regard them from the outside, the process of trying to place ourselves in the world leads eventually to thoughts that we cannot think of as merely "ours." If we think at all, we must think of ourselves, individually and collectively, as submitting to the order of reasons rather than creating it (Nagel 1997, p. 143).

Reason, Nagel says, has to have the last word. However, this statement needs to be qualified. As a reviewer notes with regard to Nagel's book: "reason has the last word – or perhaps only the last but one, since reality, reason tells us, has always the absolutely last word"' (Lindström, p. 3-6). Let us in the next two chapters see how this last word may make itself visible among all the words we use.

Reference list

Bernstein RJ. *Beyond Objectivism and Relativism*. Basil Blackwell. Oxford 1983.

Blackburn S. *Truth. A Guide for the Perplexed*. Penguin. London 2004.

Bliss M. *The Discovery of Insulin*. The University of Chicago Press. Chicago 1982.

Broad W, Wade N. *Betrayers of the Truth. Fraud and Deceit in the Halls of Science*. Touchstone Books. New York 1983.

Bunge M. *Scientific Research*, 2 vol. Springer-Verlag. New York 1967.

Bunge M. *Emergence and Convergence. Qualitative Novelty and the Unity of Knowledge*. University of Toronto Press. Toronto 2004.

Bunge M. *Chasing Reality. Strife Over Realism*. University of Toronto Press. Toronto 2006.

Feyerabend P. *Against Method. Outline of an Anarchistic Theory of Knowledge*. Verso. London, New York 1993.

Fleck L. *Genesis and Development of a Scientific Fact*. The University of Chicago Press. Chicago 1977.

Gould SJ. *The Mismeasure of Man*. Norton & Company. New York, London 1996.

Haack S. *Manifesto of a Passionate Moderate: Unfashionable Essays*. University of Chicago Press. Chicago 1998.

Haack S. *Defending Science: Between Scientism and Cynicism*. Prometheus Books. Amherst NewYork 2005.

Hacking I. *The Social Construction of What?* Harvard University Press. Cambridge Mass. 1999.

Hanson NR. *Patterns of Discovery*. Cambridge University Press. Cambridge 1958.

Harvey W. *Anatomical Studies on the Motion of the Heart and Blood* (translated by Leake C). Springfield 1978.

Johansson I. *A Critique of Karl Popper's Methodology*. Scandinavian University Press. Gothenburg 1975.

Johansson I. Bioinformatics and Biological Reality. *Journal of Biomedical Informatics* 2006; 39: 274-87.

Judson HF. *The Great Betrayal. Fraud in Science*. Harcourt, Inc. Orlando 2004.

Kuhn TS. *The Structure of Scientific Revolutions*. University of Chicago Press. Chicago 1979.

Keuth H. *The Philosophy of Karl Popper*. Cambridge University Press. Cambridge 2004.

Latour B, Woolgar S. *Laboratory Life. The Construction of Scientific Facts*. Princeton University Press. Princeton 1986.

Lindström P. 'Näst sista ordet' ('The Last Word But One'). *Filosofisk tidskrift* 2001; no 1: 3-6.

Nagel T. *The View From Nowhere*. Oxford University Press. Oxford 1986.

Nagel T. *The Last Word*. Oxford University Press. Oxford 1997.

Niiniluoto I. *Critical Scientific Realism*. Oxford University Press. Oxford 2002.

Peirce CS. *The Philosophy of Peirce. Selected Writings* (ed. J Buchler). Routledge & Kegan Paul. London 1956.
Popper K. *The Open Society and Its Enemies*. Routledge & Kegan Paul. London 1966.
Popper K. *The Logic of Scientific Discovery*. Hutchinson. London 1968.
Popper K. *Conjectures and Refutations*. Routledge & Kegan Paul. London 1969.
Popper K. *Objective Knowledge*. Oxford University Press. London 1972.
Porter R. *The Greatest Benefit to Mankind. A Medical History of Humanity from Antiquity to the Present*. Fontana Press. London 1999.
Proctor R. *Racial Hygiene. Medicine Under the Nazis*. Harvard University Press. Cambridge Mass. 1988.
Roll-Hansen N. *The Lysenko Effect: The Politics of Science*. Humanity Books. New York 2005.
Service RF. Bell Labs Fires Star Physicist Found Guilty of Forging Data. *Science* 2002; 298: 30-1.
Semmelweis IP. *Die Aetiologie, der Begriff und die Prophylaxis des Kindbettfiebers*. Hartlen's Verlag-Expedition. Pest, Wien and Leipzig 1861.
Soyfer VN. *Lysenko and the Tragedy of Soviet Science*. Rutgers University Press. New Jersey 1994.
Williams B. *Truth and Truthfulness*. University Presses of California. Columbia and Princeton 2004.

4. What Does Scientific Argumentation Look Like?

Some arguments and inferences are theoretical in the sense that they are meant to show that a certain view is true or probably true, whereas others are practical in the sense that they are meant to show that one ought to act in a certain way. In both kinds of cases a certain conclusion (C) is meant to follow from some other things that are called premises (P). Schematically: 'P, hence: C'. Between the premises and the conclusion there is always a link of some kind; sometimes this 'hence' is called an 'inference', sometimes not. In this chapter we shall present different such links. We will mainly focus on theoretical argumentation, but some space will be devoted to practical inferences too. The former dominate in medical science, but the latter in the everyday life of clinicians.

What is a conclusion in one argument may be a premise in another. Arguments can sometimes be put together into long chains of argumentations. Even for such immaterial kinds of chains, it can hold true: no chain is stronger than its weakest link. However, arguments of different kinds can also be wired around each other, and then the weakest-link metaphor breaks down. According to the father of fallibilism, Peirce, our "reasoning should not form a chain which is no stronger than its weakest link, but a cable whose fibers may be ever so slender, provided they are sufficiently numerous and intimately connected (Peirce, p. 229)." It should be added that new epistemological situations may require new 'cables', and that this view makes argumentations situation-bound (the same kind of 'particularism' holds true also for moral reasoning, as we will explain in Chapter 9.4).

Taken in isolation, some arguments are strong and some are weak. The strongest possible theoretical link is the deductive inference, but it is only in sciences such as mathematics and logic that it makes sense to try to rely merely on deductive arguments. All the empirical sciences have to rely on weaker – and much weaker – forms of argumentations. In the empirical sciences there is fallibility and epistemic uncertainty, but, on the other

hand, in pure logic and pure mathematics there is no contact with the world in space and time. The basic form of empirical argumentation is *argumentation from perceptions*. Explicitly or implicitly, we argue by saying ‘I saw it!’ Everyday life is impregnated with such arguments. Especially, we can be said to argue with ourselves in this way. If you want to know if it is raining, you look out the window; if you want to know if there is milk in the fridge, you open the door and look; and so on. This form of argumentation, with its tacit presupposition that most perceptions are veridical, will merely form the background for most of the things we will say in this chapter.

Perhaps in a not too distant future, this chapter will be deemed to suffer from a neglect of ‘arguments by pictures’ and ‘arguments by diagrams’; even though we have already surpassed the view that all argumentation takes place by means of sentences. We have just noted the existence of immediate perception-assertion transitions that can take place in arguments from perceptions, and we have earlier (Chapter 3.2) noted the phenomenon of ‘perceptual structuring’. Since long, physicians have argued by means of X-ray pictures; recently, computer tomography, magnetic resonance imaging (MRI), functional neuroimaging, and even other imaging technologies have entered the medical scene on a broad scale; and there seems to be good reasons to think that in the future one can even reason with the help of ‘pictures’ that are not real pictures but computer simulations. This being noted let us start our presentation.

4.1 Arguments ad hominem

Many modern advertisements tell the onlookers: ‘This celebrity is doing A, hence: you ought to do A!’ The point of such practical ‘ads-arguments’ is that we should trust the famous person in question, and behave as he is said to do. Compare with old-time medicine: the famous Hippocrates says that a good physician should do this and avoid that, therefore as a physician you ought to do so as well. This kind of argumentation has also a theoretical form: ‘This celebrity claims that S is true, hence: you ought to believe that S is true!’; ‘Einstein claims that S is true, hence: you ought to believe that S is true, too!’ In logic, arguments of this kind are sorted out as a special kind of logically *invalid* arguments called *arguments ad hominem*, i.e., they are arguments ‘to a man’ instead of arguments that present axioms,

principles, or empirical evidence. Arguments ad hominem may be persuasive, but they cannot be convincing; at least not if to be convinced of a view means being able to offer evidence in favor of this view. Marketers can rest content with persuasion, but scientists should aim at convincing. Nonetheless, science cannot do without arguments ad hominem.

Arguments ad hominem are of course often needed in the relation between scientists and laymen; much research can be judged only by experts. But they are also needed between scientists themselves. The reason is that science, like most modern enterprises, rests on a far-reaching division of labor. Not even skeptical research groups can rely only on their own reasoning and observations; they have also to trust other scientific knowledge authorities. Even though the scientific community as a whole has only two epistemological sources, observation and reason, each individual scientist and research group has three:

- observation
- reason
- trust in information from others.

All these three sources are fallible. We can make wrong observations, reason wrongly, and trust the wrong people. But we shall try to avoid it.

Perhaps trust comes primarily to mind when one thinks of small groups such as families, clans, and friend circles, but trust is central to human societies at large and to scientific communities, too. Not to speak of the relation between patients and physicians. Depending on context, an argument ad hominem can be either an attempt at pure persuasion or a necessary move in a rational process that aims at convincement. In human life, the general ad hominem structure is inevitable, but what authority we shall put in the premise as ‘the man’ can very much be a matter for occasional discussion. Here are four different ad hominem schemas:

	(1) celebrity does A	(2) parent says ‘H is true’
hence:	----------	----------
	(1) I will do A	(2) I believe ‘H is true’

	(3) physician says ‘do A’	(4) scientist says ‘H is true’
hence:	---	---
	(3) I will ‘do A’	(4) I believe ‘H is true’

Let us mention one famous example of medical trust and one of mistrust. In the 1970s, double Nobel Prize winner Linus Pauling (1901-1994) claimed that massive doses of vitamin C are good both for treating and preventing colds. As a result, rather immediately, many people started consuming vitamin C when having a cold. They simply trusted Pauling. When Max von Pettenkofer (see Chapter 2.5) claimed that he had drunk a culture of cholera bacteria without becoming ill, there was no other person who could testify this. Pettenkofer simply declared that he had done it, and he expected to be trusted. His assistant wrote to Koch:

> Herr Doctor Pettenkofer presents his compliments to Herr Doctor Professor Koch and thanks him for the flask containing the so-called cholera vibrios, which he was kind enough to send. Herr Doctor Pettenkofer has now drunk the entire contents and is happy to be able to inform Herr Doctor Professor Koch that he remains in his usual good health.

But Koch seems not to have considered his letter.

Before randomized controlled trials (RCTs) in the second half of the twentieth century became the golden standard for assessing medical technologies, new medicines and surgical methods were launched only on the authority of some prominent physician. The treatments in questions were mostly based on long personal experience and not gratuitous, but they were nonetheless not evidence-based in the modern sense. This older argumentative mode, with its appeals to authoritative figures, is still frequently used when so-called complementary or alternative medical treatments are launched on the market. Personal experience is very important for both clinicians and researchers (see Chapter 5), but it is not – by far – in itself a sufficient tool for assessing the effects of medical treatments.

From a social perspective, the increased use of RCTs might be regarded as a democratization of medical assessments. Since more than one person

may influence this kind of evaluative process, both openness and corrigibility of the assessments increase.

There are since long different forms of RCTs, e.g., ‘single-blind’ and ‘double-blind’ (see Chapter 6.3), but a wholly new step in the development of RCTs was taken (starting in the early nineties) by the Cochrane Collaboration groups. These try to take account also of the fact that, often, different medical research groups that evaluate the same treatment obtain different results. Therefore, there is a need also for groups that perform *meta-analyses* of such conflicting RCTs. Today, the medical community is inclined to trust representatives from Cochrane groups or centers more than others when the effects and the efficacy of medical technologies are under discussion.

The first philosophical advocate of the industrial revolution, Francis Bacon, is famous for the saying ‘knowledge is power’. He meant that knowledge of nature allows us to master nature. Today, it should be noted, this phrase is often given quite another meaning; or, rather, two meanings that are confounded. It is used (a) to claim that power and knowledge are socially inseparable, and in this sense the saying contains a kernel of truth. Having power in some respect often implies being able to tell what should be regarded as true in the corresponding area. Conversely, to be regarded as someone who can tell the truth is often to have a considerable amount of power. The Cochrane Collaboration groups have a kind of power. However, the saying ‘knowledge is power’ is also in radical social constructivist writings used (b) to claim that there is no knowledge at all in the traditional sense, only power disguising itself as knowledge. From such a curious perspective, *all* arguments are necessarily arguments ad hominem. We regard it as disastrous. If it is adopted, one can call all scientific argumentation, and even clinicians’ diagnoses, ‘symbolic terror’. Humpty Dumpty, from *Alice in Wonderland*, was in this sense ahead of his time when he argued with Alice:

> - But ‘glory’ doesn’t mean ‘a nice knock-down argument’, Alice objected.
> - When *I* use a word, Humpty Dumpty said, in rather a scornful tone, it means just what I choose it to mean - neither more nor less.

> - The question is, said Alice, whether you *can* make words mean so many different things.
> - The question is, said Humpty Dumpty, which is to be master – that's all.

Questions of power and legitimacy are often relevant in medical scientific discussions. Think, for instance, of the recurring discussions about whether or not the power of pharmaceutical companies influences the production of clinical knowledge. For instance, it was recently reported in *The Journal of American Medical Association* (May 17, 2006) that concerning vascular diseases, clinical trials funded by profit based organizations appeared more likely to report positive findings than those funded by non-profit organizations. Such facts, however, are quite compatible with the fact that medical research has a truly epistemological and logical side to it too. Some arguments in science have to be allowed to be arguments ad hominem, but all arguments cannot be allowed to be of this kind.

There are also *negative* ad hominem arguments. They claim that a certain view is false since it is defended by a person that is regarded as being a bit crazy, belonging to some enemy camp, or as simply benefiting from putting forward such a view.

4.2 Deductive and inductive inferences

Some arguments are able to show that from views already accepted as true, one can infer that, necessarily, a certain other view is true, too. In such cases, the argumentative link is a deductive inference. In the other kind of inferences that we shall discuss in this section, e.g., inductive inferences, there is no similar necessity to be found between premises and conclusions (this kind of induction has, note, nothing to do with 'mathematical induction', not to speak of 'electromagnetic induction', 'enzyme induction' and 'immunological induction'). If two persons share some premises but differ with respect to a conclusion that follows deductively, then at least one of them is in serious argumentative trouble; if the conclusion follows only inductively, this need not be the case.

In both deductive and inductive inferences, premises and conclusions are entities (beliefs, assertions, sentences, statements, propositions) that can be

ascribed a truth-value, i.e., they can be regarded as being true or false, even though their *actual* truth-value in fact makes no difference to the validity of the inference in question. Inferences that are presumed to be deductive inferences but are not, are often called invalid deductions. A very simple (valid) deductive inference is:

premise 1:	*if* there are red spots all over the body, *then* there is the measles
premise 2:	there are red spots all over the body
hence:	----------
conclusion:	there is the measles

No physician would admit the first premise to be true without qualifications, but this is beside the point at issue. A deduction is characterized by the fact that, necessarily, *if* the premises are true *then* the conclusion is true, too. A deduction is truth-transferring and truth-preserving, but not truth-guaranteeing. If there is no truth to transfer and preserve, the truth of the conclusion cannot be guaranteed. Deductions tell us neither (a) whether the premises are true or false, nor (b) whether the conclusion is true or false when the premises are false. But they tell us (c) that the premises cannot be true and the conclusion false. Here comes another deduction:

premise 1:	*if* there are red spots all over the body, *then* there is cancer
premise 2:	there are red spots all over the body
hence:	----------
conclusion:	there is cancer

These two deductive inferences acquire their truth-preserving power from the *form* of the assertions involved, not from the substantive content of the assertions. Both inferences have the same form. This can be seen in the following way. If we let p and q be variables for assertions (or ‘propositions’, as logicians prefer to say), the form of the examples can be represented by this schema:

premise 1: *if* p, *then* q
premise 2: p
hence: ----------
conclusion: q

Whatever specific assertions we put in as values of p and q in this schema, it will be the case that the premises cannot be true and the conclusion false. Even though this schema does not contain any substantial semantic content, it is of course not completely semantically empty. First, the variables are variables only for entities that can be true or false; second, there is the semantic content of the expression ‘if … then’; third, since we have premise 1 *and* premise 2, there is also the semantic content of this implicitly present ‘and’. Expressions such as ‘if … then’, ‘and’, ‘not’, ‘all’, ‘some’, etc. are called logical constants. To claim that the inference schema above represents a *deductive inference* is equivalent to claim that the statement below is true:

- necessarily, if [(if p then q) and p], then q.

The statement ‘if [(if p then q) and p], then q’ is one of several *formal-logical truths*. It – and the corresponding inference schema earlier presented – is since medieval scholasticism called ‘modus ponens’. Another famous deductive inference schema is called ‘modus tollens’:

premise 1: *if* p, *then* q
premise 2: *not*-q
hence: ----------
conclusion: *not*-p

This form can be exemplified by:

premise 1: *if* there are red spots all over the body, *then* there is the measles
premise 2: it is not the case: there is the measles
hence: ----------
conclusion: it is not the case: there are red spots all over the body

We would also like to present two schemas that do not represent deductive (nor inductive) inferences; we will make use of them later on. We call them (a) 'inverse modus ponens' and (b) 'implication-equivalence conflation', respectively:

	(a)	(b)
premise:	*if* p, *then* q	
premise:	q	*if* p, *then* q
hence:	---------- (INVALID)	---------- (INVALID)
conclusion:	p	*if* q, *then* p

So far, we have presented parts of *propositional logic*, i.e., a logic where the variables in the inference schemas are variables for whole assertions. If a presumed inference is deductively valid in propositional logic, it is deductively valid – period. However, even if a presumed inference that is concerned with everyday assertions is *invalid* in propositional logic, it might nonetheless be valid in a logic that also takes the internal structure of the assertions/propositions at hand into account. Look at the following classic inference:

premise 1: all human beings are mortal
premise 2: Socrates is a human being
hence: ----------
conclusion: Socrates is mortal

When this inference is formalized in propositional logic, we arrive at the invalid formal schema below (from two arbitrary assertions one cannot deduce a third arbitrary assertion):

	propositional logic
premise 1:	p
premise 2:	q (= an assertion that may be distinct from p)
hence:	---------- (INVALID)
conclusion:	r (= an assertion that may be distinct from p and q)

However, if the same inference about Socrates is formalized in Aristotelian subject-predicate logic or term logic, we obtain the more fine-grained and valid formal schema below:

	Aristotelian term logic
premise 1:	all H are M
premise 2:	*a* is H
hence:	----------
conclusion:	*a* is M

In everyday life, we are not only following grammatical rules without noticing, we are also following many deductive inference rules without noticing. In the Socrates-is-mortal example, most people realize immediately that it follows from the premises that Socrates has to be mortal, but few are able to state explicitly that this is due only to the exemplified logical form. This fact of unconscious deductive inferences means, among other things, that in medical research there is usually no need to make simple deductions as those above explicit. Nonetheless, it is important for medical researchers to know explicitly what is characteristic of deductive inferences. The reason is that not only formal-logical derivations but also mathematical derivations are deductions, and mathematics plays a great role in modern medical science.

Even though purely mathematical truths such as ‘1+2 = 3’ and ‘7·8 = 56’ and the corresponding inference rules seem to differ radically in form from formal-logical truths such as ‘if [(if p then q) and p], then q’ and the corresponding inference rules, there is a certain similarity. Neither kind of assertions refers to anything in the world. It is impossible to point at something in the world and claim that it is simply ‘2’; it has to be 2 *of something*, like ‘2 chairs’, ‘2 cm’, and so on. Some famous philosophers, in particular Gottlob Frege (1848-1925) and Bertrand Russell (1872-1970),

have argued, appearances notwithstanding, that mathematical truths at bottom simply *are* formal-logical truths, but no consensus to this effect has emerged in the philosophical community. However, everyone agrees that if from some pure mathematical statements (premises) one can mathematically derive a certain conclusion, then (in two equivalent formulations):

- necessarily, if the premises are true, then the conclusion is true;
- not possibly, the premises are true, but the conclusion is false.

If someone claims that the premises of a certain mathematical inference are true, but that the conclusion is false, then he is as illogical as someone who denies a formal-logical deductive inference.

Let us now turn to inductive inferences. Even if an inductive inference has been made in the best possible way, it may nonetheless be the case that the premises are true but the conclusion false. This fact may easily be overlooked in empirical-statistical research when mathematical methods are used in the transition from descriptions of the sample (the premises) to stating some truth about the corresponding population (the conclusion); we will return to this below in Section 4.7.5.

In an inductive inference, the premises describe a number of (presumed or real) facts of a certain kind; normally, they are the results of a number of already made observations, e.g., ‘some swans, namely the observed ones, are white’. The conclusion describes cases – *of the same kind* – that, to start with, were not regarded as facts. An inductive conclusion can describe either

(i) a single case of the kind under consideration
(‘*the next* swan will be white, too’) ,
(iia) a finite number of such cases
(‘*all* the swans in the zoo are white’), or
(iib) all cases of this kind whatsoever
(‘*all* swans are white’).

In the last case we have a real (unlimited) generalization. In an inductive inference a step is taken either from *some* to *the next* of the same kind as in

(i), or, more importantly in research, from *some* to *all* of the same kind as in (iia) and (iib). For instance, from the fact that all men so far (= some men) have died, I can try to infer inductively that (i) I will die, that (iia) all now living persons will die, and (iib) all men whatsoever will die. Let us now see how this reasoning can be applied to a medical example.

Assume that a clinician observes that his depressed male patients who are treated with SSRI (serotonin selective reuptake inhibitors) as a side effect are relieved also of another problem: premature ejaculation. Based on his observations (sample) he first draws the limited inductive inference (i) that the same will happen with his next patient, but later on he makes a true inductive generalization (iii) to the effect that SSRI always has a beneficial effect on men suffering from premature ejaculations. Schematically:

premise:	*some* men, suffering from premature ejaculation, are cured by SSRI
hence:	----------
conclusion:	*all* men, suffering from premature ejaculations, will be cured by SSRI

This ‘hence’ of inductive generalizations can mean three different things, which, from an epistemological point of view, have to be kept distinct. We will use only the third one. First, ‘hence’ might mean only that the premise makes it reasonable to put forward *the conclusion as a hypothesis* that, in turn, ought later to be tested (e.g., by means of randomized controlled trials; see Chapter 6.3). Looked upon in this way, i.e., as being only hypotheses generators, inductive generalizations are of course epistemically unproblematic. Second, an inductive ‘hence’ might be read as saying that the truths of the premises are transferred to the conclusion. This reading should not at all be allowed, since there simply never is such a truth-transference; if there were, we would have a deductive inference.

Third, and reasonably, ‘hence’ might be interpreted as saying that the inductive inference schema makes – if the premises are true – the truth of the conclusion epistemically probable. Or, in other words, that the premises *give inductive support* to the conclusion, i.e.:

- probably, if the premises are true, then the conclusion is true.

This understanding of induction is quite consistent with the fact that the premises might be true but the conclusion false. The epistemic probability mentioned take degrees, i.e., in particular cases it may be so small that the conclusion should not be considered. When the epistemic probability of the conclusion comes *only* from the forms of the inductive inference (‘*some* to the *next*’ or ‘*some* to *all*’) it is almost certainly too small to be of interest. In interesting inductions, some epistemic probability has in each concrete case to come also from intuitions connected to the substantive content displayed in the premises and the conclusion of the induction; here, so-called tacit knowledge (Chapter 5) may play a part.

The essence of induction in research is a transition from ‘*some* of a certain kind’ to ‘*all* of the same kind’; in everyday life it is mostly a transition to ‘*the next* of the same kind’. However, it is seldom a matter of only an immediate transition ‘some → all’, as when people in Europe once upon a time made an inference from ‘some (all observed) swans are white’ to ‘all swans are white’. Mostly, there is a pre-existing ‘all-premise’ as well, and the skeleton of the inductive procedure had better be represented as: ‘(all →) some → all’. In modern research, something takes place before the inductive transition from sample to population is made. The observations that ground the ‘some-assertions’ in question are not mere chance observations. Let us explain by means of a simple non-medical example. Assume that we want to make an inductive forecast about how people will vote in the next election. How should we proceed? Answer: the two arrows in the form ‘all (1) → some (2) → all’ ought broadly to look as follows:

1. based on our knowledge of how *all* citizens (population) in various regions, occupations, ages, and so on have voted in earlier elections, we try to find a group of *some* citizens (sample) that mirror the overall voting pattern; when this is done, we ask the persons in the sample how they will vote, whereby we obtain the facts that constitute the premises in our inductive inference;

2. from these facts about how *some* citizens will vote, we inductively infer how *all* will vote.

Of epistemological importance is the fact that what step 2 concretely looks like is dependent on what step 1 looks like. At the same as one decides what the sample of the population should look like, one decides how one should return from the sample to the population. Therefore, to people who are taught only to make the second step, and to make it by means of mathematical formulas, it may falsely look as if they are simply deducing facts about the population (the conclusions) from the facts in the sample (the premises). The inductive inferences become hidden in the hypothesis that the sample is representative.

Most philosophers and many scientists living before ‘the fallibilistic revolution’ (Chapter 3) wanted and searched for certain and incontestable knowledge. To them, it was quite a problem that inductive inferences do not transfer truth from premises to conclusions; it was called ‘the problem of induction’. Since this problem was first clearly stated by the Scottish philosopher David Hume (1711-1776), it is also called ‘Hume’s problem’. But already the so-called Ancient Skeptics of the Graeco-Roman culture had seen it.

Consider now, to end this section, the following inference:

premise 1:	everyone suffering disease D is cured by treatment T
premise 2:	Paul is suffering disease D and has received treatment T
hence:	----------
conclusion:	Paul will be cured

This inference is not an induction. It is a deduction. Nonetheless it is by no means certain that Paul will be cured. Why? Since the first premise is based on an inductive inference it might well be false; therefore, premise 1 contains no certain truth that can be transferred to the conclusion.

Whereas formal-logical and mathematical deductive inferences cannot possibly expand the empirical content of the premises, such an expansion is precisely the aim of inductive inferences. The remarkable fact is that empirical assertions, especially in the form of mathematically formulated laws, can have an empirical content that the speakers are not fully aware

of. Therefore, even though deductive inferences cannot expand the empirical content of the premises, they can – and often do – expand the *explicitly known* such content. Formal-logical inferences cannot expand the content because their necessity relies only on the form of the premises; mathematical inferences cannot expand empirical content because they are concerned only with relations between space- and timeless numbers.

4.3 Thought experiments and *reductio ad absurdum* arguments

Aristotle (384-322 BC) is the first philosopher to realize that logical inferences ground their validity *only* in the *form* of the assertions made, and that therefore such inferences are valid independently of the truth of the premises. But he also noted another and related thing. Even when we make inductive inferences, and even when we put forward arguments ad hominem, we have to talk and think in conformity with *the law of contradiction*, i.e., in situations where we are interested in truth, we are not allowed to regard as true assertions that have the logical form of a contradiction. (The law is sometimes, and more adequately, also called 'the law of non-contradiction'.) In propositional logic, the basic form for a contradiction is 'p and not-p', an example being 'Jones has fever and it is not the case that he has fever'; in Aristotelian term logic, the basic form for a contradiction is 'S is P and not-P', an example being 'Jones' body temperature is 39°C and 37°C'. All substantial assertions that fit into the logical forms 'p and not-p' and 'S is P and not-P' are necessarily false because of their logical form. To reject contradictions is a basic precondition for *all* form of rational argumentation, however weak the form is in other respects. To deny a valid deduction is to state a contradiction. Of course, when joking and when talking metaphorically, contradictions can make very good sense.

Allowing contradictions means leaving the realm of argumentation completely behind. The peculiar status of logical contradictions can be made visible as follows. Assume that someone wants to argue that logical contradictions in spite of all need not be false. Then, in relation to at least one contradiction 'p_1 and not-p_1' this person has not only to

- maintain that 'p_1 and not-p_1' is true,
  he also has to
- deny that 'p_1 and not-p_1' is false.

That is, he has to take it for granted that 'p_1 and not-p_1' cannot be both true and false simultaneously. In his denial of the law of contradiction he has nonetheless to presuppose that the statement 'p_1 and not-p_1' conforms to the law of contradiction. The law has this feature: in order to question it one has to presuppose it. It can be neglected, but it can never be argued away. If one clinician says 'this patient has a lung infection', but another says 'no, he has not', and both think that there is no conflict, they have entered the realm of the absurd. If one physicist says without any hidden qualifications 'Newtonian mechanics is true', but another says 'no, the special relativity theory is true', and both think that there is no conflict, they have also entered the realm of the absurd. Note, however, that if the first physicist says 'Newtonian mechanics is false for velocities close to and above that of the velocity of light, but approximately true for lower velocities' then there is no contradiction or absurdity.

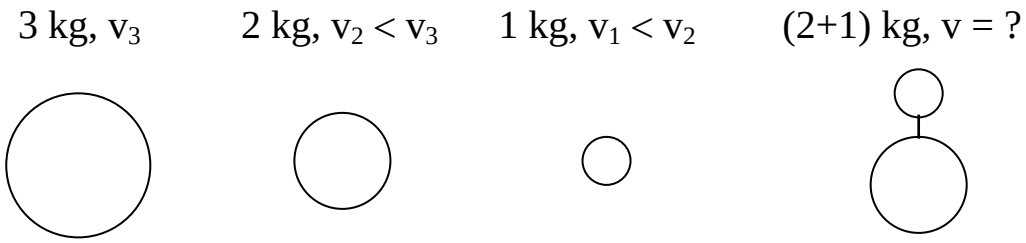
Now and then, both in science and philosophy, people make use of a line of argumentation that is termed 'reductio ad absurdum'. In such an argument, a seemingly reasonable view is shown to contain or imply, unexpectedly, a logical contradiction or some other kind of wholly unreasonable view. Since the first view (the premise, p) is 'reduced' to something absurd, it cannot be accepted. The whole argumentation with its 'reduction to the absurd' and the eventual rejection of the contested view (p) has the form of modus tollens:

premise 1:	*if* p, *then* q
premise 2:	*not*-q (since q is absurd; perhaps is in itself a contradiction)
hence:	----------
conclusion:	*not*-p

The reductio ad absurdum pattern is often part of thought experiments that are intended to show that a certain view *must* be false. Here comes a famous example from the history of physics, Galileo Galilei's arguments to

the effect that the kinetic theories of his predecessors could not possibly be true.

According to the view of the causes of movements that Galilei criticized, heavy bodies fall faster than lighter ones. This presumed law of nature was apparently supported by many simple empirical observations. If we simultaneously drop a book (heavy thing) and a sheet of paper (light thing), the book reaches the floor before the paper. Therefore, according to this law, the following is true: 'a stone weighing 3 kg will fall with a velocity (v_3) that is higher than that of a stone of 2 kg (v_2), and one of 2 kg will fall faster than one of 1 kg (v_1)'. Let us now in thought connect a 2 kg stone to a 1 kg stone, and ask: 'How fast will this new stone fall?' Since it weighs 3 kg, we can deduce from our premises that 'it will fall faster than a 2 kg stone' (p). However, since its 1 kg part will fall more slowly than the 2 kg part, the 1 kg part ought according to our premises slow down the speed of the 2 kg part, and so: 'the combined 3 kg stone will move slower than a 2 kg stone' (not-p).



From the premises of the old theory (= t), a contradiction (p and not-p) is deduced: a 3 kg body falls both more rapidly and more slowly than a 2 kg body. We have the inference schema:

premise 1:	*if* t, *then* (p *and not*-p)
premise 2:	*not*- (p *and not*-p)
hence:	----------
conclusion:	*not*-t

According to the argument, the old kinetics has to be regarded as false and, consequently, it ought to be abandoned or at least revised. Galilei's own alternative successful hypothesis was that all the four bodies mentioned, quite independently of their weight, fall at the same speed – but

only in a vacuum. A sheet of paper falls more slowly to the ground than a book because the air resistance functions differently in the two cases; it becomes greater on the sheet of paper.

4.4 Hypothetico-deductive arguments

The next argumentative pattern will be called ‘hypothetico-deductive arguments’. In twentieth century philosophy of science, there was much talk about the hypothetico-deductive *method*. Famously, in their search for a unified science the logical positivists (Chapter 3.4) hailed it as *the* method of science. By now, however, it is quite clear that it is merely one of several kinds of methods or argumentative patterns within the sciences. In itself, the hypothetico-deductive method is neither a sufficient nor a necessary condition for calling an investigation or argumentation scientific. Its name does not contain the term ‘inductive’, but, as we will make clear during the exposition, hypothetico-deductive arguments/inferences can, just like *inductive inferences*, supply no more than *inductive support* for the hypotheses and generalizations under consideration.

In order to make the presentation simple, we will take our main illustrative example from proto-science. We want to display a simple abstract pattern, but one that in more complicated and hidden forms exist also in real modern science. It is in principle easy to implement numerically formulated laws in the schemas below. Assume that (living long ago) we have suddenly noticed that some metal pieces expand when heated. If we are old-fashioned inductivists, we might then reason as follows:

premise:	when heated, these (*some*) pieces of metal expand
hence:	----------
conclusion:	when heated, *all* pieces of metal expand

If, instead, we want to use a hypothetico-deductive argument, we shall start with what is above the end point, the conclusion. In this kind of argument, one always starts with a *general hypothesis*. Normally, such a hypothesis is based on some observations, but from the point of view of the argument itself, it is completely uninteresting what has made the hypothesis arise in someone’s mind. The point is how to test it, and such a test has three steps.

In the first step, from the general *hypothesis* and descriptions of some spatiotemporally specific *initial conditions* we deduce a *test implication* (for the hypothesis). Since the law has a conditional form, *if* heating *then* expanding, the categorically formulated so-called initial conditions are needed in order to come in contact with a determinate portion of reality. When the law has numerical variables, the initial conditions assign values to the variables, but here comes the proto-scientific example:

hypothesis:	when heated, *all* pieces of metal expand
initial conditions:	these (*some*) pieces of metal are heated
hence:	----------
test implication:	these (*some*) pieces of metal expand

According to deductive logic (modus ponens), *if* the premises (the general hypothesis and the initial conditions) are true, then the test implication is true too. But is the test implication true? Answer: we cannot know. Even if we take it for granted that the premises constituted by the initial conditions are absolutely true, we cannot know that the test implication is true since, by definition, it is partly derived from a *hypothesis*. Whether the test implication is true or false has to be tested empirically and independently of the inference displayed; this is of course the reason why it is called a ‘test implication’ (for the hypothesis).

The second step of the test consists in trying to find out whether the test implication is true or not; the third step in finding out what conclusion we can then draw with respect to the truth-value of the hypothesis itself. Let us first investigate the third step on the assumption that the second step has made us believe that the test implication is true. In a way, but with a difference, we are now back in the inductive situation described in Section 4.3. Since we want to say something about the truth of the general hypothesis, we have the following schema (where the premises are taken to be true):

initial conditions:	these (*some*) pieces of metal are heated
test implication:	these (*some*) pieces of metal expand
hence:	----------
hypothesis:	when heated, *all* pieces of metal expand

This schema does not represent any deductive inference; here, the truth of the premises is *not* transferred to the hypothesis. They only give *inductive support.*

Out of context, the conjunction of the two premises is equivalent to the single premise in the inductive inference described at the beginning of this section: ‘when heated, these (*some*) pieces of metal expand’. However, the contexts make them differ. The premise of the pure inductive inference is only backward-looking, i.e., people doing pure inductive inferences are only relying on what they have already got, but the second premise of the last schema was once a matter of prediction. That is, the hypothetico-deductive procedure contains a stage where one consciously has to *search for new observations*. This might seem a negligible difference, but from a methodological point of view it is not. We are quite prone to see what we would like to see, and most scientists are somewhat anxious that their hypotheses may turn out to be false. Therefore, it is important to make some careful new observations even when one’s hypothesis has been created on the basis of quite accurate old observations.

Another important difference between inductive arguments and hypothetico-deductive arguments is that only the latter can give inductive support to hypotheses about unobservable entities. Many sciences are preoccupied with things and properties that are not directly observable, but no inductive argument can take us from observables to unobservables since the premises and the conclusions have to talk about entities of the same kind. The hypothetico-deductive method contains no such constraint. It allows the introduction of hypotheses about unobservables on the requirement that test implications about observables can nonetheless be deduced. The hypotheses about the proton, the electron, and the neutron are classical examples.

In order to make the ensuing presentation brief, we need to increase the level of abstraction. If ‘H’ is a variable for hypotheses, ‘I’ a variable for

initial conditions, and ‘T’ for test implications, then the hypothetico-deductive schema can be presented either as in (a) or as in (b) below:

(a)		(b)
	H	
	I	
hence:	----	necessarily, if (H and I) then T
	T	

We can now abstractly examine what happens both when T for empirical reasons is true and when it is false. To start with, we assume that I (= the conjunction of the initial conditions) is beyond reasonable doubt true. Then, if T (our test implication) is false, for reasons of deductive logic we have to conclude that H (our hypothesis) is false, too. The inference can be presented as a simple case of modus tollens if the initial conditions (which are assumed to be true) are left out of account:

premise 1:	*if* H, *then* T
premise 2:	*not*-T
hence:	----------
conclusion:	*not*-H

In this situation, we can truly say that the test implication *falsifies* the hypothesis. What then to say about the case when T is true?

premise 1:	*if* H, *then* T
premise 2:	T
hence:	----------
conclusion:	?

This is an abstract form of the metal-expands-when-heated case already discussed. That is, deduction cannot here tell us whether H is true or false. From the truth only of the premises we cannot *deduce* anything about the truth or falsity of H (the schema is a case of the invalid *inverse* modus ponens). Briefly put: the fact that a hypothesis is not falsified by a certain test does not imply that it is true.

If the truth-values of the initial conditions and the test implication are easy to determine, then there is a logical asymmetry between the verification and the falsification of a hypothesis. On the assumptions stated, *a true generalization (one not confined to a finite number of cases) might be falsified but never verified*. However, the methodological importance of this logical asymmetry becomes almost negligible when the complications of real research are taken into account. In actual science, it is hardly ever the case that test implications can be deduced only from the hypothesis under consideration and some relevant initial conditions. Many *auxiliary hypotheses* are also needed. In order to make our abstract hypothetico-deductive schema come a bit closer to real research, it has to be expanded into the one below (H_{An} is short for 'auxiliary hypothesis number n' and I_{An} is short for 'initial conditions for auxiliary hypothesis number n'):

hypotheses:	H and H_{A1} and H_{A2} and H_{A3} and … H_{An}
initial conditions:	I and I_{A1} and I_{A2} and I_{A3} and … I_{An}
hence:	----------
test implication:	T

If this T is false (and, again, all the initial conditions are true), it follows deductively only that the *whole conjunction* 'H and H_{A1} and H_{A2} and H_{A3} and … H_{An}' has to be false; and so it is as soon as only one of these Hs is false. Why, we might ask with good reasons, blame H when there are n auxiliary hypotheses that can be blamed too? As the hypothetico-deductive method is traditionally described, it contains no method of choice for this kind of situation. And, from our presentation of the concept of paradigms (Chapter 2.4), it follows that there can be no simple such method.

In some cases, even the last and complex hypothetico-deductive schema has to be made more complex. Assume that we want to test the hypothesis that metals expand when heated, and that we have a simple case with no auxiliary hypotheses and no doubtful initial conditions:

hypothesis:	always, if a piece of metal is heated, then it expands
initial condition:	this piece of metal is heated
hence:	----------
test implication:	this piece of metal expands

Assume now that the heated piece of metal referred to is situated in a clamping device that does not permit it to expand. That is, the metal piece is heated but it does not expand. The test implication is false and the initial condition is true, but is the hypothesis thereby falsified? If the schema captures the whole of the experimental situation then, for deductive reasons, the hypothesis has to be regarded as false. But this seems odd. Better is to interpret the hypothesis as saying ‘always, if a piece of metal is heated, it has a *tendency* to expand’. The absence of expansion can then be explained as being due to the counteracting forces of the clamping device. There exists, as predicted, a tendency, but it cannot be *realized* because of counteracting tendencies.

Assume that you have two conflicting desires. For instance, you want to sit down in order to read a book, but you would also like to take a walk. In this situation, your two behavioral tendencies may for a while counteract each other in such a way that you just stand still; and this in spite of the fact that you have no rock-bottom wish to stand still. In the way now explained, tendencies can exist unrealized in both persons and inanimate nature. Hypotheses that predict such tendencies to arise can only be proven false under a very special assumption. One that claims: no other factors than those stated in the auxiliary hypotheses (H_{An}) can here make counteracting or reinforcing tendencies arise. Such an assumption is not an ordinary generalization, since it has not the form ‘all A are B’; nor is it an initial condition for an auxiliary hypothesis (I_{An}), since it has not the form ‘this is A’ or ‘this is B’. It has been called both ‘provisoe’ and ‘closure clause’, and it has the form ‘there are no more relevant factors in the situation at hand’. That such an auxiliary clause is true is in many cases just as uncertain as the hypothesis and the auxiliary hypotheses on trial.

In Newtonian mechanics, the need for provisoes is more easily seen than in most other theories. Apart from the three laws of motion and the law of gravitation, this theory contains an explicit *superposition principle for*

forces; it says that when making a calculation or an experiment, then all forces at play have to be added. But since the principle does not tell how many forces there are, one has on each and every occasion to ‘close it’ by means of an auxiliary (closure) clause that says: ‘no more forces are relevant’.

Normally, when there is talk about unobservable entities in science, one refers to entities that are either too small (sub-molecular entities) or too far away (extremely distant galaxies) to be able to be observed through microscopes or telescopes, respectively. In the same way, tendencies that exist unrealized in inanimate nature are also unobservable, but for another reason, namely that they can exist unrealized.

Tendencies are posited in the medical sciences too. Let us give three examples. First, the disorder thrombophila is normally *defined* as a disorder of the hemopoietic system in which there is an increased *tendency* for forming thrombi. Second, there is research around the question whether some people may have inborn or acquired psychopathic *tendencies*. Third, there are good reasons to speak of people’s *tendencies* to react on a certain medical treatment. On one patient a treatment may have a considerable effect but on another none. And this difference may be due to the fact that the latter patient is in a situation where the tendency created by the treatment is counteracted by some other tendency.

The existence of tendencies can explain why a certain treatment might trigger a specific reaction in some patients, but leave others unaffected. Therefore, many medical tests and predictions may require, just like predictions with Newtonian mechanics, an auxiliary assumption to the effect that nothing in the situation at hand counteracts or reinforces the tendency described by the main hypothesis (H). In cases involving tendencies, the abstract hypothetico-deductive schema looks as follows:

hypothesis:	H
auxiliary hypotheses:	H_{A1} , H_{A2} , H_{A3} , … H_{An}
initial conditions:	I , I_{A1} , I_{A2} , I_{A3} , … I_{An}
closure clause:	no other auxiliary hypotheses are relevant
hence:	----------
test implication:	T

If T is true, then everything that is above the inference may be true too, but it need not be. If T is false, then for formal-logical reasons *something* above the inference line has to be false too. But what? There is much to choose between. To make the best choice may require long scientific training, in some cases even a grain of genius. Neither in its original verificationist form nor in its later falsificationist variety (see presentation of Popper in Chapter 3.5) can a hypothetico-deductive argument supply more than inductive support for the hypothesis under test. As we can never be certain that we have verified a theory, we can never be certain that we have falsified it. But we can receive positive and negative empirical evidence, i.e., positive and negative inductive support. Theories are always empirically underdetermined, but seldom completely empirically vacuous.

The general hypothetico-deductive form of argumentation now presented is given a specific form in the randomized controlled trials of medical science; we present it in Chapter 6.3.

4.5 Arguments from simplicity, beauty, and analogy

Even when brought together in hypothetico-deductive arguments, deductive and inductive arguments cannot supply a conclusive verdict on the truth or falsity of empirical hypotheses and theories. This empirical underdetermination implies fallibilism, but not epistemological relativism or social constructivism. Why? Because it admits that nature can make some test implications false and thereby tell us that *something* is wrong. However, this underdetermination explains why scientists sometimes have recourse to arguments from simplicity, beauty, and analogy.

Ernest Rutherford (1871-1937), famous for putting forward the hypothesis that atoms contain electrons that orbit around a central nucleus (the proton), said: “Considering the evidence as a whole, it seems *simplest* to suppose that the atom contains a central charge distributed through a very small volume” (Hanson p. 216). Such considerations can be viewed as a way of making the invalid inverse modus ponens schema a bit more reasonable. Let T represent a description of all the evidence assembled (the test implications found to be true), and let H represent the hypothesis in question. If we leave auxiliary assumptions aside, we arrive at this deductively invalid inference schema:

premise 1: *if* H, *then* T
premise 2: T
hence: ---------- (INVALID)
conclusion: H

Rutherford can be said to have used the next also deductively invalid but more reasonable schema:

premise 1: *if* H, *then* T
premise 2: T
premise 3: H is simpler than the alternatives
hence: ---------- (INVALID)
conclusion: H

In our opinion, such arguments can be allowed to play some role in the short run of research, but in order to be strong they have to be connected to some overarching principle that says ‘nature is simple’. But this creates a new problem: how is such an overarching principle to be epistemologically justified? Because of this predicament, we will only say a few words about what scientists can mean when they speak of simplicity, beauty, and analogies.

In physics, Maxwell’s equations, which describe how properties of electromagnetic waves are related to electric charges and currents, are often described as beautiful, over-archingly symmetrical, and simple. They look as follows:

Maxwell’s equations:

$$\nabla \cdot D = \rho$$
$$\nabla \cdot B = 0$$
$$\nabla \times E = -\partial B/\partial t$$
$$\nabla \times H = J + \partial D/\partial t$$

We agree; there is something appealing in this symbolism even if one does not understand it exactly and, therefore, cannot see its simplicity against the background of the complexity of the reality described.

Sometimes even beauty is given the role that simplicity has above. In such cases we obtain this inference schema:

premise 1:	*if* H, *then* T
premise 2:	T
premise 3:	H is more beautiful than the alternatives
hence:	---------- (INVALID)
conclusion:	H

In the discovery of the DNA molecule, several famous researchers seem to have thought of their discoveries in terms of beauty; often being beautiful means being symmetric. “What a beautiful molecule!”, was the reaction of Rosalind Franklin (1920-1958) when she was the first person to observe the DNA molecule from an X-ray diffraction picture. J. D. Watson (b. 1928) characterized Linus Pauling’s discovery of the structure of proteins as ‘unique and beautiful’. Although they were competitors in the discovery-of-the-DNA race, Linus Pauling had inspired ‘the winners’ Watson and Francis Crick (1916-2004) both in their puzzle-solving strategy and in their research outlook: the truth should be recognized as simple, straightforward, and beautiful. Watson and Crick succeeded, and most people seem to find the outlines of double helix model beautiful (see Figure 1).

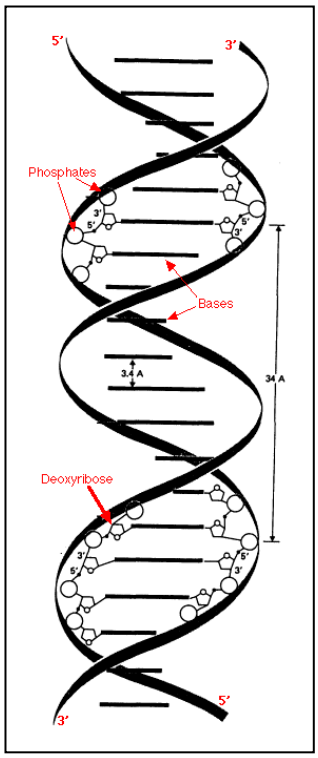


Figure 1: *The DNA model.*

Let us now turn to arguments from analogy. To say that *A* and *B* are analogous is sometimes just to say that they are similar, but often something more complex is meant, namely that '*A*s are to *C*s what *B*s are to *D*s'. To claim that handlebars are analogous to rudders is to claim that handlebars (*A*) are to bicycles (*C*) what rudders (*B*) are to ships (*D*). We will start with simple similarity. Once again, we will quote from *Alice in Wonderland* – 'A Mad Tea-party'.

> - I'm glad they've begun asking riddles. I believe I can guess that, said Alice aloud.
> - Do you mean that you think you can find out the answer to it? said the March Hare.
> - Exactly so, said Alice.
> - Then you should say what you mean, the March Hare went on.
> - I do, Alice hastily replied; at least - at least I mean what I say – that's the same thing, you know.
> - Not the same thing a bit! said the Hatter. You might just as well say that 'I see what I eat' is the same thing as 'I eat what I see!
> - You might just as well say, added the March Hare, that 'I like what I get' is the same thing as 'I get what I like'!
> - You might just as well say, added the Dormouse, who seemed to be talking in his sleep, that 'I breathe when I sleep' is the same thing as 'I sleep when I breathe'!'

Alice maintains that 'I say what I mean' is synonymous or equivalent to 'I mean what I say':

I say what I mean (= if I mean something then I say it
hence: ---------- (ANALOGY)
I mean what I say (= if I say something then I mean it)

Her friends then retort that to make such a claim is as odd as to argue as follows:

	the Hatter	the March Hare	the Dormouse
	I see what I eat	I like what I get	I breathe when I sleep
hence:	----------	----------	---------- (ANALOGY)
	I eat what I see	I get what I like	I sleep when I breathe

All these arguments can in propositional logic be given the invalid form of 'implication-equivalence conflation' (premises: '*if* I eat something (p) *then* I see it (q)', '*if* I get something *then* I like it', '*if* I sleep *then* I breathe', respectively):

premise:	*if* p, *then* q
hence:	---------- (INVALID)
conclusion:	*if* q, *then* p

This invalid inference might be strengthened by adding an analogy premise; this premise adds *something*, but it does not make the inference deductively valid.

premise:	*if* p, *then* q
premise:	in p and q, analogous things are talked about
hence:	---------- (INVALID)
conclusion:	*if* q, *then* p

In the cases of Alice and her friends, however, there is no similarity between what is talked about in p (in turn: saying, seeing, liking, and breathing) and in q (in turn: meaning, eating, getting, and sleeping). There is only a similarity between the forms 'if p then q' and 'if q then p'. Let us now look at the following inference schema (from *X* to *A*):

	there are *X*s that are to *C*s what *B*s are to *D*s
hence:	---------- (ANALOGY)
	*A*s are to *C*s what *B*s are to *D*s

A good example of this schema is related to Darwin's discoveries; it looks as follows:

there is something (*X*) that is
to the development of wild animals (*C*)
what breeding (*B*) is
to the development of domestic animals (*D*)
hence: ---------- (ANALOGY)
natural selection (*A*) is to *C* what *B* is to *D*

Questions such as 'What (*X*) is to the foot (*C*), what the palm (*B*) is to the hand (*D*)?' are sometimes used to test the intellectual development of children. It is an intellectual achievement to be able to recognize similarities such as these. But, of course, as human beings we are also able to project onto the world resemblances that are not there. However, after the fallibilistic revolution one can well let arguments by analogy play some role in the context of justification as well as in the context of discovery. What role they will come to play in computer science is at the moment hard to tell, but since many decades there are in AI analogy-finding programs.

When we mentioned Pasteur's discovery that bacteria can cause infections (Chapter 2.4), we also mentioned that part of the process that led to this discovery was Pasteur's earlier discovery that yeast cells are not an accidental product of fermentation processes but the cause of them. In analogy with this reversal, he reversed also the view that microorganisms in infected wounds are accidental by-products into the view that, in fact, they cause the infection. We can now see the structure of the argumentation more clearly:

there is something (*X*) that is to infections (*C*)
what yeast cells (*B*) are to fermentations (*D*)
hence: ---------- (ANALOGY)
bacteria (*A*) are to *C*s what *B*s are to *D*s

We will end this section with a little story; be it a myth or not. The surgeon who started to treat burn injuries with cold water discovered the

treatment by means of an analogical inference made when he had breakfast. (Prior to this discovery, one avoided to pour water on burn injuries in order to minimize the risks of infection.) His wife normally served him a soft-boiled egg. One morning, however, she forgot to rinse the cooked hot egg in cold water. A few minutes later, when the surgeon started to eat the egg, it was hard. He asked his wife why the egg was hard-boiled, and she explained what had happened. On a little reflection, the surgeon came to the conclusion that cold-water cooling of eggs inhibits an ongoing coagulation process that is caused by the heat from the boiling water. Further reflections, made him quickly believe that cold water can inhibit coagulation processes in all albumins. Then, thirdly, by means of an analogical inference he reached the conclusion that what cold water is to egg coagulation is probably to blood coagulation too. The form of this inference is a bit simpler than the earlier one, for it has the form: what *A* is to *C*, must be to *D* as well.

4.6 Abductive reasoning and inferences to the best explanation

In deductions, one is moving from premises to conclusions in such a way that *if* the premises are true then the truth of the conclusions is guaranteed. In pure inductions and in hypothetico-deductive arguments, not to speak of arguments from simplicity, beauty, or analogy, this is not the case. Here, true premises can only make the conclusion more or less epistemically probable. This is true also of the next kind of reasoning to be presented: *abduction*. What we have already said about the formal inference schemas for induction can now be repeated for abduction: some epistemic probability has in each concrete case to come also from intuitions connected to the substantive content displayed in the premises and the conclusion; here, so-called tacit knowledge (Chapter 5) may play a part. Other names for this kind of reasoning are ‘retroduction’ and ‘inference to the best explanation’.

In our exposition of abduction, we will develop thoughts from C. S. Peirce and N. R. Hanson. We will distinguish two forms of abduction:

- abduction to a known kind
- abduction to an unknown kind.

The inference schema for *abduction to a known kind* has, in Aristotelian term logic, a formal structure distinct from, but similar to, schemas for deductions and inductive generalizations that involve the same three sentences:

	Deduction	Induction	Abduction
premise 1:	all D are S	*a* is D	all D are S
premise 2:	*a* is D	*a* is S	*a* is S
hence:	----------	----------	----------
conclusion:	*a* is S	all D are S	*a* is D

This form of abductive reasoning has the deductively invalid form that we have called ‘inverse modus ponens’. This means that abductions rely for their kind of validity on the particular content of each case, not on the general form displayed above. Even if the very concept of abduction is unknown to clinicians, abductions play a large role in their life. The inference schema below exemplifies the form for abduction above (note that logic has to abstract away differences between verbs such as ‘is’ and ‘have’):

premise 1:	all persons having disease D have symptoms S
premise 2:	patient *a* has symptoms S
hence:	---------- (ABDUCTION)
conclusion:	patient *a* has disease D

Even if patient *a* has all the symptoms S of D, he may not suffer from D. To make a diagnosis is not to make a deductive inference; nor is it to make an inductive generalization. It is to make an abduction. In order to be able to make such inferences, clinicians need practical experience or what has been labeled ‘tacit knowledge’, but now we merely want to show that its formal structure is distinct from those of deduction and induction. The schema above explains why abduction is also called ‘retroduction’ and ‘inference to the best explanation’. When a clinician diagnoses someone

(*a*) as having a disease (D), he argues 'backwards' from the symptoms (the effects) to the disease (the cause), i.e., from the symptoms he does not *deduce* the fact that there is a disease, he *retroduces* it. Based both on the disease-symptoms generalization and his practical skills, he makes 'an inference to (what he thinks is) the best explanation' of the symptoms in the case at hand.

As an example of *abduction to an unknown kind*, we will use what both Peirce and Hanson thought of as "the greatest piece of Retroductive reasoning ever performed (Hanson p. 85, Peirce p. 156)", namely Johannes Kepler's discovery that the planets do not have circular but elliptical orbits. When one already knows that *all* planets move in ellipses, one can *deduce* that a certain planet moves in an ellipse:

premise 1:	all planets move in ellipses	(all D are S)
premise 2:	Mars is a planet	(*a* is S)
hence:	----------	
conclusion:	Mars moves in an ellipse	(*a* is D)

When one knows only that *some* (say, three) planets move in ellipses, one can try the following *inductive* inference:

premise 1:	these three bodies are planets	(*a*, *b*, *c* are D)
premise 2:	these three bodies move in ellipses	(*a*, *b*, *c* are S)
hence:	----------	(INDUCTION)
conclusion:	all planets moves in ellipses	(all D are S)

Kepler could, when he started his work, make none of these inferences; nor the kind of abduction we have already presented. Why? In all these schemas the concept of an elliptical orbit figures in one of the premises, but neither Kepler nor anyone else had then entertained the idea that planets may move in ellipses. Rather, Kepler's way to the conclusion 'all planets move in ellipses' should be seen as consisting of the following two steps, the first being the truly abductive step:

these are the plotted positions for Mars (the planet) this year
hence (1): ---------- (ABDUCTION)
Mars (the planet) has moved in an ellipse
hence (2): ---------- (INDUCTION)
all the planets move in ellipses

The second step came for Kepler automatically, since he simply took it for granted that he investigated Mars as a representative for all the six planets then known. Otherwise, this step has to take the form of an inductive generalization from *one* or *some* planets to *all* planets. Let us focus on the first step.

Mars' orbit is not directly observable from the earth. At the times of Kepler, the astronomers related the places the planets occupied at different times *to the stars as seen from the earth*. Since the earth itself moves, it is no simple task to extract from such 'inside' representations the corresponding orbits as (let us say) seen from most places outside the solar system. But things were even harder. Kepler had to free his mind from the view that – *of course* – planets move around *one* centre; ellipses have two. Therefore, it was quite a feat to be able to abduct ellipses from the data available. The most prominent astronomer at the time, Tycho Brahe (1546-1601), for whom Kepler worked as an assistant a short time, and whose data Kepler could use, was not able to do it. Before Kepler no one seems to have even suspected that the planets could be moving in ellipses.

Abductions to an unknown kind have quite a place in the history of medicine. Often, the discovery of distinct diseases such as Alzheimer's and Parkinson's has consisted in an abduction from symptom data to a disease. For example, what Brahe's data was for Kepler in relation to the elliptic orbit of Mars, the symptoms of Alzheimer's disease was for Emil Kraepelin (1856-1926) in relation to the disease as such. This is abduction at the observational level. To find unobservable causes of diseases already observationally identified, a second abduction is needed. In relation to Alzheimer's disease, this work was done by the man after whom the disease is now named, Aloysius Alzheimer (1864-1915).

In connection with abduction, some words about so-called *ex juvantibus* treatments and diagnoses are needed, too.

According to some medical norms, physicians are always supposed to present a specific diagnosis before they start a treatment, but this is simply not always possible. Here is an example. According to the theory of deficiency diseases, a treatment with the relevant deficient substance will cure a patient. Let us take the case of iron deficiency based anemia. Sometimes a patient suffering from anemia has to be treated before it can be established what specific kind of anemia he has. If then, the anemia patient is treated with iron just provisionally, i.e., *ex juvantibus*, this choice of treatment does not deserve the name abduction. However, if in fact the patient becomes cured, and the physician therefore in retrospect makes the diagnosis that the patient actually suffered from iron-deficiency anemia, then this ex-juvantibus-treatment-based *diagnosis* is an abduction (to a known kind). Schematically:

premise 1:	if patients with iron deficiency anemia are treated with iron, they will be cured
premise 2:	this patient had anemia, was treated with iron, and was cured
hence:	---------- (ABDUCTION)
conclusion:	this patient suffered from iron deficiency based anemia

Our general points about abduction in relation to medical science can be put as follows:

(a) all the first physicians to see true disease patterns (where others saw only disconnected illnesses or symptoms) made *abductions to unknown (disease) kinds*;

(b) all the physicians who every working day (based on some known disease-symptom correlation) diagnose patients as having known diseases are making *abductions to known (disease) kinds*.

4.7 Probabilistic inferences

Due to considerations of many different kinds – metaphysical (see Chapter 2), logical (see the sections above), technological (all measuring devices are somewhat uncertain), and ethical (see Chapters 9 and 10) – it is impossible in medical research to control all the factors that affect the outcomes of laboratory experiments, clinical research, and epidemiological

investigations. Therefore, normally medical researchers have to talk in terms of probabilities when they present their results, be it that a certain disease or illness has a specific cause, that a certain treatment has a good effect, or that a certain kind of situation or habit contains a high risk for a disease. This fact, in turn, makes it important for everyone interested in medicine to have some understanding of some of the features that are specific to probability statements and to probabilistic inferences. Especially for clinicians, it is good to become aware of the complexity that surrounds probabilities for singular cases.

4.7.1 Four kinds of probabilistic inferences and four kinds of probability statements

For the philosophical-scientific purposes of this book, four kinds of *probabilistic inferences* and four kinds of *probability statements* have to be kept distinct. A probabilistic inference is an inference that involves at least one probability statement. Most of such inferences conform to the patterns of deduction, induction, and abduction already distinguished, but there are also other some patterns that have to be presented separately; they have to do with the fact that there are several kinds of probability statements. Therefore, we will divide *probabilistic inferences* into:

- deductive probabilistic inferences
- inductive probabilistic inferences
- abductive probabilistic inferences
- cross-over probabilistic inferences

while the *probability statements* themselves will be divided into:

- purely mathematical
- frequency-objective
- singular-objective
- epistemic (subjective).

All statements in the *calculus of probability* are as purely mathematical as each statement in the multiplication table is. Therefore, they are not

subject to empirical testing. However, all the other kinds of probability statements are empirically true or false. The most conspicuous objective probability statements are those describing relative frequencies, for example, the relative frequencies of heads and tails in coin tossing. We call them frequency-objective probability statements in contradistinction to singular-objective ones, which claim that there is an objective probability in a singular event or situation as such. Epistemic (or 'epistemological') probability statements are about how certain a person can be that a belief or an assertion is true, i.e., epistemic probability statements make claims about the *relationship* between knowledge subjects and the world. Therefore, they may also be called 'subjective' probability statements. Objective (or 'ontological') probability statements, on the other hand, make direct claims (true or not) about how *the world in itself* is in some respect.

The difference between objective and epistemic probability statements is most easily seen in relation to universal statements such as 'all men are mortal' and 'all metal pieces expand when heated'. Assertions such as '*probably*, all men are mortal' and '*probably*, all metal pieces expand when heated' are not meant to say that the relative frequency of mortals among men, and the relative frequency of expanding metal pieces among heated metal pieces, is high. They are meant to say that it is *probably true* (epistemology) that all men are mortal, and that it is *probably true* that all metal pieces expand when heated, respectively.

Note that it can be extremely hard, and often impossible, to transform with good reasons purely qualitative epistemic probability statements such as 'probably, all metal pieces expand when heated' into numerical epistemic probability statements such as 'with a probability of 0.8 it is true, all metal pieces expand when heated'.

An assertion such as 'probably, we will be visited by Joan on Friday' is out of context ambiguous. Depending on context, it can state either an epistemic probability or a frequency-objective probability. In one context it can mean that the speaker claims to have reasons (epistemology) to think that Joan, as it happens, will visit on the upcoming Friday, but in another it can mean that the speaker states that Joan usually visits on Fridays. In some cases, however, it is hard to distinguish between objective and epistemic probability statements. We will comment on this problem later

when we present cross-over inferences (4.7.5). To begin with, we will explain in more detail the difference between frequency-objective and singular-objective probability statements.

4.7.2 The die, the voter, and the medically ill-fated

If a die that is made of a very hard material is thrown 60 million times, each of the sides comes up with a relative frequency of approximately 1/6, i.e., 10 million times each. If 60 million such dice are thrown once simultaneously, each side will again come up with a relative frequency of approximately 1/6. Think next of the following situation. There is a nation where 60 million citizens are allowed to vote, all of them do in fact vote, and none votes blank. There are six parties, and, according to a scientific investigation, each party will in the upcoming election get 10 million votes each. That is, the relative frequency of voters for each party is expected to be 1/6. Let us now compare the statements

- the probability that *this* throw of *this* die will yield a side-2 is 1/6
- the probability that *this* vote of *this* voter will be for the so-called party-2 is 1/6'.

Both these statements are, no doubt, *formally* singular-objective, i.e., from a purely grammatical point of view they talk about a singular throw and a singular vote, respectively. However, everyday language allows what is formally singular to be short-hand for what in fact is substantially non-singular. For instance, 'Smith has 1.9 children' is formally singular but it makes no sense as a substantial singular-objective statement, and it means that the average number of children among couples in Smith's country is 1.9. Similarly, the singular-objective statement 'the probability that *this* throw of this die will yield a side-2 is 1/6' *can* be short-hand for the frequency-objective statement 'the relative frequency of side-2 results is 1/6', and the statement 'the probability that *this* vote of this voter will be for the party is 1/6' *can* be short-hand for the frequency-objective statement 'the relative frequency of party-2 voters is 1/6'.

Now, let it be perfectly clear, our concept of 'singular-objective probability statements' is not meant to cover such cases of frequency statements in disguise. When we are talking about singular-objective

probability statements, we are talking about statements that *literally* ascribe a probability to a singular event. Hereby, we are rejecting the positivist rejection of the concept of causality (Chapter 3.4); in relation to the interpretation of singular-objective probability statements, the overlap between science and philosophy becomes obvious. That there is a distinction between merely formal and substantive singular-objective statements can be made clear in several ways.

Assume, as before, that the relative frequency of party-2 voters is 1/6 in the community as a whole, but that for people more than 50 years old it is 1/4 and for males 1/5. This means that for a male person over 50 (call him John), *all* the three following statements are in one sense true:

(a) 'the probability that John will vote for the party-2 is 1/6'
(b) 'the probability that John will vote for the party-2 is 1/4'
(c) 'the probability that John will vote for the party-2 is 1/5'.

But John cannot have all three probabilities as real singular-objective probabilities. This would be like weighing 75, 70, and 65 kg simultaneously. The three statements discussed have to be regarded as being merely short-hands for the corresponding three relative frequencies that are about (a) citizens in general, (b) citizens older than 50, and (c) male citizens, respectively. (And, according to positivism, this is all there is to be said.)

The distinction between singular-objective and frequency-objective statements concerned with voting behavior can also be highlighted in the following way. Obviously, a certain voter can hate all what party-2 represents. Therefore, there is no chance whatsoever that he will vote for it, i.e., the singular-objective statement 'the probability that this voter (v_a) will vote for the party-2 is 0' is true. Another voter can be a fanatic party-2 member. He will vote for this party whatever happens, i.e., the singular-objective statement 'the probability that this voter (v_b) will vote for the party-2 is 1' is true. All cases in-between probability 0 and 1 are also possible. In particular, there may be a voter that is searching for reasons to prefer one single party, but at the moment thinks that there are equally strong reasons for and against all the six parties. Since he (v_3) thinks that as a citizen he nonetheless has a duty to vote, it holds true of him: 'the

probability that this voter (v_3) will vote for the party-2 is 1/6'. He is equally strongly torn in six different directions, i.e., he has six equally strong tendencies, but he can realize one and only one. He can grab a vote blindfolded, or let a random number generator make his choice.

With respect to each of the 60 million dice the situation is quite different. No die is pre-determined to come up with a side-2, and no die is pre-determined *not* to come up with a side-2. Why? Because all the dice are perfectly symmetrical. What can make a difference is the way the die is thrown, not the die itself. But since there are an infinite number of different such ways, it can be assumed that the ways the throws are made are as if they were chosen by a random number generator. If the dice are looked upon as human beings with inner reasons to act, then all the dice would be in the predicament of the voter that had equally good reason to vote for six different parties. Both such a voter and the dice *tend* in six different directions equally; the voter because his reasons for acting are symmetrical, the dice because its material structure is symmetrical.

What about singular persons in relation to medical probability statements? Well, unfortunately we have no definite answer, and we will only make some remarks hoping that future philosophy and science can make this area more clear. In our opinion, the only thing that at the moment can be said with certainty is that it is often hard to understand whether singular-objective medical probability statements are merely meant as short-hands for frequency-objective statements or as being literally meant. In the latter case, there ought to be some kind of similarity with dice, which inherit a definite singular-objective probability from their material structure. Of course, one should then think of a very asymmetrical die.

Let us assume that in a certain community, as far back as statistics are available, 1/6 of all the citizens between 45 and 65, the middle-aged ones, is infected by what is called 'disease-2'. One might then say about a particular such citizen John:

1. 'John runs a risk of 1/6 to get disease-2'.

However, if no more is said this statement has to be interpreted as being only formally singular-objective and short-hand for the mentioned frequency-objective statement:

1a. ‘1/6 of all the middle-aged persons, of which John is one, get disease-2’.

Despite the fact that statement 1a is true, John may run no risk at all to get the disease; he may be completely immune to it (compare with the voter that hates party-2). Conversely, he may be pre-determined to get the disease (compare with the fanatic party-2 voter). Both these states of affairs are quite consistent with the frequency-objective statement being true.

If the concept of tendency is allowed to be introduced, then statement 1 can be interpreted as a claim that each and every singular middle-aged person, John included, in fact has a *tendency* to contract or develop disease-2, but that there are unknown factors that prevent the realization of the tendency in five out of six cases. We should write:

1b. ‘John has a tendency and real risk to get disease-2, but for some reason there is only a probability of 1/6 that the tendency will be realized’.

Now the singular-objective statement ‘John runs a risk of 1/6 to get disease-2’ ascribes a real property, a tendency, to John, but it allows no sure prediction of illness to be made.

Often in medical science, knowledge of one relative frequency or correlation triggers the search for an even more significant correlation. Let us introduce such a move in our probabilistic thought experiment. Here is a list of usual risk factors in epidemiological research: age, sex, race, ethnicity, family history, workplace chemicals, smoking habits, drinking habits, diet, weight, waist size, blood type, blood pressure, blood sugar, and cholesterol. Assume that the next investigation shows disease-2 to correlate with, say, workplace chemicals in such a way that people exposed to the chemical called ‘Ch’ for ten years are said to run a risk of 4/6 to get

disease-2. What are we to say about this new probability statement? As far as we can see, the old interpretative problem simply reappears, even though now connected to a new factor with a higher probability. We arrive at the following three statements:

2. ‘John runs a risk of 4/6 to get disease-2’
2a. ‘4/6 of all persons exposed to Ch, of which John is one, get disease-2’
2b. ‘John has a tendency and real risk to get disease-2, but for some reason there is only a probability of 4/6 that the tendency will be realized’.

One may search for still another new factor or combination of factors, but for every risk statement that contains a probability smaller than one, the problem of how to interpret the corresponding singular-objective probability statements will reappear. No relative frequency and correlation smaller than one can imply that a singular-objective risk statement is literally true.

We now will turn to the four different kinds of probabilistic *inferences* that we want the reader to become able to distinguish.

4.7.3 Deductive probabilistic inferences

If in a deductively *valid* schema like that to the left below, one exchanges the general premise for a corresponding premise about a relative frequency, then one obtains the deductively *invalid* schema to the right. Sometimes the latter is called a statistical syllogism (the term ‘syllogism’ is often used in philosophy as a synonym for ‘valid inference’), but we will call it an inductive probabilistic inference. Instead of the usual inductive form, ‘from *some(-and-perhaps-all)* to *the next*’, it has the form ‘from *some(-but-not-all)* to *the next*’. Here are the schemas:

premise 1:	all human beings are mortal	70% of all human beings are mortal
premise 2:	Socrates is a human being	Socrates is a human being
hence:	----------	---------- (INDUCTION)
conclusion:	Socrates is mortal	Socrates is mortal

This does not mean that there are no probabilistic deductions. Mathematical inferences are deductive, and inferences made by means of the mathematical *calculus of probability* are proper deductions. Of course, as always in deduction, when it comes to evaluating the *truth-value of the conclusion*, one has to keep in mind what truth-value the premises have. Here is an example of a simple deductive probabilistic inference concerned with purely mathematical probability statements (but with one premise so far hidden):

$$\begin{aligned} & P(A) = 0.6 \\ & P(B) = 0.4 \\ \text{hence: } & \text{----------} \\ & P(A \cdot B) = 0.24 \end{aligned}$$

It should be read: if the probability for an event of type A is 0.6, and the probability for an event of type B is 0.4, then the probability for both events to occur is 0.24. If there is a chance of 0.6 to win in one lottery and a chance of 0.4 to win in another lottery, then there is a chance of 0.24 to win in both lotteries. However, as the inference explicitly stands, it is not deductively valid. It is so valid only on the condition that the probability stated for A is independent of the probability stated for B, and vice versa. This requirement is part of the mathematical probability calculus, and it is mathematically expressed as ‘$P(A|B) = P(A)$’. It could just as well have been written ‘$P(B|A) = P(B)$’; these two equalities imply each other. The formula ‘$P(A|B) = P(A)$’ should be read:

- the (relative) probability of an event A given an event B equals the (absolute) probability of an event A.

It states that the probability of an event of type A remains the same whether or not an event of type B has occurred or obtains. The following probabilistic inference is without qualifications deductively valid:

$$
\begin{aligned}
& P(A) = 0.6 \\
& P(B) = 0.4 \\
& P(A|B) = P(A) \\
\text{hence: } & \text{----------} \\
& P(A \cdot B) = 0.24
\end{aligned}
$$

How shall this inference be looked upon if ‘P(A) = 0.6’, ‘P(B) = 0.4’, and ‘P(A|B) = P(A)’ are not purely mathematical but *objective* probability statements? Answer: the inference is still valid. If the premises states truths about some relative frequencies, then the conclusion states a truth about a relative frequency too. However, since the premises are empirical statements they might be false and, therefore, one may feel uncertain about the truth of the conclusion. Deductions are, as we have said, truth-preserving inferences, but if one is uncertain about the truths of the premises, such inferences may be called ‘uncertainty-preserving’. A deductively valid inference cannot reduce the epistemic uncertainty of the premises.

What then if ‘P(A) = 0.6’, ‘P(B) = 0.4’, and ‘P(A|B) = P(A)’ are *subjective* (epistemic) probability statements? In one sense, what has just been said about objective probabilities applies to subjective probabilities too. However, two things have to be noted. First, the probability calculus is about numerical probabilities, but, normally, it is difficult, not to say impossible, to ascribe numerical values to epistemic statements. In many situations, it is easy to put forward vague but true subjective probability statements such as ‘probably, it will soon start to rain’, but senseless to state ‘P(it will soon start to rain) = 0.6’; or some other definite number.

Second, deductive inferences can only relate sentences that express something that can be regarded as true or false. This means that if an epistemic assertion such as ‘probably, all men are mortal’ is put into a deductive inference, then it has to be read as saying ‘it is true: probably, all

men are mortal'. This is not odd. To assert simply 'all men are mortal' is also to assert, but implicitly, 'it is true: all men are mortal'. Usually, this implicit level is made explicit only when the truth of an assertion is questioned and people are forced to reflect on and *talk about* their assertions.

Some illustrating words now about deductive inferences in relation to the classical example of relative frequencies: playing dice. What is the probability of getting two '5' when throwing a die twice? As premises, we assume that the probability to get '5' on the first throw, A_1, is one sixth, and the same holds true for the second throw, A_2. We have:

- $P(A_1) = 1/6$, and $P(A_2) = 1/6$.

If the probability calculus is applied, we can deductively infer that the probability of getting two '5' in succession is 1/36:

- $P(A_1 \cdot A_2) = P(A_1) \cdot P(A_2) = 1/36$.

Since the general (objective) probability statement 'P(A) = 1/6' is an empirical statement about the die that one starts to use, one can never know for sure that it is true. Furthermore, even if the die is absolutely symmetric and has the probability one sixth for each side when one starts to throw, it is possible that the first throw changes the symmetry and the probabilities of the die and, thereby, makes '$P(A_2) = 1/6$' false. Another way of expressing this last possibility is to say that the probability for the throw A_2 may *depend on* A_1. If one of the premises '$P(A_1) = 1/6$' and '$P(A_2) = 1/6$' is false, then the conclusion '$P(A_1 \cdot A_2) = 1/36$' need not be true, but the probabilistic *inference* in question is nonetheless deductively *valid*.

Singular-objective probability statements such as 'the probability of getting a five (A) in the next throw of this die is 1/6' ('P(A) = 1/6') can just like many other empirical hypotheses be argued for by means of hypothetico-deductive arguments. The reason is that from such a singular-objective statement one can (if the number n is large) deduce a relative frequency:

hypothesis:	P(A) = 1/6
initial condition 1:	die D is thrown *n* times
initial condition 2:	during the test, D does not change any relevant properties
hence:	----------
test implication:	approximately, $n \cdot (1/6)$ of the throws will yield A

As we have explained earlier, such a deductive schema has to be complemented with an inductive inference that goes in the opposite direction, i.e., from the test implication and the initial conditions to the hypothesis; because from a relative frequency (the test implication) one cannot deduce a singular-objective probability (the hypothesis). Let us now proceed to induction in probabilistic contexts.

4.7.4 Inductive probabilistic inferences

We have distinguished between two kinds of induction: 'from *some* to *the next*' and 'from *some* to *all*'. Both have in their own way probabilistic versions. In relation to the first kind, one version has the form 'from *some-but-not-all* to *the next*' or 'from *x%* to *the next*', as below:

	deduction	probabilistic induction
premise 1:	all persons having disease D have symptoms S	x% of the persons having disease D have symptoms S
premise 2:	patient *a* has disease D	patient *a* has disease D
hence:	----------	----------
conclusion:	patient *a* has symptom S	patient *a* has symptoms S

If x is a small number, the inductive conclusion would rather be 'patient *a* does not have symptoms S'.

The next two schemas represent other versions of inductive probabilistic inferences. In the first, the conclusion is expressed by a singular-objective probability statement; in the second, it is expressed by a frequency-objective statement. It has to be added that 'some (= *n*)' now has to mean 'many'; the number *n* has to be large. Since the inferences are inductive there is in neither case any truth-preserving, i.e., the frequency-objective

probability statement of premise 1 can in both cases only give inductive support to the conclusion:

premise 1:	1/6 (approximately) of the *n* throws with the die D have yielded A
premise 2:	during the test, the die D did not change any relevant properties
hence:	---------- (INDUCTION)
conclusion:	in the next throw of D: P(A) = 1/6

An inductive generalization of dice throwing has to go in two directions. It has to generalize to all throws of the die spoken of and it has to generalize to all dice of the same kind. We then have this schema:

premise 1:	1/6 (approximately) of the *n* throws with the dice have yielded A
premise 2:	during the test, the dices did not change any relevant properties
hence:	---------- (INDUCTION)
conclusion:	in any large number of throws, approximately 1/6 will yield A

In statistics, one talks about inferences from a random sample to a whole population. In such terminology, the first premise in the last schema describes a sample and the conclusion describes the corresponding population. This population is in principle infinitely large, as they are in all probabilistic natural laws, but populations can just as well be finite. In either case, however, statistical data-material can provide no more than inductive support to the relative frequencies in the populations. The fact that the statistical move from sample to population is not deductive and truth-preserving is sometimes overlooked. One reason is that it is quite possible to use the hypothetico-deductive method in actual research without keeping the deductive and the inductive inferences explicitly separated in the manner we are doing in this chapter. What in fact is a mix of induction and deduction and, therefore, contains the uncertainty of

induction, may then falsely appear as having the certainty of a pure deduction.

Another reason that the statistical move from sample to population is falsely regarded as deductive might be that some people who use probability statements do not distinguish between substantial singular-objective probability statements and merely formal such statements. We explained this distinction in Section 4.7.2, but let us repeat the point with a new example:

premise 1: 2% of all delivering women at hospital H have been infected with childbed fever
premise 2: the woman W is going to deliver at H
hence: ----------
conclusion: the probability (risk) that W might be infected by childbed fever is 2%

The conclusion allows two different interpretations, one which turns the preceding inference into a deduction, and one which keeps it inductive. On the one hand, the conclusion can be interpreted as another way of stating *only and exactly* what is already said in the premises. While it sounds as if there is talk only about the next patient (singular-objective statement), in fact, there is talk only about a group of earlier patients (frequency-objective statement). Such an interpretation of the conclusion makes of course the inference in question deductively valid, but also uninformative.

On the next interpretation, whereas the first premise (as before) says something about the past, the conclusion says something about the future. The conclusion is a singular prediction, and it may turn out to be false even if the premises are true. That is, W may in fact be immune to childbed fever and have no risk at all to become infected.

4.7.5 Abductive probabilistic inferences

We have distinguished between two kinds of abduction: to a known and to an unknown kind, respectively. Both have probabilistic versions. If we replace the general premise of an ordinary clinical abduction to a known kind by a corresponding premise about a relative frequency, we arrive at the following probabilistic abduction:

	Prototypical abduction	Probabilistic abduction
premise 1:	all persons having disease D have symptoms S	x% of the persons having disease D have symptoms S
premise 2:	patient *a* has symptoms S	patient *a* has symptoms S
hence:	----------	----------
conclusion:	patient *a* has disease D	patient *a* has disease D

Of course, a probabilistic abduction is more unreliable than the corresponding prototypical one, and the lower the relative frequency is, the more unreliable is the inference. To take a concrete example: should a general practitioner ever send a patient suffering from headache to a brain MRI (magnetic resonance imaging)? The overwhelming majority of patients suffering from headaches have no brain tumor. But the single patient in front of the doctor might be the rare case. Probabilistic abduction has to rely very much on the content at hand. The abductive inference form does not in itself transmit much epistemic credibility from the premises to the conclusion.

Abduction to an unknown kind has also its place in statistics related to medicine and some other sciences. What is unknown is then *the statistical model* that is fitting for some kind of research. After a statistical model has been chosen, one knows fairly well both how to find the sample and how to move from the sample to the population, but what kind of statistical model to choose is at the start an open question. Often, the situation is such that one can choose an already existing model, but sometimes the model itself has to be created. In the latter case there is true abduction to an unknown kind – to an unknown kind of statistical model.

4.7.6 Cross-over probabilistic inferences

From the preceding three sections it should be clear that the traditional formal schemas for deductive, inductive, and abductive inferences allow the introduction of probability statements. In these sections it was tacitly assumed that the premises-statements and the conclusion-statement are of the same general character: (a) purely mathematical, (b) objective, or (c) epistemic. Now we will present some inferences where the premises are

frequency-objective (ontological) statements and the conclusion is a subjective (epistemological) probability statement. We call them cross-over probabilistic inferences, and we present them because if they are not clearly seen, then some of our earlier remarks may look odd. Here are the first two varieties:

cross-over probabilistic inference (i)

premise 1:	the relative frequency of persons with disease D that get cured by recipe R is high (say, 0.95)
premise 2:	patient *a* has disease D and gets R
hence:	----------
conclusion:	the epistemic probability that patient *a* will be cured by R is 1

cross-over probabilistic inference (ii)

premise 1:	the relative frequency of persons with disease D that get cured by recipe R is small (say, 0.05)
premise 2:	patient *a* has disease D and gets R
hence:	----------
conclusion:	the epistemic probability that patient *a* will be cured by R is 0

Both these inferences may at first look spurious, turning ‘0.95’ into ‘1’ and ‘0.05’ into ‘0’, respectively, but they aren’t. For two reasons. First, often when we act it is impossible to have a distanced attitude thinking that the outcome is only probable and not ensured. Therefore, transitions from various frequency-objective statements to the *epistemic and subjective* probabilities one and zero are often needed. Second, the relative frequencies in question do not rule out the cases that the patient *a* may surely be cured and not cured, respectively.

In the third variety, a cross-over inference assigns the numerical value of the relative frequency to the epistemic probability in the conclusion. This is of course also possible. In this case, it is interesting to make a comparison with the inductive transition from a frequency-objective statement to a

realistically interpreted singular-objective statement. Here are the two schemas:

	cross-over probabilistic inference (iii)
premise 1:	the relative frequency of persons with disease D that get cured by recipe R is 0.x
premise 2:	patient *a* has disease D and gets R
hence:	----------
conclusion:	the epistemic probability that *a* will be cured by R is 0.x

	inductive probabilistic inference
premise 1:	the relative frequency of persons with disease D that get cured by recipe R is 0.x
premise 2:	patient *a* has disease D and gets R
hence:	----------
conclusion:	the singular-objective probability that *a* will be cured by R is 0.x

Some of the confusions that surround probability talk has, we think, its root in the fact that sometimes two different kinds of probability statements with the same numerical probability – such as the statements ‘the *epistemic* probability that *a* will be cured by R is 0.x’ and ‘the *singular-objective* probability that *a* will be cured by R is 0.x’ – can be inferred from exactly the same premises. That is, the brief singular statement ‘the probability that *a* will be cured by R is 0.x’ may be regarded as fusing two theoretically distinct meanings. It may have practical advantages, but it may also have the negative consequence that the problem of the interpretation of singular-objective probability statements (highlighted in Sections 4.7.1 and 4.7.2) becomes hidden. We guess that there is much more to say and make clear here, but we will not try to do it.

4.8 Harvey’s scientific argumentation

We will now, by means of an example from the history of medical science, highlight how different types of arguments can be woven together when a

radically new theory is argued for. Our example is taken from William Harvey's once revolutionary book about the movement of the heart and blood circulation in animals', *Exercitatio Anatomica de Motu Cordis et Sanguinis in Animalibus* (1628), usually abbreviated as *De Motu Cordis*. This work has rightly been praised as an excellent example of how medical research should be conducted. But, of course, all medical research cannot have such a pioneering character.

As we have earlier mentioned (Chapter 2.3), Harvey's views became essential parts of the criticism and eventual rejection of Galen's theories. According to Galen, food is convened into blood in the liver, from which the blood goes to the right part of the heart, where it is heated, and then it moves through the other blood vessels in a centrifugal direction out into the organs, extremities and tissues where the blood is consumed. Moreover, the left part of the heart is, according to Galen, assumed to contain the so called vital spirit, which is supposed to be distributed from this side of the heart to the rest of the body.

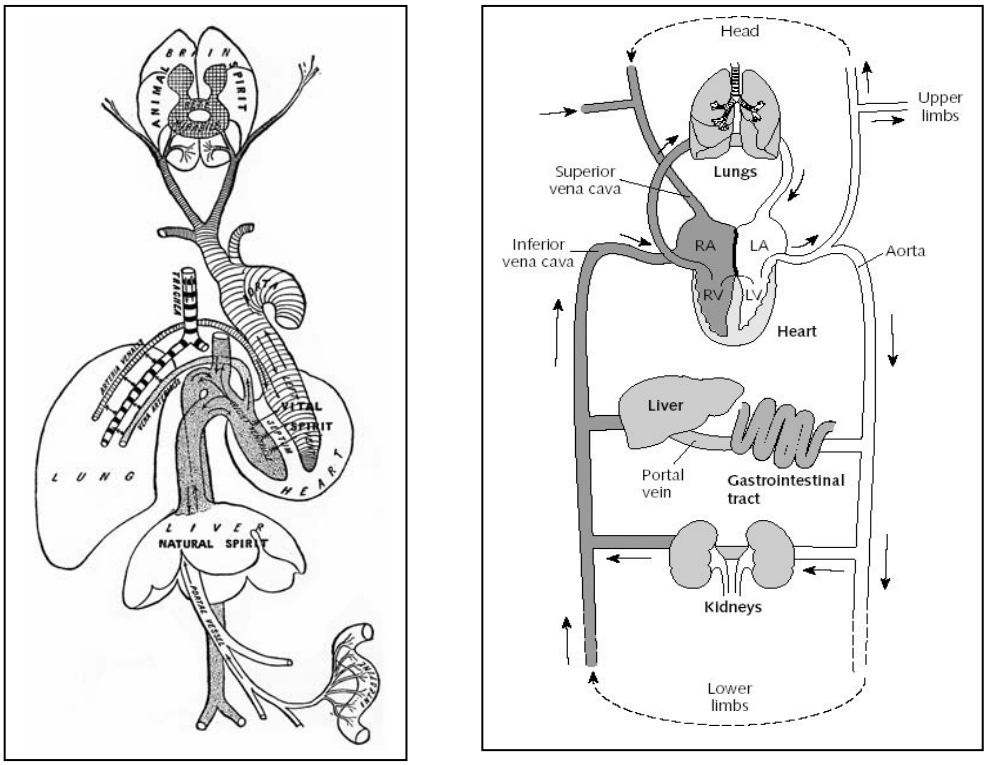


Figure 2: *Galen's and Harvey's views of the body.*

According to Harvey, the blood moves in two circles in relation to the heart, a minor and a major circulation. In the minor circulation, the blood moves from the right side heart-chamber or ventricle via the pulmonary artery to the lungs, passes through the lungs, and goes via the pulmonary veins back to the left side atrium or auricle of the heart. From here, the blood is pumped into the left side heart-chamber, and from this ventricle the major circulation starts. The blood is pumped into the large body arterial vessel (the aorta) and further into the different organs and tissues. When the bloodstream has passed the organs and tissues it returns to the right side auricle of the heart via the veins. From this auricle, the blood is pumped into the right side chamber, and back to the minor circulation, i.e., back to where it started from (see Figure 2).

According to modern terminology, *arteries* are vessels that lead the bloodstream away from the heart, and *veins* are vessels that lead the bloodstream towards the heart. All arteries except the pulmonary artery are carrying blood saturated with oxygen, and all veins except the pulmonary vein are carrying deoxygenated blood. It is important to underline the fact that the function of the lungs and oxygenation was not yet discovered. Malpighi discovered the capillary circulation in the lungs in 1661, which before this discovery, provided an anomaly in Harvey's theory. But it was not until the chemical theory of oxygen was discovered (around 1775) that we got the modern explanation of the function of the lungs. (Compare Harvey's and Galen's views as represented in Figure 2.)

The partly new situation that Harvey could benefit from can be summed up in the following way:

1. It had become accepted that physicians could question the old authorized knowledge, and that there were reasons to make one's own observations and experiments.
2. It had become legitimate for physicians to perform dissections of human corpses.
3. The idea that blood may circulate had already been suggested by Realdo Colombo (1516-1553) and Andrea Cesalpino (1519-1603), but it had not become accepted.
4. The existence of the minor circulation of blood had already been described by Colombo and Michael Servetus (1511-1553).

5. There were some speculations concerning the possibility of capillary passages in the tissues.
6. During dissections of corpses, blood had been observed both in the left heart chamber and in the large arterial vessels.
7. Small pores in the walls between the right and the left chambers of the heart, as assumed by Galen, could not be observed.

Harvey was an Englishman, but he had studied in Padua under the famous anatomist Hieronymus Fabricius, who discovered the valves in the veins. For a brief time, he was contemporary Francis Bacon's (see Chapter 3.4) physician. In all probability, Harvey was influenced by Bacon's philosophy. Especially, it has been claimed, he was inspired by the *Novum Organum* (1620) in which Bacon argued that one should make "induction upon data carefully collected and considered."

Harvey made meticulous and repeated observations in order to obtain inductive support for his theories. However, before we present Harvey's ways of argumentation, we would like to give a summery of Galen's views. In order to understand Harvey, one has to know what he was arguing against (see also Chapters 2.3 and 2.4).

According to Galen:

1. The heart is neither a pump nor a muscle, it is a passive reservoir; due to the movements of the chest it has a certain suction-capacity that helps the blood to go from the liver to the heart.
2. The heart is also a thermal unit whose function is to keep the blood thin and liquid.
3. The blood vessels contain, apart from the sanguine fluid, even black bile, yellow bile, and phlegm. Together these make up the four basic bodily fluids (not to be conflated with the three kinds of spirits or pneuma mentioned in point 6 below).
4. The right heart-chamber keeps the blood in a lapping movement whereby the four fluids are mixed.
5. There is blood mainly in the veins - the vessel system connected to the right part of the heart - and the bloodstream moves towards the periphery slowly according to the pace at which it is produced in the liver and consumed in organs, extremities and tissues. It departs from

the liver and goes to the right part of the heart, where it is heated, and then further from the heart to the periphery of the body.

6. There are three kinds of spirits: a) *spiritus animalis*, which is produced in the brain, is distributed along the assumed hollow nerves, and whose functions are related to perception and movement; b) *spiritus vitalis*, which is produced in the left part of the heart, is distributed by the arteries, and whose function is to 'vitalize' the body and to keep the body warm; c) *spiritus naturalis*, which is produced in the liver, but goes to and passes through the right part of the heart, and whose function is to supply the organs and tissues with nutrition.

7. The left side of the heart chamber and the arterial vessels containing spiritus vitalis are together considered to be merely a pneumatic system connected with the lungs and air-pipe, which was also categorized as an artery.

8. There are pores in the wall (septum) between the two heart chambers. By means of these pores blood can percolate from the right side to the left side, where it is mixed with the air and then sublimated into spiritus vitalis.

9. There are two kinds of vessels: arteries and veins. Arteries depart from the left side of the heart and contain spiritus vitalis and these vessels are pulsating by means of a kind of (peristaltic) movement in the artery walls. The vessel system made up of veins contains blood, for instance, 'the artery-like lung vein', which we currently define as an artery (a. pulminalis). What we today refer to as the lung veins was according to the Galenic theory referred to as a vein-like artery. (In some languages, arteries and veins have names that can be translated into English as 'pulsating vessel' and 'blood vessel'; in German, "Pulsader" and "Blutader", respectively.)

Now, against this background, how did Harvey proceed and argue? Those who have observed a living heart know how hard it can be to have a clear picture of what definite pattern that the rather fast movements of the heart constitute. It is by no means immediately obvious that the heart functions like a pump. Harvey, however, developed an experimental technique that made it possible to examine dying animal hearts. Since these perform only

a few beats per minute, it became easier to see that the heart might function as a pump.

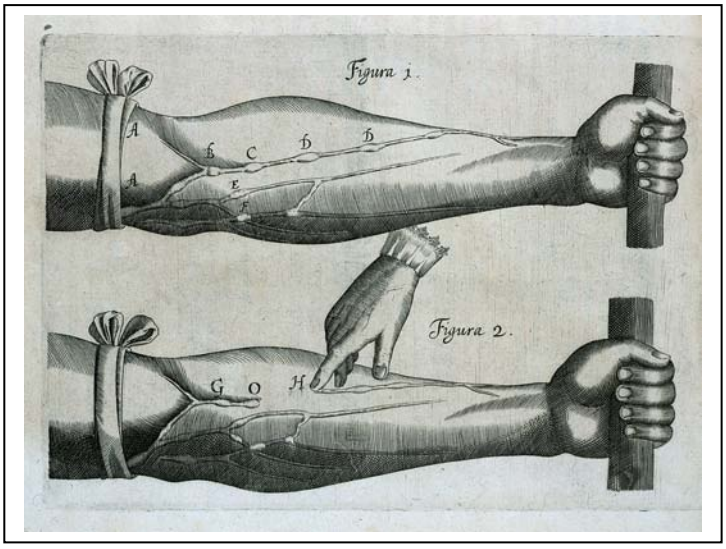


Figure 3: *Experiment made by Harvey.*

Harvey also figured out the following simple but ingenious experiment. He bound a string around an arm and observed what happened. The experiment is illustrated in one of the few figures in *De Motu Cordis* (Figure 3).

Considering the theories of Galen, his own hypothesis, and his own observations, Harvey might have reasoned somewhat as follows:

> I. <u>What to expect according to Galen?</u> According to Galen's theory, the direction of the bloodstream in the arms is only outwards, and we might thus expect an accumulation of blood proximally to the string. Distally to the string, the blood vessels should fade away and finally collapse.
> II. <u>What do we observe?</u> The blood vessels proximally to the string fade away, while the blood vessels distally to the string become swollen. When we the string is taken away, the proximal blood vessels appear again whereas the swelling disappears.

III. What conclusions can be drawn? As we have explained in Section 4.6, definite conclusions need auxiliary hypothesis. Now, some hundreds years later, it is easy to say that there were no good auxiliary hypotheses to be found, but at the moment of this experiment things could not be regarded as definitely settled. Nonetheless one might say that *at first sight* the outcome of Harvey's experiment falsifies Galen's theory. Obviously, Harvey demonstrated a serious anomaly within the Galenic paradigm. The same observation, conversely, supported Harvey's own view that the blood in the veins is streaming towards the heart; quite contrary to the Galenic teachings.

In another experiment Harvey lanced the main pulse artery - the aortic vessel - in a living animal. Let us look at the argumentation:

I. What do we expect according to Galen? We might expect that something air-like, spiritus vitalis, would stream out; or, at least, a mixture of something air-like and blood.
II. What do we observe? Blood, and only blood, is splashed out.
III. What conclusions can be drawn? Again, hypothetico-deductive argumentation tells against the Galenic theory, i.e., that there is again a serious anomaly. Harvey's experiment displayed a considerable discrepancy between what the Galenic theory (in its actual form) predicted and what was observed. It also generated the hypothesis that the left part of the heart and the arterial system contains blood.

Harvey made more experiments with the same kind of conclusions. The Galenic theory became confronted not only with one or two, but with several unexpected empirical phenomena. In this phase, however, Harvey had only begun his journey towards his theory. Many things remained to be explained before he was able to provide a coherent and reasonable alternative. In this work, Harvey also uses some thought experiments and reductio ad absurdum arguments.

If, Harvey asked himself, the function of the lung vein (today we refer to this as the pulmonary artery as the bloodstream depart from the heart) is to supply the lungs with spiritus naturalis (nutrition), why are they so big

(thick) compared to other veins with a similar function, e.g., those in the legs? Ought not vessels that have the same function have somewhat similar dimensions? Furthermore, if nonetheless there can be a difference, shouldn't it go in the other direction? Since the legs are larger than the lungs, shouldn't the vessels supplying the legs with nutrition be bigger than those supplying the lungs with nutrition? Merely by thinking about Galen's theory, Harvey managed to discover an absurdity.

Harvey used an abductive and (at the same time) analogical inference when he argued that the heart is a muscle. If we touch a muscle, e.g., the biceps, we will find it soft and relaxed when it is not used, but hard and tensed when it is used. Correspondingly, when we touch a small and hard heart it feels like a tensed biceps, and when it is enlarged and dilated it feels like a relaxed biceps. If we call the phenomenon 'feels like biceps', we can write:

premise 1:	all muscles feel like biceps
premise 2:	the heart feels like biceps
hence:	---------- (ABDUCTION and ANALOGY)
conclusion:	the heart is a muscle

Simplicity arguments come in when Harvey argues against the view that there are invisibles pores in the wall between the heart chambers. The transport of blood from the right side to the left side of the heart is explained more simply by the assumption that the bloodstream pass the lung veins and the porous lungs.

When Harvey argued for the existence of the minor circulation, he used analogy reasoning. According to Harvey's assumption, the blood passes the lungs. How is this possible? Harvey gave the following answers:

there might be an *X* that is to (can pass through) the porous lungs
what blood is to the compact liver
hence: ---------- (ANALOGY)
blood is to (can pass through) the porous lungs
what blood is to the compact liver

there might be an *X* that is to (can pass through) the porous lungs
what urine is to the compact kidneys
hence: ---------- (ANALOGY)
blood is to (passes through) the porous lungs
what urine is to the compact kidneys

In his description of the relationship between the functions of the auricles and the ventricles of the heart, Harvey used a simple analogy with a pistol. He maintained that the auricles and ventricles together was “as a piece of machinery, in which, though one wheel gives motion to another, yet all the wheels seem to move simultaneously”. If things can work in this manner in a weapon, things may work in the same way in the body.

In a special argumentation, Harvey made an analogy between, on the one hand, the heart and the circulation of the blood and, on the other hand, the sun and the circulation of water in nature. When heated by the sun, the humidity of the fields and the water of the oceans rise into the air. When in the air, the water is condensed into rain clouds that eventually produce raindrops that fall down on the earth; thereby creating humidity in the fields and water in the oceans. This analogy made a strong impact, since many people at this time believed in a general correspondence between micro-cosmos and macro-cosmos.

Harvey also uses ad hominem arguments. Indeed, when Harvey highlights an argument, or some precondition for an argument, he often refers to authorities and uses phrases such as ‘the prominent anatomist Colombo’. So he did, for instance, when he argued that it is possible for the bloodstream to pass the compact liver tissue, which was a premise in one of the analogy arguments mentioned above.

One kind of ad hominem argument was the ad divinum argument. For example, one of Harvey’s arguments in support of the hypothesis that blood was circulating was that circular movements are more perfect, beautiful, divine and therefore desirable than straight ones. According to Harvey, this is a manifestation of God’s great creativity. In these times, a reference to God was understood as a special kind of ad hominem argument, or, more correctly, ad divinum argument.

When qualitative reasoning is converted into quantitative, new and interesting conclusions often appear. According to Galen, the velocity of

the bloodstream was mainly determined by the speed in which the liver produced the blood; and the volume of blood streaming from the heart to the other organs and tissues was assumed to be rather small. Harvey, as we have seen, assumed that the heart is a muscle, which – due to contractions – pumps the blood into the artery system; and that due to the shape of the heart valves, the blood cannot return to the heart chambers. Having looked at hearts, Harvey assumed that the left chamber contains approximately half an ounce of blood, and then he multiplied this volume with the number of heartbeats per minute. The interesting conclusion was that during half an hour the heart pumps out a volume of blood that is much larger than that estimated for the body as a whole. No one thought that such a quantity of blood could be produced during that interval in the liver, and consumed in the organs and tissues. That is, thanks to Harvey's quantitative thinking the Galenic theory was faced with still another difficult anomaly.

The fact that the Galen's theory of the blood and the heart was shown to contain both many empirical anomalies and some theoretical inconsistencies, and eventually was rejected in favor of Harvey's, does not mean that Harvey would have been better off starting from scratch. Mostly, it is better to start from a false theory than from no theory at all. It should also be remembered that Harvey did not break with Galen in all aspects. For instance, he never questioned the concept of 'spiritus', which is a bit astonishing since he assumed that all arteries, veins, and both sides of the heart contain one and the same substance, blood. Nor did he question the view that blood is composed of the four old bodily fluids. Furthermore, like Galen, Harvey regarded the heart as also being a kind of heater – a sun in the microcosmos constituted by the human body. He thought that the percolation of blood through the lung also had the function to temper the blood and make sure that it should not start to boil.

Now leaving his theory of blood circulation behind, it is also interesting to note that Harvey, despite his careful empirical observations, never questioned the old but false view that nerve fibers are hollow. On the other hand, he did question the new and true view that there is a lymph system. This system was discovered by Thomas Bartholin (1616-1680) in 1653, but Harvey refused to accept this finding. The existence of lymph vessels had been demonstrated by vivisection on dogs; they had been described as 'white veins' containing a milky fluid. In order to provide an explanation

of this to him seemingly odd phenomenon, Harvey maintained that the vessels in question were nothing but nerve fibers. The fact that Harvey primarily had studied fishes and birds – animals with no lymph systems – might also have influenced his interpretation of the 'white veins'. Empirical examinations were important in Harvey's scientific work, but his denial of the existence of a third system, the lymph system, illustrates the view that observations are theory-laden.

Harvey did not bother about any distinction between a 'context of discovery' and a 'context of justification'. At the same time as he provided arguments against the Galenic theory, he also generated his own hypotheses and produced arguments in favor of these. He did not first produce a well formulated and in detail specified theory, which he later started to test in accordance with the pattern for hypothetico-deductive arguments.

Hopefully, our presentation of Harvey has given credence to our view that scientific argumentation is often composed of different but intertwined kinds of argument. Harvey used interchangeably:

1) Arguments from perceptions (i.e., direct ocular and palpating observations as well as observations based on experiments)
2) Inductive inferences (what is true of some human bodies are true of all)
3) Hypothetico-deductive arguments
4) Abductive inferences
5) Thought experiments and reductio ad absurdum arguments
6) Arguments from simplicity, beauty, and analogy
7) Ad hominem arguments

Arguments can, as we said at the beginning of this chapter, be connected to each other both as in a chain and as in a wire. Harvey, we would like to say, mostly wired his deductions, inductions, abductions, and other kinds of arguments. But even such a wiring cannot change the fact that empirical theories and hypotheses are epistemologically underdetermined. Fallibilism reigns, and not even a huge dose of tacit knowledge, which will be discussed in Chapter 5, can alter this fact. Tacit knowledge is fallible knowledge too.

4.9 Has science proved that human bodies do not exist?

We started this chapter by presenting formal-logical deductive inferences. In passing, we said that there is seldom need in actual science to make these explicit. But ‘seldom’ is not ‘never’. Science overlaps with philosophy, and we will end this chapter by returning to logical contradictions. Medical science takes the existence of material bodies for granted, but in philosophy this assumption is questioned both by traditional idealists (claim: there are only souls and mental phenomena) and modern linguistic idealists (claim: we can only know conceptual constructs). Sometimes, even science itself is referred to as having shown that there are no material bodies. But then, mostly, the critics contradict themselves as in the quotation below, which comes from the advertisement of a book entitled *Matter. The Other Name For Illusion* (author: Harun Yahya). The advertisement says:

> All events and objects that we encounter in real life—buildings, people, cities, cars, places—in fact, everything we see, hold, touch, smell, taste and hear—come into existence as visions and feelings in our brains.
>
> We are taught to think that these images and feelings are caused by a solid world outside of our brains, where material things exist. However, in reality we never see real existing materials and we never touch real materials. In other words, every material entity which we believe exists in our lives, is, in fact, only a vision which is created in our brains.
>
> This is not a philosophical speculation. It is an empirical fact that has been proven by modern science. Today, any scientist who is a specialist in medicine, biology, neurology or any other field related to brain research would say, when asked how and where we see the world, that we see the whole world in the vision center located in our brains.

These paragraphs imply, to put it bluntly, (a) that everything we encounter in perception exists in an existing material thing called brain, and (b) that

there exist no material brains. But it is a logical contradiction to claim both that there are material brains and that there are not. Hopefully, most medical scientists and clinicians will continue to believe that our brains are material things and that we know at least something about how they are structured and function.

Reference list

Aristotle. *Metaphysics* (book IV). (Many editions.)

Bunge M. *The Myth of Simplicity*. Prentice Hall. Englewood Cliffs, N.J. 1963.

Chalmers AF. *What is This Thing Called Science?* Hackett Publishing. Indianapolis 1999.

Coady CAJ. *Testimony: A Philosophical Study*. Oxford University Press. Oxford 1992.

Davidson A. Styles of Reasoning. In Galison P, Stump G (eds.). *The Disunity of Science*. Stanford University Press. Palo Alto 1996.

Hanson NR. *Patterns of Discovery*. Cambridge 1958.

Harrowitz N. The Body of the Detective Model – Charles S. Peirce and Edgar Allan Poe. In Eco U, Seboek TA (eds). *'The Sign of Three. Dupin, Homes, Peirce.'* Indiana University Press. Bloomington 1983.

Harvey W. *Anatomical Studies on the Motion of the Heart and Blood* (translated by Leake C). Springfield 1978.

Hempel C. *Aspects of Scientific Explanation*. The Free Press. New York 1965.

Hempel C. Provisoes: A Problem Concerning the Inferential Function of Scientific Theories. *Erkenntnis* 1988; 28: 147-64.

Jansen L. The Ontology of Tendencies and Medical Information Sciences. In Johansson I, Klein, B, Roth-Berghofer T (eds.). *WSPI 2006: Contributions to the Third International Workshop on Philosophy and Informatics*. Saarbrücken 2006.

Johansson I. Ceteris paribus Clauses, Closure Clauses and Falsifiability. *Zeitschrift für allgemeine Wissenschaftstheorie* 1980; IX: 16-22.

Kalman H. *The Structure of Knowing. Existential Trust as an Epistemological Category*. Swedish Science Press. Uppsala 1999.

Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *British Medical Journal* 2003; 326: 1167-70.

Mowry B. From Galen's Theory to William Harvey's Theory: A Case Study in the Rationality of Scientific Theory Change. *Studies in History and Philosophy of Science* 1985; 16: 49-82.

Peirce CS. *The Philosophy of Peirce. Selected Writings* (ed. J Buchler). Routledge & Kegan Paul. London 1956.

Porter R. *The Greatest Benefit to Mankind. A Medical History of Humanity from Antiquity to the Present*. Fontana Press. London 1999.

Rosenthal J. *Wahrscheinlichkeiten als Tendenzen. Eine Untersuchung objektiver Wahrscheinlichkeitsbegriffe*. mentis Verlag. Paderborn 2002.

Sebeok TA, Umiker-Sebeok J. "You know my method": A Juxtaposition of Charles S Peirce and Sherlock Holmes. In Eco U, Seboek TA (eds). *'The Sign of Three. Dupin, Homes, Peirce.'* Indiana University Press. Bloomington 1983.

Semmelweis IP. *Die Aetiologie, der Begriff und die Prophylaxis des Kindbettfiebers*. Hartlen's Verlag-Expedition. Pest, Wien, and Leipzig 1861.

Sowa JF, Majundar AK. Analogical Reasoning. In Aldo A et al. (eds.). *Conceptual Structures for Knowledge Creation and Communication. LNAI 2746*. Springer-Verlag. Berlin 2003.

5. Knowing How and Knowing That

In contrast to some other languages, English mirrors immediately the fact that *knowing how* (to do something) and *knowing that* (something is the case) have, despite their differences, something in common; both are forms of knowledge. In this chapter, we will present what is peculiar to know-how and argue that optimal knowledge growth in medicine requires much interaction between knowing-how and knowing-that. Note that to 'know how a thing functions' is to know *that* it functions in a certain way, i.e., this kind of knowledge is a form of knowing-that (something is the case); what might be called *knowing-why* – i.e., a knowledge of explanations why something happened, why something exists, and why something stopped functioning – is also a form of knowing-that. *Knowing-what* (kind of thing something is) may be both knowing-that and know-how. When it is the latter, it is an ability to identify in perception something as being of a certain kind.

5.1 Tacit knowledge

Most of us have much know-how without much of a corresponding knowing-that. We can make phone calls, but we do not know much about how telephones work; we can write, but we do not know in detail how we hold the pen and how we move our hand when writing; we can talk, but we hardly know anything at all about how we move our mouth and our tongue. Examples can be multiplied almost to infinity; even though, of course, the list differs a little from person to person. There are know-hows that concern only the body, some that concern the use of tools, and others that concern the use of machines and computers; there are also know-hows that are very important in interactions with animals and persons. In medicine, the last kind of know-how is important when one tries to understand and/or improve doctor-patient and nurse-patient relations.

Michael Polanyi (1891-1976), the scientist and philosopher who has coined the term 'tacit knowledge', has an illustrative example of what a discrepancy between know-how and knowing-that can look like. Most of the people are able to ride a bike, but very few are able (a) to correctly

describe how they move their body when biking and (b) to correctly explain why humans can bike at all. Most people think falsely that they do not move their arms and hands when they are biking without turning, and that biking should primarily be explained by our sense of balance. In fact, when biking we rely only to a small extent on this sensory system. Biking is in the main made possible by centrifugal forces. When we turn to the left, a centrifugal force arises that will tilt us to the right; we then move our hands and make a turn to the right, whereby a force arises that will tilt us to the left. We then make a new little turn to the left, and so on. Tiny movements of the hands enable us to keep the bike upright by creating centrifugal forces of opposite directions. The bike is actually tottering from left to right, even though we may think that the bike is continuously in a stable upright position. Even when we are biking straight ahead, we are unconsciously making small turns; and we have to do these turns. The reader who does not believe in this explanation can easily test it. Weld the handlebars and the frame of your bike together, and try to ride in a straight line by means of your sense of balance. After a few meters you will inevitably fall to one side. That is, having know-how about bicycling not only goes well together with a complete lack of a corresponding knowing-that; it often even lives in peaceful co-existence with false knowing-that.

Knowing-that can by definition only exist in symbol systems; of course, mostly, this is a natural language. Therefore, knowing-that might also be called ‘spoken (or non-tacit) knowledge’. Know-how, on the other hand, can exist both with and without symbol systems. Children can learn many skills such as walking, tying shoes, sawing, sewing, and biking before they learn to speak properly. This is one reason why know-how deserves to be called tacit knowledge. Another reason is that even if language and knowing-that have in fact been useful when a non-language skill has been acquired, such a skill can later be put to work ‘tacitly’. Adults, let it be said, mostly have some knowing-that about each of their skills. But they must stop thinking about this knowledge when actually using their know-how. Because if one concentrates on the knowing-that aspect of a know-how when using the latter, one is often thereby obstructing or getting in the way of this skill. For instance, if when riding a bike one starts to think of how to move the handlebars in order to create good centrifugal forces, then

one will impair one's actual biking. Similarly, orators should not think when speaking.

As know-how can exist without any knowing-that about this very know-how, conversely, one might know-that about a certain know-how without being able to do anything at all in this respect. One can read much about how to perform heart operations without becoming able to perform one – not even a relatively bad one.

Know-how is not restricted to knowledge about how to use our body and how to use tools. This fact has to be stressed because some older literature on tacit knowledge gives the contrary and false impression. First, as we said at once, there is tacit knowledge in relation to the use of machines and interactions with computers, and there is such knowledge in our interactions with other human beings and animals. There is even know-how in relation to intellectual skills. Reading, writing, and performing mathematical calculations cannot be efficiently done without tacit knowledge. Normally, when reading and writing, we are not aware of anything that has to do with grammatical, semantic, and pragmatic language rules, not to speak about being aware of how the eyes and the hands are moving. We simply read and write. Similarly, without bothering about axioms and theorems in number theory, we simply add, subtract, multiply, and divide numbers with each other.

Sometimes inventions based on new knowing-that make old know-how superfluous. But then some new know-how has arisen instead. When old-fashioned handicraft based production was replaced by machines and industrial production, tool skills were replaced by machine-managing skills. One kind of know-how was replaced by another.

Tacit knowledge is action-related knowledge. Therefore, despite some similarities, tacit knowledge must not be put on a par with completely automatic reactions, adaptations, and behaviors of the body. Even though tacit knowledge is tacit, it has a connection to consciousness and agency. A necessary condition for an action (activity, behavior, or process) to be an expression of know-how is that its overarching pattern is governed by the person's will. This fact is reflected in ordinary talk about abilities and skills. It makes good sense to ask whether a certain person is able to ride a bike, cook a meal, play tennis, make diagnoses, or perform surgical operations, but it makes no sense to ask whether a person's heart is able to

beat, or whether the person is able to make his heart beat – the heart beats independently of our will. Nonetheless, the heart's pulse, just like many other bodily processes, adapts to new conditions; in a sense, the heart can learn how to beat under various conditions, but it has no tacit knowledge.

A last warning, nor must the concept of tacit knowledge presented be conflated with any psychoanalytic or otherwise psychological concept of 'the unconsciously known'. What in this sense is said to be unconscious are memories and desires that are assumed to be actively 'repressed', since they are assumed to fill the mind with agony if they were suddenly to become conscious. Also, in psychoanalytic theory, they are supposed to be reflected in dreams and be responsible for neurotic behavior. Tacit knowledge is by no means identical with 'repressed knowledge', be such knowledge existent or not.

The fact that a person has good tacit knowledge is in non-philosophical discourses expressed by sentences such as 'he has it in his fingers', 'he has a good feeling for it', and 'he has a good clinical glance'.

5.2 Improving know-how

Those who have developed a skill to a very high degree are sometimes said to have developed their skill into an art. Physicians of all kinds may well in this sense try to develop their specific skills into arts. But then they have better to learn (as knowing-that) that know-how is not just a special form of knowledge, it has its own methods of improvement, too. Even though new knowing-that can be used to introduce and to improve an already existing know-how, we want to emphasize the fact that know-how can also be improved independently. Becoming proficient at handicrafts, sports and music requires years of practice and personal experience; the same goes for proficient handling of some machines and computers; and it is also true of conducting good laboratory experiments, making good diagnoses, and performing good operations. There are four general ways in which know-how can be improved:

1. practicing on one's own
2. imitation
3. practicing with a tutor
4. creative proficiency.

In cases 1, 2, and 4, know-how is improved independently of any reading of manuals or other kind of apprehension of relevant knowing-that. Now some words of explanation.

1. Practicing on one's own. The more we practice an activity, the more we improve the corresponding skill. As the old adage says: 'practice makes perfect'. Or, more moderately: 'practice greatly improves proficiency'. Obviously, the human body and the human brain have an in-built capacity of self-learning through trial and error. It functions in relation to kids (that for instance learn to ride a bike) as well as in relation to adults (e.g., medical students who train clinical skills). A remarkable fact is that this kind of tacit learning by repetition also can function across the gap between real activities and simulations of these activities. Since long, pilots are trained in simulator cockpits. The computer revolution may in the future make physicians train many things on computer simulations, which is actually already the fact within some areas such as anesthesia, internal medicine, and surgery.

2. Imitation. Simply looking at and/or listening to other people performing a certain activity, can improve one's own skill in this respect. Small children's ability to imitate is remarkable. But even adults can learn new activities and improve previously acquired skills by means of imitation. In cases where one can learn an activity both by imitating and by reading a manual, it is often easier to learn it by imitation. It is against this background that the cry for 'positive role models' should be understood. The fact that know-how can be improved by imitating shows that there is a close connection between our ability to perceive and our ability to act. Our perceptual system does not exclusively process information; by means of this information, it also improves our actions. Imitation and practice on one's own can be fused in a peculiar way that has been developed by sports psychologists. Some kinds of know-how can be improved on by imitating a

repeatedly created mental picture of oneself performing very successfully the activity in question. For instance, if you are a basketball player, you may improve your penalty shooting by visualizing yourself – over and over again – making perfect penalty scores.

3. Practicing with a tutor. Neither practicing on one's own, nor imitating, nor creative proficiency requires language to describe the new know-how in question. But when a tutor (includes teachers, supervisors, trainers, coaches, and masters of all kinds) enters the scene, language and knowing-that are also brought in. When a driving instructor teaches a novice to drive a car, he begins by describing how the steering wheel, the pedals, and the stick-shift should be used. Thus he first gives some knowing-that of the know-how that the pupil shall learn. Then the pupil tries to follow this oral instruction and practice begins. But even later in the process the driving instructor uses his knowing-that. He makes remarks like 'relax your hands on the steering wheel', 'listen to the sound of the motor before you change gears', 'press the gas pedal more slowly', and so on. Common to all these knowing-that transfers is their approximate character. They are very abstract in relation to the wished for performance that constitutes 'flow'; they might be said to supply necessary but by no means sufficient descriptions of good know-how. However, despite being only rules of thumb, they can function well in interaction with the practitioner's own practice. And what in these respects goes for learning how to drive goes for most know-how learning.

4. Creative proficiency. Independently of all imitation and all prior pictures of an activity, a person may start to perform an already known activity in a completely new way. He so to speak 'creates in action'. We have chosen to call this phenomenon 'creative proficiency'. There is and has been much literature about 'creative thinking'. Sometimes this talk gives the false impression that creativity is an exclusively intellectual thinking-phenomenon; one consequence being that all *radical* know-how improvements have to come about indirectly via radically new knowing-that. But just as there are two forms of knowing, knowing-that and know-how, there are two forms of radical creativity, 'creative thinking' and 'creative proficiency'. For example, Jimi Hendrix did not create his new

way of playing guitar by first creating a mental picture of how to play guitar his own way. Here is an example of clinical proficiency. A mother with her four year old boy is consulting a general practitioner (GP). The boy is suffering from an ear disease that makes an examination of the internal part of the ear (an otoscopy) necessary. GPs know that it is difficult to have the child's permission to examine his ear, and the present GP suddenly on impulse asks whether the child is able 'to hear the light' when the doctor looks into the ear. The boy becomes curious and asks the doctor to perform the examination in order to see whether he can hear the light.

In most actions one can discern part-actions. That is, when we act we concentrate on an overarching goal even though we are in some sense aware of the part-actions. With Polanyi one might say that we act *from* the parts *to* the whole action. This from-to structure is important in some learning situations. It is sometimes possible, but not always, first to learn the part movements of an activity and then integrate these into a homogeneous 'Gestalt', as when learning to drive a car. Sometimes, when we already can perform a certain activity, we can improve on it by first repeating merely a detail many times, and then try to let the consciousness of the detail disappear in the consciousness of the larger integrated whole that constitutes the activity in question. Polanyi exemplifies with a piano teacher that interrupts his sonata playing pupil in order to make him touch a key slightly more softly in a certain passage. The pupil, under strong concentration, is forced to touch the key repeatedly in order to obtain the right softness of touch. Later on, when the whole sonata is to be played, concentration and consciousness has to be directed towards the whole. If this does not happen the transitions between the different keys will be incorrect. A particular note sounds right, but the music sounds wrong.

Tacit knowledge is also present in what we earlier have called 'perceptual structuring' (Chapter 3.2). When we concentrate on something in perception, we experience this something as having parts even if we cannot in detail see what the parts are like and describe them. We might say that we perceive *from* the parts of a percept *to* the whole percept. What part actions are to a whole action, perceptual details are to a perceived whole. One of Polanyi's examples comes from radiology. When a layperson looks at an X-ray, it is usually impossible for him to differentiate

between different anatomical details. The radiologist, on the other hand, immediately observes theses details in the same way as the layperson sees details in an ordinary photo. It is even as hard for the radiologist *not* to see anatomical details in the X-ray as it is for him and the layman *not* to see ordinary things in an ordinary picture. The radiologist possesses skilled perception.

A person at the outset of his education is a layperson. In the beginning of his studies, the radiology student only saw black and white spots on radiographies. When children learn their first language, initially they understand nothing; they only hear sounds. Nevertheless, they eventually become able to speak fluently and to understand immediately what other persons are saying.

Tacit knowledge is firmly anchored in the body and the brain. We know that the movement of the eyes of a radiologist that looks at an X-ray differs from those of a layperson looking at the same picture. The brain of an expert is probably able to receive and adapt to certain kinds of perceptual data which the novice's brain is not yet able to deal with. Therefore, the brain of an expert can send signals to the muscles that the brain of the novice cannot yet send. Such signals move extremely fast – in a billionth of a second – and without any awareness on our part. It is this fact that might have misled some thinkers to *identify* tacit knowledge with the bodily automatics that this knowledge is dependent on.

5.3 Interaction between knowing-how and knowing-that

In one specific sense, some kinds of scientific knowing-that are disconnected from all knowing-how: they completely lack practical application. A good example is cosmogony, the theory of the genesis of the universe. However, even such knowing-that is for its existence dependent on know-how – other than that which is always required by language itself. The theory is based on observations with instruments; hence all the skills and structured perceptions necessary for handling the instruments are necessary for the theory. No knowing-that concerned with the world in space and time can exist without many different types of know-how. Improved know-how can be a necessary requirement for new knowing-that. Lens grinding in relation to the microbiological paradigm (Chapter 2.5) is merely one of a huge number of examples in the history of science.

Conversely, new knowing-that can lead to improved know-how. The relation between lens grinding and microbiological discoveries is obvious, but so is the relation between new knowing-that about how systems and organs in the body functions and improved know-how about how to cure and prevent various diseases and illnesses. For instance, without detailed knowledge about how the heart and the blood system works, bypass operations would be impossible. Often, to see how something functions in detail (knowing-that) is enough for receiving cues about how to improves one's ability to repair it (know-how).

The purpose of basic research is to obtain knowing-that; this goes also for basic medical research. But the overarching knowledge purpose of the whole healthcare system, of which much of the medical research is a part, is to develop know-how. It shall embody knowledge about *how* to prevent diseases and illnesses, *how* to diagnose diseases and illnesses, *how* to treat diseases and illnesses, *how* to alleviate pain, and *how* to comfort a patient. From what has been said and indicated above, it ought to be clear how the general relationship and interaction between new knowing-that and improved know-how can look like – and that such an interaction is important. Below, we will show how the interaction between knowing-that and know-how can look like in a special case of 'practicing with a tutor'; one in which the practitioner so to speak becomes his own tutor. It concerns medical consultation.

As stated by Hippocrates, "Life is short, art long; the crisis fleeting; experience perilous, and decision difficult", and the GP is probably the first to recognize this. Since the GP is supposed to deal with numerous unselected patients per day, in many situations his skill requires quick adaptation, improvisation, and vigilance. Consultations have to be optimally efficient in relation to the problems for which the patients consult the doctor. This requires, apart from medical knowledge, communication skills and empathy. Not only the novice has to try to develop his consultation skills, now and then even the expert clinician has.

In a typical GP consultation there is only the doctor and one patient. A newly licensed doctor may have a senior colleague with him, but his work is mostly done alone. Our own performance of know-how is hard and often impossible to observe. One may be acutely aware that one is performing poorly, but one can nonetheless not see exactly what goes wrong.

Accordingly it is difficult to correct such sub-optimal or counterproductive actions. But videotapes have radically changed this predicament. Now, it is sometimes possible to observe one's performance in retrospect.

Medical consultations can be videotaped and the doctor (or other health care providers) can afterwards reflect on it, i.e., acquire knowing-that about his own know-how or about others. He can do it alone and try to be his own tutor, or he can do it together with colleagues or a senior tutor. The latter may then make various apt comments from which the former can benefit. Such video studies by novices as well as experts have given rise to some rules of thumb for GPs. There exist many different lists of such rules and relevant questions, and the subsequent list (developed by GP Charlotte Hedberg) is a model used in some continuing professional development programs in Sweden. It is called 'the Prismatic Model'.

As a piano pupil can be requested to play and concentrate on just one key at a time, the participants (students, nurses, or physicians) of prismatic-model-training are requested to concentrate on merely one aspect of a videotaped medical consultation at a time. Each aspect is named by a color and is associated with a corresponding pair of colored glasses.

1. The white glasses focus, to start with, on the health care provider's general perspective. Before the video is being played he is asked questions such as 'Had you seen the patient before or was it the first visit?', 'Was it a planned or unplanned consultation?', 'Did workload, work condition, and schedules for the day influence the actual consultation?' After having seen the video, the health care provider at hand is supposed to make some comments of his own and to say something about his feelings; then the tutors (or some colleagues) are giving their comments. Later on, all participants are asked to focus on the patient's perspective and try to imagine themselves as being the patient. They should think of themselves as having the same kind of body, and they shall describe the patient's life situation and physical illnesses/symptoms as comprehensibly and vividly as possible in sentences such as '*I* feel …' and '*My* illness make *my* life troublesome because …'.

2. The red perspective focuses on the conversation between the health care provider and the patient and, e.g., on who dominated the conversation.

The participants are asked to pay particular attention to the first three minutes of the consultation. Special focus is on the health care provider's questions. Do they open up for real discussion or are they leading questions? How does he react to the patient's answers? For instance, is he often 'humming' and/or often saying 'yes, I understand'. Does he ever say, e.g., 'Tell me more'? One also focuses on facts such as whether the health care provider summarized the conversation and whether he interrupted the patient. Also pauses and their importance for the conversation are discussed among the participants.

3. The pink glasses deal with the patient's and the provider's agenda, respectively. Relevant questions are: 'What is the patient's/provider's problem?', 'What are the patient's/provider's expectations?', and 'Do provider and patient agree on the actual problem and is there a common basis to the conversation?'

4. The orange perspective is concerned with explanations. Did the patient receive any explanations at all? If not – is there an explanation?; if yes – what did it look like and was it comprehensible? Did the patient receive information about diagnosis, prognosis, preventive measures, and treatments? If not – was it possible and relevant to provide such information?

5. The yellow color represents the body language between the two actors. What is the position of the provider and the patient, respectively? The participants are asked to describe the distance between the provider and patient, the latter's mimics and eye contact, as well as whether or not they copy each other's movements during the conversation. Questions about possible inconsistency between the body language and the spoken language are also relevant.

6. The green color represents emotional aspects. Is there a tension in the conversation? Has the provider given any emotional response to the patient's problem? Are there any key replies? Do we observe any 'laden' or 'golden' moments during the conversation?

7. The turquoise color focuses on the medical content of the conversation and how the doctor understood the medical task, the physical examination included. The participants are asked to focus on the description of the medical complexity of the case history rather than on a possible right answer.

8. The blue aspects concern gender, social issues, and taboos. Did the provider's/patient's gender or social background influence the content and development of the consultation? Would the consultation have been different if the provider and/or patient had been male/female, or vice versa? Did the provider or patient avoid certain topics such as sexuality, drinking habits, smoking habits, and death?

9. The violet color deals with the ethical aspects of the consultation. Was it a fair negotiation? Did the provider patronize the patient or did the patient patronize the provider? Did the provider respect the patient's autonomy and integrity? Did the patient respect the provider's professional autonomy?

10. The purple color deals with the time aspects. How did doctor and patient use the time? Was there a correspondence between clock-time and attention-time?

As a piano pupil is expected to have only background awareness, if any awareness at all, about the specific keys when he is playing in public, the practitioners that take a consultation course are expected later in their actual work to have at most background awareness of the ten consultation aspects listed. The pianist should when not training give the melodies played a good Gestalt, and the health provider should give all real consultations a good Gestalt. Know-how should be exercised with flow.

5.4 Tacit knowledge and computer science

The computer revolution has affected the discussion of tacit knowledge. On the one side we find determinist philosophers and (many) computer scientists who think that human beings are just a very complex kind of machine-with-computers that we ourselves have not yet been able to build.

For these people, to create a man is only a matter of implementing the right kind of software in a hardware that is capable of processing this software. On the other side we find philosophers and (a few) computer scientists who think that there is something special about human beings that never can be mirrored by any machine or computer whatsoever. That is, to them it is certain that there will never ever be expert systems, artificial intelligences, and robots that will be able to perform exactly like human experts. The main argument of these humans-are-unique defenders can be schematized as follows, using artificial intelligence (AI) as our example:

premise 1:	all artifical intelligencies perform only by means of rule following
premise 2:	expert tacit knowledge cannot be reduced to rule following
premise 3:	human beings can acquire expert tacit knowledge
hence:	----------
conclusion 1:	human beings have a capacity that cannot be reduced to rule following
conclusion 2:	human beings cannot be wholly substituted by robots

Before we make some brief remarks on the debate, let us say some words about experts and stages of know-how. In Chapter 5.2, we presented four different *ways* in which know-how can be improved. Now we will present five *stages* of know-how. At most stages, all the four ways of improving discerned can be useful, but, by definition, there are two exceptions. When someone is the number one expert, he does not need to imitate anyone; and if somebody has reached the absolutely highest possible level, he can’t improve at all. However, the different ways may have a more or less prominent role to play at various stages. According to the American philosophers and AI researchers Hubert and Stuart Dreyfus, skill acquisition relies much on rules (knowing-that) in the lowest stage but not at all on rules in the highest stage. According to the Dreyfus brothers, when adults develop skillful behavior there are five possible emerging stages that ought to be distinguished:

1. novice stage
2. advanced beginner stage
3. competence stage
4. proficiency stage
5. expertise stage.

1. Novice. The novice is instructed by means of strict rules about what to do. Persons that act only by applying such rules work rather slowly; and in many situations their strict rule-following leads to bad or very inefficient actions.

2. Advanced beginner. As the novice gains experience by trying to cope with real situations, he either notes himself or is told by his instructor about various aspects of the situations. The strict rules become transformed into maxims or rules of thumb that the advanced beginner knows how and when to apply. Nonetheless, the actions are performed in a detached analytic frame of mind where the individual thinks about rules and examples.

3. Competence. In this stage the individual is able to note an overwhelmingly number of potentially relevant aspects of various situations. Therefore, apart from the strict rules and the maxims, he starts in many situations to devise plans and perspectives that can determine what aspects are important. He becomes *as a person* involved in his activity. When something goes bad or well he can no longer blame or praise only the rules and maxims, he feels personal responsibility. He can feel remorse for mistakes, and he can experience a kind of elation when being successful. To quote H. Dreyfus (2006): "And, as the competent student becomes more and more emotionally involved in his task, it becomes increasingly difficult for him to draw back and adopt the detached maxim-following stance of the beginner. Only at the level of competence is there an emotional investment in the *choice of action*."

4. Proficiency. The emotional involvement that comes about in the former stage causes an automatic strengthening of successful responses and an inhibition of unsuccessful ones. Thereby, the rules, maxims, and

plans will: “gradually be replaced by situational discriminations, accompanied by associated response. Only if experience is assimilated in this embodied, atheoretical way do intuitive reactions replace reasoned responses (ibid.).”

5. Expertise. “The proficient performer, immersed in the world of his skillful activity, *sees* what needs to be done, but must *decide* how to do it. The expert not only sees what needs to be achieved; thanks to a vast repertoire of situational discriminations he sees immediately what to do. Thus the ability to make more subtle and refined discriminations is what distinguishes the expert from the proficient performer. […] What must be done, simply is done (ibid.).”

According to the Dreyfus brothers, experts simply do not follow any rules, and that is the reason why knowledge engineers who try to develop perfect expert systems are bound to fail. Knowledge engineers use textbook knowledge and try to get experts to articulate their rules and principles for both bodily and intellectual actions – but what the experts or masters of a discipline are really doing is discriminating thousands of special cases. Now, the five Dreyfus-stages give a good description of how things look like from the point of view of the consciousness of the performer. But in itself the description of the last stage begs the question whether or not the brain and the body, unknowingly to the performer, are following extremely complicated rules and are the causes of the actions that are personally experienced as not being instances of rule-following.

(Let us here add that our earlier remarks about ‘creative proficiency’ have an interesting consequence. Traditionally, the philosophy of tacit knowledge is surrounded by an authoritarian aura. Even if an expert sometimes has to say to people on lower stages ‘I cannot tell you why, but this is simply the way we have to act!’, it may turn out that the latter because of creative proficieny was right and the expert wrong.)

The proof of the pudding is in the eating. The limits of the artificial chess players, of the medical expert systems, and of what actions robots can perform are probably to be found empirically. If there will be robots that can bike, then the constructors have to program them to take account of the centrifugal forces that we earlier mentioned.

Simulators and computerized programs may probably in the future be fruitful means when medical novices develop into medical experts; they are already used in certain specialties such as anesthesia and surgery. Also, simulators and computerized programs may be used as time saving tools for the experts. But so far we have not seen any computers that can replace medical experts, be these clinicians or researchers.

5.5 Tacit knowledge and fallibilism

At the end of the nineteenth century, there arose in many Western societies a strong and widespread belief that science affords us certain knowledge, that science progresses linearly, and that the scientific mode of thinking should be generalized to all areas of life. When this 'scientism' became shattered in the late 1960s, some people tried to still their quest for certainty by starting to rely on tacit knowledge instead of science. If science is fallible, they seem implicitly to have argued, we have to rely completely on common sense, practical people, and our own spontaneous feelings of what to do. However, these knowledge sources are equally fallible. As knowing-thats can be more or less truthlike and even completely false, knowing-hows can succeed more or less and even fail completely. If biomedically trained clinicians can – based on expertise know-how – make false diagnoses and give wrong treatments, this is surely equally true for homeopaths and acupuncturists that regard themselves as having know-how expertise within their respective field (compare Chapter 6.4). There is no other way out then to get rid of the quest for absolute certainty. When this is done, one can in a new way retain the trust in both science (knowing-that) and tacit knowledge (know-how). Both, however, have to be regarded as fallible kinds of knowledge.

At the beginning of Chapter 4 on scientific argumentation, we said that we regard *arguments from perception* as a kind of zero point for empirical-scientific argumentation. Later, we have claimed that such arguments rely on structured perceptions that, in turn, rely on fallible tacit knowledge. Knowing-that by means of perception is a kind of know-how. Induction schema and abduction schema, we have also said, are mere forms for inferences that cannot transfer truth from premises to conclusion. Put briefly, observations are theory-laden and dependent on fallible tacit knowledge, and generalizations are empirically underdetermined and

dependent on fallible inductions and abductions. Fallible tacit knowledge seems to be relevant also for inductions and abductions. Such knowledge from experienced scientists can fill the inference schemas with content and in each particular case make them more reasonable, but it cannot possibly take fallibility away.

Reference list

Benner P, Tanner C. Chesia C. *Expertise in Nursing Practice: Caring, Clinical Judgement, and Ethics*. Springer Publishing Company. New York 1995.

Chopra V, Engbers FH, Geerts MJ, Filet WR, Bovill JG, Spierdijk J. The Leiden Anaesthesia Simulator. *British Journal of Anaesthology* 1994 Sep; 73: 287-92.

Csíkszentmihályi M. *Flow: The Psychology of Optimal Experience*. Harper Collins. New York 1991.

Dreyfus H, Dreyfus S. *Mind Over Machine*. Free Press. New York 1986.

Dreyfus H. *What Computers Still Can't Do: A Critique of Artificial Reason*. The MIT Press. Cambrdige Mass. 1992.

Dreyfus H. A Phenomenology of Skill Acquisition as the Basis for a Merleau-Pontian Non-representational Cognitive Science. On Dreyfus' web site 2006.

Hedberg C. The Prismatic Model – how to improve the doctor-patient communication. Manuscript < http://www.sfam.se/documents/prisma-stbrev060822.pdf >.

Hippocrates. *Works by Hippocrates On Airs, Waters, and Places Written 400 BCE*. Translated by Francis Adams. The Internet Classics Archive.

Lehmann ED. The Freeware AIDA Interactive Educational Diabetes Simulator. *Medical Science Monitor* 2001; 7: 516-25

Nyiry JC, Smith B. *Practical Knowledge: Outlines of a Theory of Traditions and Skills*. Croom Helm. London 1988.

Pendleton D, Schofield T, Tate P, Havelock P. *The New Consultation: Developing Doctor-Patient Communication*. Oxford University Press. Oxford 2003.

Polanyi M. *Personal Knowledge*. Routledge & Kegan Paul. London 1958.

Polanyi M. *The Tacit Dimension*. Routledge & Kegan Paul. London 1967.

Polanyi M. *Knowing and Being*. Routledge & Kegan Paul. London 1969.

Rudebeck CE. General Practice and the Dialoque of Clinical Practice. On Symptom, Symptom Presentations, and Bodily Empathy. *Scandinavian Journal of Primary Health Care*. Supplement 1/1992.

Ryle G. *The Concept of Mind*. Penguin Books. Harmondsworth. Middlesex 1966.

Sanders AJ, Luursema JM, Warntjes P, et al. Validation of Open-Surgery VR Trainer. *Stud Health Technol Inform.* 2006; 119: 473-6.

Schön D. *The Reflective Practitioner*. Basic Books. New York 1982.

Schön D. *Educating the Reflective Practitioner*. Jossey Bass. San Francisco 1987.

Stewart M, Belle Brown J, Weston WW, McWhinney IR, McWilliam CI, Freeman TR. *Patient Centered Medicine. Transforming the Clinical Method*. Sage Publications. London 1995.

Wallis C. Consciousness, Context, and Know-how. *Synthese* 2008; 160: 123-53.

6. The Clinical Medical Paradigm

In Chapters 2-5 we have presented epistemological issues. We have distinguished between deductive inferences, inductive support (in the form of induction, abduction, and hypothetico-deductive arguments), and some other kinds of reasoning that can function as arguments in science. We have stressed that inductive support is never pure; empirical data are always theory impregnated. There is no hypothetico-deductive method that can be cut loose from other kinds of arguments, and there is no hypothetico-deductive method that can produce crucial experiments that literally prove that something is a scientific fact. Nonetheless, empirical data constrains what we are able to obtain with the aid of formal logic, mathematics, and theoretical speculations. This makes know-how and tacit knowledge important, too. We have introduced the concepts of paradigms and sub-paradigms and explained how they function from an epistemological point of view. In the next section we will briefly present the modern medical paradigm, i.e., the biomedical paradigm, from an ontological point of view, i.e., we will present what kind of claims this paradigm makes about the structure of the human body, not how we can know whether these claims are true or false. In the ensuing sections we will mix epistemology and ontology and focus on the sub-paradigm that might be called ‘the clinical medical paradigm’. Central here is the randomized controlled trial.

6.1 Man as machine

The view that clinical symptoms are signs of disease processes assumes, implicitly or explicitly, that there are underlying mechanisms that give rise to the symptoms. According to positivism, science always ought to avoid assumptions concerning underlying processes, but we will henceforth take fallibilist epistemological and ontological realism for granted. That is, we will take it for granted that there are mechanisms by means of which we can explain how the body functions, what causes diseases (pathoetiology), and how diseases develop (pathogenesis). Through the ages, medical researchers have asked questions about the nature of these mechanisms,

but never whether or not there are mechanisms. The discovery of mechanisms such as the pump function of the heart and the circulation of the blood, the function of the central nervous system, the way micro-organisms causes diseases, the immune system, DNA and molecular genetic (including epigenetic) theories, and so forth, have been milestones in the development of medical science.

Even though, for instance, Semmelweis' central hypothesis that the mortality rate of childbirth fever should diminish if doctors and medical students washed their hands in a chlorine solution can be evaluated in a purely instrumentalist and positivist manner (i.e., tested without any considerations of assumptions about underlying mechanisms), this is not the way Semmelweis and his opponents looked upon the hypothesis. Both sides also thought in terms of theories of underlying mechanisms. Semmelweis was referring to organic living material (cadaveric matters) and the theory of contagion, and his opponents referred to the miasma theory or theories presupposing other mechanisms.

Modern medicine is part of the scientific revolution and its stress on experiments, observations, and human rationality. When this revolution started, the change was not only epistemological, but also ontological (as we made clear in Chapter 2.3). Nature and human bodies did now become regarded as purpose-less mechanical units. Earlier, the human body was understood in analogy with the appearances of animals and plants. Seemingly, a plant grows by itself with the aid only of water. Superficially seen, there is an internal growth capacity in the plant. Stones do not grow at all. Neither plants nor stones can move of themselves, but animals seem to have such an internal capacity to move to places that fit them. That is, plants seem to have an internal capacity to develop towards a certain goal, become full-grown; animals seem to have the same capacity to grow but also a capacity to move towards pre-determined goals. Dead matter, on the other hand, is just pushed around under the influence of external mechanical causes.

At the outset of the scientific revolution, explanations by final causes were banned from physics. The clock with its clockwork became the exemplar and model for how to understand and explain changes. Behind the visible minute and hour hands and their movements there is the invisible clockwork mechanism that makes the hands move.

The clock metaphor entered medicine later than physics, but already at the beginning of the seventeenth century René Descartes developed an influential ontology according to which the body and the mind existed in different ways, the former in both space and time, but the latter only in time. According to Aristotle and his medieval followers, mind and body are much more intimately interwoven, and mind requires for its existence the existence of a body. According to Descartes, the human body is a machine, but a machine connected to a mind (via the epiphysis). This connection was meant to explain why some processes in the body can produce pain and how willpower can make the body act. Animals were regarded as merely machines. Since they were assumed to lack a connection to a mind, they were also assumed not to be able to feel pain. Thus experimentation on animals was not an ethical issue at all. Descartes even stressed that it was important for medical researchers to rid themselves of the spontaneous but erroneous belief that animals are able to feel pain.

The French Enlightenment philosopher Julien de La Mettrie (1709-1751) created the famous expression ‘Man as Machine’, but his book, *L’Homme Machine,* is not easy to interpret in detail. It is quite clear that he denies the existence of purely temporal substances such as Descartes’ souls, but his human machines do not fit the mechanical cog-wheel metaphor. He also wrote a book ‘Man as Plant’ (*L’Homme Plante),* which he regarded as consistent with the first book. He seems to accept the existence of mental phenomena, but he argues that all spiritual and psychological functions of human beings should be explained solely by means of the function of the body. We will, however, use the expression ‘man-is-a-machine’ in its ordinary sense.

The machine metaphor might be used in order to highlight the paradigm to sub-paradigm relation. If we merely claim that man is a machine, we have not said anything about what kind of machine man is. Within the same paradigm different machine conceptions can compete with and succeed each other. The machine metaphor – with its metaphysical assumptions – has dominated modern medicine to the present day, but new inventions and discoveries have now and then altered the more concrete content. Initially, the body was regarded as on a par with a cog wheel system combined with a hydraulic system. But with the development of the

modern theories of chemistry and electricity and the accompanying inventions, the body was eventually regarded as a rather complicated physico-chemical machine – almost a chemical factory. To regard the brain as a computer is the latest step in this evolution of thought, even though computers are not normally classified as machines. Instead of saying only ‘man is a machine’, we can today say ‘man is a computer regulated moving chemical factory’.

To claim that modern medicine has been dominated by the ‘man as machine’ paradigm is not to claim that clinicians and medical researches in their non-professional lives have looked upon human beings in a way that radically differs from that of other people. It is to stress that in their professional theoretical framework there is no real causal place allotted to psychological phenomena. This notwithstanding, probably only a few physicians have entertained the view that there simply are no mental phenomena at all, and that to think so is to suffer from an illusion; philosophers refer to such a view as ‘reductive materialism’. This ontological position, let us remark in passing, has the incredible implication that even this presumed illusion (the thought that there are mental phenomena) is itself a purely material phenomenon. We wonder how a purely material phenomenon can be an illusion. More popular, especially today, is the view that body and mind are actually *identical* (the main philosophical proponents of this ‘identity theory’ have been the couple Patricia S. and Paul M. Churchland).

However, for medicine the practical consequences of reductive materialism and the identity theory are more or less the same. On neither position is there any need to take mental aspects into consideration when one searches for causes of diseases and illnesses. According to reductive materialism, this would be wholly unreasonable, since there are no mental phenomena; according to the identity theory, this may well be done, but since every mental phenomenon is regarded as identical to a somatic condition or process, there is no special reason to bring in mental talk when discussing causal relations.

There is one mental phenomenon that physicians have always taken seriously: pain. Pain is the point of departure for much medical technology and many therapies. From a pure man-is-a-machine view, anesthetics of all kinds are odd inventions. Therefore, in this respect, physicians have

embraced an ontological position that says that mental phenomena really exist, and that events in the body can cause mental phenomena. Conversely, however, it has been denied that mental phenomena might cause and cure somatic diseases. Mental phenomena such as will power and expectations have played an almost negligible role in modern causal medical thinking, even though the placebo effect is admitted and even investigated by means of neuro-imaging techniques (Chapter 7). Therefore, the ontological position of the traditional biomedical paradigm has better be called ‘epiphenomenalist materialism with respect to the medical realm’. An epiphenomenon is an existing phenomenon that cannot cause or influence anything; it is a mere side effect of something else. Shadows are an example of epiphenomena. A shadow reacts back neither on the body nor on the light source that creates it. According to epiphenomenalist materialism, mental phenomena are real and distinct from material entities, but they are assumed not to be able to react back on any somatic processes.

In the biomedical paradigm, treatments of diseases, illnesses, fractures, and disabilities are often likened to the repairing of a machine; preventions are looked upon as keeping a machine in good shape. The direct causes of the kind of health impairments mentioned can easily be divided into the five main groups listed below. Indirectly, even the immune system can be the cause of some diseases, so-called autoimmune diseases; examples might be rheumatoid fever and some types of hypothyroid (struma or goiter). But here is the list of the direct causes:

1. wear (normal aging as well as living under extreme conditions)
2. accidents (resulting either directly in e.g., fractures and disabilities, or indirectly causing illnesses and diseases by directly causing the factors under points 3 and 4)
3. imbalances (in hormones, vitamins, minerals, transmitter-substances, etc.)
4. externally entering entities (microorganisms, substances, sound waves, electromagnetic waves, and even normal food in excessive doses)
5. bad construction/constitution (genetic and chromosomal factors)
6. idiopathic (note: to talk about ‘idiopathic causes’ is a kind of joke, see below).

Some brief words about the labels in the list:

(1) All machines, the human body included, eventually deteriorate. Some such deterioration and tear seems to be unavoidable and part of a normal aging process, but the development of some degenerative diseases such as Alzheimer's and Amyotrophic lateral sclerosis (ALS) is also regarded as being due to normal deterioration. The results of some kinds of extreme wear, e.g., changes in the muscles and the skeleton that are due to very strenuous physical work, are regarded as proper for medical treatments and prevention measures, too.

(2) Preventive measures against *accidents* do sometimes have their origin in medical statistics, e.g., speed limitations in traffic in order to prevent car-accidents, and the use of helmets when bicycling or when working on construction sites. Being accidentally exposed to pollution, poison, radiations, starvation, or to certain medical treatments that influence the reproductive organs might even have genetic consequences for the offsprings. That is, accidents may even cause bad constitutions (5) in the next generation.

(3) *Imbalances* of body fluids were regarded as the most important causal disease and illness factor in Galen's humoral pathology (Chapter 2.3). When there is talk about imbalances in modern medicine, one refers to the fact that there can be too much or too little of some substance, be it hormones, vitamins, or elements such as sodium and iron; imbalances are lack of homeostasis. They can of course, in their turn, have various causes; many imbalances are regarded as due to either external causes or internal causes such as genetic conditions (bad constructions); or a combination of both. For instance, deficiency diseases like a vitamin deficiency might be due both to lack of a particular vitamin in the diet or due to an inborn error in the metabolic system. Imbalances can also give rise to some psychiatric disorders; for instance, depressions might be caused by a lack of serotonin.

(4) Wear, accidents, and imbalances have played some role in all medical paradigms, ancient as well as modern. But with the emergence of the microbiological paradigm modern medicine for a long time put a particular stress on the fourth kind of causes of health impairments, i.e., on diseases and illnesses caused by *intruding entities* such as bacteria, viruses, parasites, fungi, prions, and poisonous substances. However, our label is

meant to cover also incoming radiation of various sorts and excessive quantity of normal food, alcohol, and sweets (which might result in dental as well as obese sequels).

(5) The label *bad construction* is a metaphoric label. It is meant to refer to diseases and disabilities caused by genetic defects and other inborn properties such as chromosome aberrations. Of course, genetic conditions might interact with external agents and external conditions as illustrated by epigenetic theories. In the last decade, genetic disease explanations have become prominent. The humane genome has been charted and most genetic diseases with a mono-genetic heredity are now genetically well-defined. There are great expectations that genetic mapping, proteomics, and gene therapy will improve the chances of inventing new kinds of specific and individually tailored treatments.

(6) To say, as physicians sometimes do, that the cause of a certain disease is *idiopathic* is not to say that it has a cause of a kind that differs from the five kinds listed. It is merely to say that the medical community does not at the moment know what kind of causes the disease in question has. Literally, in Greek, ‘idiopathic’ means ‘of its own kind’.

6.2 Mechanism knowledge and correlation knowledge

In philosophy of science, realist philosophers sometimes distinguish between two types of theories (or hypotheses), *representational* theories and *black box* theories, respectively (Bunge 1964). When representational theories not only describe static structures but also dynamic systems and events that can function as causes in such systems, they can be called *mechanism* theories. That is, they contain descriptions of mechanisms that explain how a certain event can give rise to a certain effect. For instance, Harvey’s theory immediately explains why the blood stops circulating if the heart stops beating. When a mechanism theory is accepted, we have *mechanism knowledge*. Engineers have mechanism knowledge of all the devices and machines that they invent. In a black box theory, on the other hand, there are only variables (mostly numerical) that are related to each other. If there is a mechanism, it is treated as if it were hidden in a black box. Instead of mechanisms that relate causes and effects, black box theories give us statistical correlations or associations between *inputs* (in medicine often called ‘exposure’) to and *outputs* (‘effects’) of the box. The

looseness or strength of the association is given a numerical measure by means of the correlation coefficient. When a black box theory is accepted, we have *correlation knowledge*.

Figure 1: *Black box model.*

Classical geometric optics provides a good example of correlation knowledge where the correlation coefficient is almost one. The classical mirror formula '$1/d_0 + 1/d_1 = 1/f$' ('d_0' represents the distance between the object and the mirror, 'd_1' the distance between the image and the mirror, and 'f' represents the focal distance of the mirror; see Figure 2) does not say anything at all about the mechanism behind the mirroring effect. Nonetheless, the formula tells us exactly where to find a picture (output) when we know the object distance and the focal distance in question (inputs).

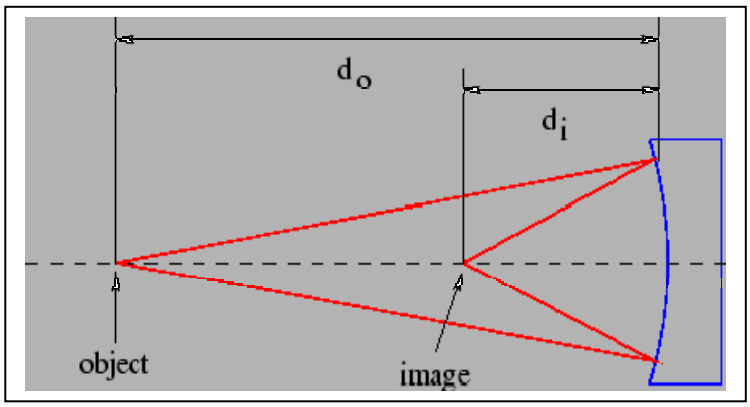


Figure 2: *The mirror formula exemplified.*

In medical correlation knowledge, input may be a certain disease treatment and output recovery or symptom reduction; or input may be exposure to events of a certain kind and output a disease or illness.

Within the biomedical paradigm the traditional ambition has been to acquire mechanism knowledge about the function of the human body, both

in order to understand the genesis of diseases and in order to develop treatments. The mechanisms might be biophysical, microbiological, biochemical, genetic, or molecular. Nonetheless many medical treatments are still based only on correlation knowledge; and the same goes for preventive measures.

Sometimes we do not know anything about the triggering cause (etiology) of a disease but we can nonetheless describe mechanisms that explain the development of the disease process (pathogenesis). We can do it for insulin dependent diabetes, which might be caused by a degenerative process in the beta-cells (in the islands of Langerhans in the pancreas) that are responsible for the internal secretion of the hormone insulin. Although the theories of the causal agents (etiology) are not yet explained – but viruses as well as genetics and auto-immunological reactions have been suggested – the pathogenesis concerning deficiency or accessibility of insulin is rather clear. The mechanisms behind late complications such as vision defects and heart diseases are still not quite clear, although it is known that well-treated diabetes with a stable blood sugar prevents complications.

Another example is the treatment of depression with selective serotonin reuptake inhibitors (SSRIs). This treatment is based on the theory that e.g., depression is caused by an imbalance (lack) of the neurotransmitter serotonin. By inhibiting the receptors responsible for the reuptake of serotonin in the synapses, the concentration of serotonin is kept in a steady state, and the symptoms of depression are usually reduced. Within the modern biomedical paradigm there are theories behind many other treatments, e.g., certain allergic conditions associated with the treatment of antihistamines, different replacement therapies (iron deficiency based anemia, Addison's disease, Graves' disease, etc.), gene therapy by means of virus capsules (although they are not yet quite safe), and of course the treatment of infectious diseases with antibiotics as for example peptic ulcer caused by the Helicobacter pylori bacteria.

Much epidemiological research, however, focuses only on correlation knowledge. Mostly, epidemiologists rest content with finding statistical associations between diseases and variables such as age, sex, profession, home environment, lifestyle, exposure to chemicals, etc. A statistically significant association tells us in itself nothing about causal relations. It

does neither exclude nor prove causality, but given some presuppositions it might be an indicator of causality. So-called lurking variables and confounding factors can make it very hard to tell when there is causality, and the epidemiological world is full of confounding factors. When compulsory helmet legislation was introduced in Australia in the nineties, the frequencies of child head injuries fell. But was this effect a result of the mechanical protection of helmets or was it due to decreased cycling? Such problems notwithstanding, spurious relationships can be detected and high correlation coefficients might give rise to good hypotheses about underlying mechanisms. Thus, improved correlation knowledge can give rise to improved mechanism knowledge. And vice versa, knowledge about new mechanisms can give rise to the introduction of new variables into purely statistical research. In this manner, *mechanism knowledge and correlation knowledge cannot only complement each other, but also interact in a way that makes both of them grow faster than they would on their own*.

Above, we have simplified the story a bit. Many medical theories and hypotheses are neither pure mechanism theories nor pure black box theories, but *grey box* theories. They contain significant correlations between some variables that are connected to only an outline of some operating mechanism. For instance, the associations found between lung cancer and smoking have always been viewed in the light of the hypothesis that a certain component (or composition of various components) in tobacco contains cancer-provoking (oncogenic) substances. But this mechanism has not played any part in the epidemiological investigations themselves. In a similar way, in the background of much clinical correlation research hovering are some general and unspecific mechanism hypotheses.

Due merely to the fact that they are placed in the biomedical paradigm, statistical associations between variables often carry with them suggestions about mechanisms. By means of abduction, one may then try to go, for instance, from correlation knowledge about the effect of a medical treatment to knowledge about an underlying mechanism. We repeat, knowing-that in the forms of mechanism knowledge and correlation knowledge can interact and cooperate – both in the context of discovery and in the context of justification. If we combine this remark with the

earlier remark (Chapter 5.3) about interaction between knowing-that and know-how, we arrive at the overview in Figure 3.

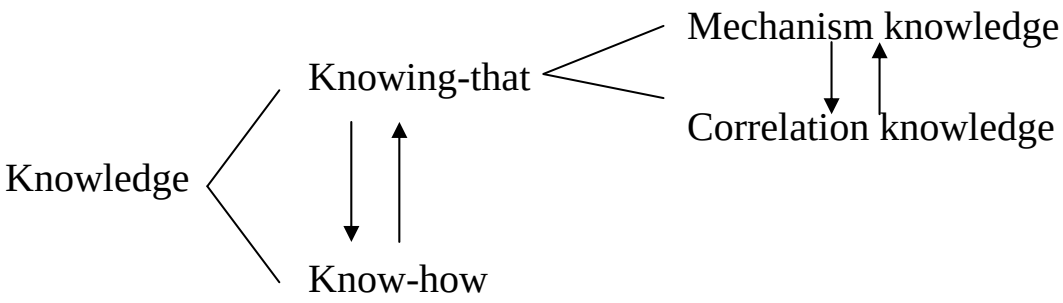


Figure 3: *Interactions between different forms of knowledge.*

In order to illustrate what it can mean to speculate about mechanisms starting from correlation knowledge, let us present a somewhat funny example. A Danish art historian, Broby-Johansen, once observed that there is a correlation between the US economy and the length of women's skirts during 1913-1953. During good times skirts were shorter, during recessions longer (Figure 4).

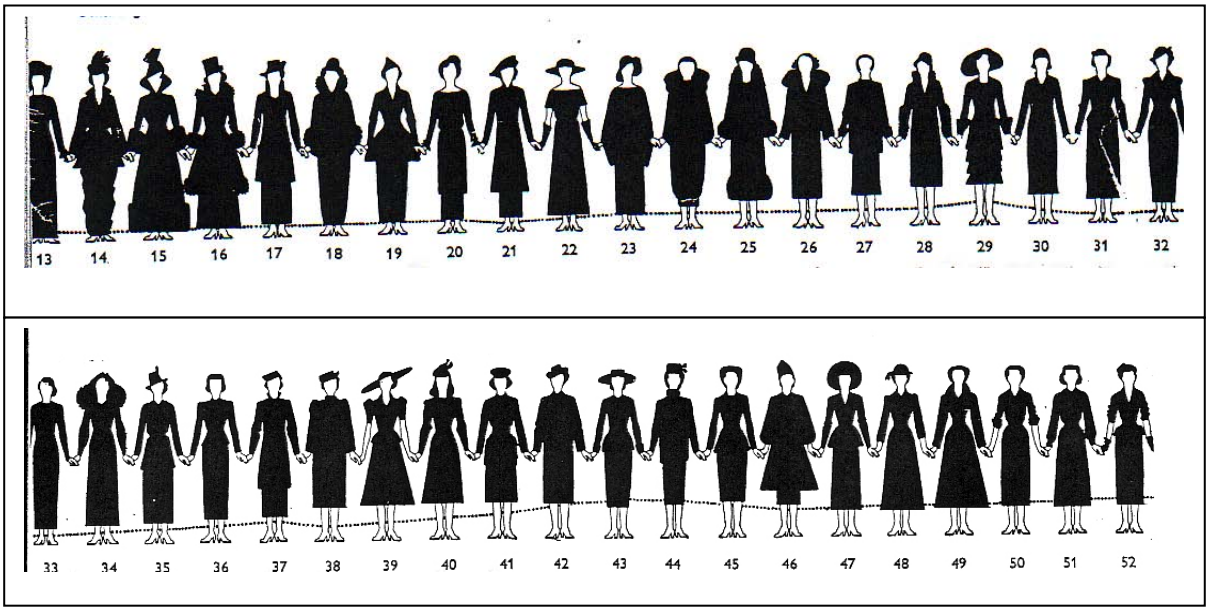


Figure 4: *Picture from Broby-Johansen R. Body and Clothes.*

Should we regard this association as merely a random phenomenon or is there a confounding variable and real mechanism? A possible

psychological mechanism might be this one: during economic recessions people are more frightened and cautious than in good times, and cautious people may feel more protected and secure when wearing more clothes; thus an economic slump might result in long skirts and vice versa.

Within the biomedical paradigm, attempts have been made to lay down a definitive list of criteria for when statistical associations can be regarded as signs of causal relations and mechanisms. The most well known list is the one launched by the English epidemiologist Austin Bradford Hill (1897-1991); a pioneer in randomized clinical trials and in studying lung cancer and cigarette smoking. In-between Hill's list and Robert Koch's four postulates specified for a mono-causal setting (Chapter 2.3) there were several proposals none of which became really famous. Hill worked out his criteria when discussing whether a certain environmental exposition might be interpreted as disease cause. He published his list in 1965, and we will present it below. But first some words of caution.

Hill did not himself use the term 'criteria' but talked of a list of 'viewpoints'. Nonetheless his list has in many later presentations taken on the character of supplying necessary or sufficient conditions for inferring a causal relation from a statistical association. But, surely, his 'criteria' cannot supply this. Today, the question is rather whether or not they can be used even as default rules or 'viewpoints'.

1. *Strength*. Strong statistical associations are more likely to be causal than weak ones. Weak associations are more likely to be explained by undetected biases or confounding variables, but a slight association does not rule out the possibility of a causal relation.

2. *Consistency*. If different investigations, conducted at different times and places and on different populations, show approximately the same association, then it is more likely that there is a causal relation than if merely one investigation is made. Lack of consistency does not rule out a causal connection, since most causes only work in certain circumstances and in the presence of certain cofactors.

3. *Specificity*. This criterion requires that one kind of cause produces only one kind of specific effect in a specific group of people, e.g., lung

cancer among a certain group of workers exposed to a certain substance. As a general rule, it is clearly invalid. Absence of specificity does not rule out the presence of a cause since diseases may have more than one cause.

4. *Temporality*. A cause must precede an effect in time. According to Hill, medical researchers should ask: 'What is the cart and what is the horse?' For instance, 'Does a particular diet lead to a certain disease or do the early stage of this disease lead to these peculiar dietic habits?' However, it is hard to rule out completely the possibility of cause and effect being simultaneous.

5. *Biological gradient*. There should be a unidirectional dose-response curve; more of a dose should lead to more of the response. The more cigarettes an individual smoke per day the more likely it is that he will die of lung cancer or chronic obstructive lung disease; the death rate is higher among smokers than non-smokers. However, the absence of such a dose-response relationship does not rule out a causal association.

6. *Plausibility*. The causal relation imposed should fit into the contemporary biomedical paradigm and the general mechanisms that it posits, i.e., be biologically plausible. The absence of plausibility does not exclude causality, and the association found might be important in order to develop new causal hypotheses.

7. *Coherence*. The idea of causation imposed should not make the association come into conflict with current knowledge about the disease. For example, the association between smoking and lung cancer is coherent with our knowledge that smoking damages bronchial epithelium. The absence of coherence does not imply absence of causality.

8. *Experimental evidence*. The strongest support for causation comes from experiments (e.g., preventive intervention studies) where the presumed causal agent can be introduced (whereupon the effect should rise or be strengthened) and removed (whereupon the effect should disappear or be weakened).

9. *Analogy*. If in a previous analogous investigation a causal relation is found, this makes it more likely that even the present association mirrors a causal relation. Hill says that since we know that the sedative and hypnotic drug thalidomide can cause congenital anomalies, it is likely that strong associations between other drugs and such anomalies are signs of causal relations, too.

Some of Hill's criteria are stronger and some weaker, but it is hard to rank them hierarchically, and this was not Hill's idea. He thought they should be considered together and might be pragmatically helpfully when assessing whether causality or non-causality was present. So interpreted, the list fits well into some of the very general views that we have propounded: (i) correlation knowledge and mechanism knowledge should interact, and (ii) arguments should be connected to each other as threads in a wire.

We would like to compare Hill's list with another one that was put forward after a controversy (at Karolinska Institutet, KI, in Stockholm) in 1992. The KI-scientist Lars Mogensen then articulated and ranked some principles for the assessment of certain empirical data as stated below. He gives Hill's plausibility criterion (5) the first place, but he also presents sociological criteria:

1) A reasonable theoretical mechanism should be available.
2) A sufficient number of empirical data should be presented – the statistical power of the study should be appropriate.
3) A research group of good reputation should support or stand behind the work.
4) Possible economic ties or interests of the researchers should not be hidden; an implementation of the results should not provide the concerned researcher with financial benefit.
5) Several independent studies should have given approximately the same results.
6) Certain methodological requirements on reliability and validity should be fulfilled; when possible, studies should be experimental, prospective, randomized, and with blinded design.

7) In case referent/control studies, i.e., retrospective studies of patients that have happened to become exposed to a certain hypothetical causal factor, one should make clear that the researchers have not influenced this hypothetical causal factor.

These principles and/or criteria overlap with Hill's list. The first principle is, as we said, similar to Hill's plausibility criterion; the second one concerning statistical power is similar to Hill's strength criterion; and the fifth one reflects Hill's consistency requirement. The sixth and seventh principles are general demands for objectivity. Principles 3 and 4 bring in the social dimension of science. The third brings in trust and ad hominem arguments, and the fourth brings in mistrust and negative ad hominem arguments (Chapter 4.1).

These principles/criteria might be looked upon as a concretization of what it means to be rational in medical research. We wanted to make it clear that they allow for, and even to some extent require, interplay between mechanism theories, black box theories, and grey box theories. True positivists can accept only black box theories; fallibilist realists can and should accept all three kinds of theories.

So far we have used the concept of causality without any philosophical comments apart from the earlier remark (Chapters 3.4 – 3.5), that we think that Hume's and the positivists' reduction of causality to correlation cannot be defended. But now we will make some remarks about the ordinary (non-positivist) concept of cause; these remarks will also explain why in general we prefer to talk of 'mechanism knowledge' instead of 'causal knowledge'.

In everyday language the concept of cause is seldom used explicitly but often implicitly; it is hidden in words such as 'be*cause*', 'so', and 'therefore'. Here are some such sentences: 'the light went on *because* I turned the switch'; 'it became very cold *so* the water turned into ice'; 'a ball was kicked into the window, *therefore* it broke'. Whatever is to be said of causality in relation to the basic laws of physics, in medical contexts causality talk has the same character as everyday causality talk. Superficially seen, such talk only relates either an action (cause) to an event (effect) or one event (cause) to another event (effect). But a moment's reflection shows that in these examples, as normally in common

sense, much is taken for granted. There are in fact almost always many causal factors in play. If the electric current is shot down, my turning of the switch will not cause light; if the water has a heater placed in its middle, the cold weather will not cause it to freeze; if the ball had been kicked less hard, or if the glass had been a bit more elastic, then the ball hitting the window would not have caused it to break. If there is a time delay between the cause and the effect (‘he got ill and vomited because he had eaten rotten food’, ‘he got cancer because several years ago he got exposed to high doses of radioactivity’, etc.), there must obviously be a mediating mechanism.

Communication about causes functions smoothly because, mostly, both speakers and listeners share much tacit knowledge about the causal background factors. What is focused on as *the* causal factor is often a factor the researcher (i) is especially interested in, (ii) can influence, or (iii) finds unusual; or (iv) some combination of these things. But can we give a more philosophical analysis of what a causal factor psychologically so chosen may look like? Yes, we can. We would like to highlight an analysis of causality made by both a philosopher and (independently) by two modern epidemiologists. We will start with the views of the philosopher, L. J. Mackie (1917-1981). He reasons (in our words and examples) as follows about causality. He starts with two observations:

1. One kind of effect can normally be caused by more than one kind of cause, i.e., what is normally called cause is not necessary for the coming into being of the mentioned effect. For example, dying can be caused by normal aging, by accidents, and by intruding poisonous substances.
2. What is normally called ‘the cause’ is merely one of several, or at least two, kinds of factors, i.e., what is normally called cause is not *sufficient* to cause the mentioned effect. For example, infectious diseases require both bacteria and low degree of immunity.

In order to capture these two observations and their implications in one single formulation, Mackie says that an ordinary cause is an *INUS-condition*, and he explains his acronym as follows:

What is normally called a cause (e.g., a bacterium) is a kind of condition

I: such that it is in itself **I**nsufficient to produce (cause) the effect (e.g., a disease);

N: such that it is **N**ecessary that it is part of a certain complex if this complex shall be able to produce by itself the effect;

U: such that the complex mentioned is **U**nnecessary for the production of the effect;

S: such that the complex mentioned is **S**ufficient for the production of the effect.

An INUS-condition (ordinary cause) is neither a sufficient (I) nor a necessary (N) condition for an effect, and it is always part of a larger unit (U and S). The simple relation ‘events of kind C causes events of kind E’ abstracts many things away. What is normally called a cause, and by Mackie an INUS-condition (‘events of kind C constitute an INUS-condition for events of kind E’), can always be said to be part of a mechanism.

Having now introduced the concept of INUS-condition, we would like to immediately re-baptize it into ‘component cause’ (‘events of kind C are component causes for events of kind E’). This term comes from the epidemiologists K. J. Rothman and S. Greenland. Like Mackie, they stress that ordinary causality is multicausality (follows from Mackie’s U and S), and they distinguish between ‘sufficient causes’ and ‘component causes’. As far as we can see, their ‘component cause’ is an INUS-condition in Mackie’s sense. Only such causes are of interest in medical science, since, as they claim: “For biological effects, most and sometimes all of the components of a sufficient cause are unknown (R&G, p. 144)”. When they shall explain what a ‘sufficient cause’ is, they find it natural to use the word ‘mechanism’:

> A “sufficient cause,” which means a complete causal mechanism, can be defined as the set of minimal conditions and events that inevitably produce disease; “minimal” implies that all of the conditions or events are necessary to that occurrence (R&G, p. 144).

We guess that many philosophically minded medical students have found the distinction between etiology and pathogenesis hard to pin down. In our opinion, this difficulty is merely a reflection of the more general problem of how to relate causal talk and mechanism talk to each other. Waiting for future scientists and philosophers to make this relationship clearer, we will in this book continue to talk about both causes and mechanisms the way we have done so far.

In summary, (what we normally call) causes are component causes and parts of mechanisms. Component causes are out of context neither sufficient nor necessary conditions for their effects. The same goes for so-called criteria for regarding a statistical association as representing a causal relation. Such criteria are neither sufficient nor necessary conditions for a causal inference. Fallibilism reigns, and fallible tacit knowledge is necessary in order to come to a definite conclusion. However, the facts highlighted do not make assessments of empirical studies impossible or causal knowledge useless. Two quotations about this:

> Although there are no absolute criteria for assessing the validity of scientific evidence, it is still possible to assess the validity of a study. What is required is much more than the application of a list of criteria. [...] This type of assessment is not one that can be done easily by someone who lacks the skills and training of a scientist familiar with the subject matter and the scientific methods that were employed (R&G, p. 150).

> If a component cause that is neither necessary nor sufficient is blocked, a substantial amount of disease may be prevented. That the cause is not necessary implies that some disease may still occur after the cause is blocked, but a component cause will nevertheless be a necessary cause for some of the cases that occur. That the component cause is not sufficient implies that other component causes must interact with it to produce the disease, and that blocking any of them would result in prevention in some cases of diseases. Thus, one need not identify every component cause to prevent some cases of disease (R&G, p. 145).

6.3 The randomized controlled trial

In physics and chemistry, it is sometimes possible to manipulate and control both the variables we want to investigate and those we want to neglect. Take for example the general gas law: $p \cdot V = n \cdot R \cdot T$. If we want to test if pressure (p) is related to temperature (T) the way the law says, we can make experiments where the volume (V) is kept constant, and if we want to check the relationship between pressure and volume, we can try to keep the temperature constant.

When the gas law with its underlying metaphysics is taken for granted, we can try to control the experimental conditions and systematically vary or keep constant the values of the different variables. This is not possible in medical research. When using rats as experimental objects we have some variables under control (at least some genetically dependent variables) and it is possible to vary the exposure (e.g., different unproven medical treatments), but human beings can for ethical reasons not be selected or exposed the way laboratory animals can. When it comes to experimentation with human beings, we have to conduct our research in special ways. We cannot just expose the research subjects to standardized but dangerous diseases or injuries and then provide treatment only to half of the group and compare the result with the other half in order to assess the effect of the intervention. In clinical research, we have to inform and ask patients before we include them in a trial; and we can only include those who happen to visit the clinic, be they young or old, male or female, large or small, smokers or non-smokers, having suffered from the disease a short time or a long time, etc.

When clinical trials are made in order to determine whether or not a certain medical or surgical treatment is effective, a special design called the ‘randomized controlled trial’ (RCT) is used; it also goes under the name of ‘randomized clinical trial’. In its simplest form, a RCT compare only two groups of patients with the same disease (illness, fracture or disability), which are given different types of treatment. Since patients should always be given the best known treatment, the new treatment to be tested must be compared with the old routine treatment, if there is one. The patients that participate in the investigation are allocated to their group by means of some kind of lottery or random numbers. The group that contains

patients that are treated with the new, and hypothetically better, treatment is called ‘the experimental group’; the other one is called ‘the comparison group’ or ‘the control group’.

When there is no routine or standard treatment available, it is considered acceptable to ‘treat’ the control group with dummy pills or placebos. Therefore, this group is sometimes referred to as the placebo group. Placebos are supposed to have no pharmaceutical or biomedical effect – often the placebos are sugar or calcium based. Nevertheless, the placebo treatments should look as similar as possible to the real treatments. If it is a matter of pills, then the placebo pills should have the same color, size, and shape as the real pills; they even ought to smell and taste the same.

Placebo RCTs can be designed as *open*, *single-blinded*, or *double-blinded*. The double-blinded design is from an epistemological point of view the best one. It means that neither the researchers nor the patients know who receives the placebo treatment and who does not. This means that the researchers cannot, when evaluating hard cases, be misled by any wishes (Baconian ‘idols’) that the new treatment is very good and will promote their careers, and the patients’ reactions cannot be influenced by beliefs such as ‘I won’t be better, I am only receiving placebos’. In single-blinded designs only the patients are blinded, and in open designs none. Mostly, when surgical treatments and psychotherapies are assessed, the trial has to be open or single-blinded.

Concretely, the blinding of the clinician giving the treatment can consist in creating a hidden code list that numbers all the treatments without telling the clinician what numbers are connected to placebos. When the study is completed and the effects and non-effects have been observed and registered, the code list is revealed and the effects in the experimental group and the control group, respectively, can be compared. Afterwards it is also possible to examine whether the randomization made in fact resulted in two groups that have similar distributions on variables such as age, sex, duration of the disease before treatment, side-effects, and lifestyle (e.g. smoking and drinking habits) that can be assumed to influence the results.

A more old-fashioned way of assessing the effect of a new treatment is to compare the experimental group with an historical group. The patients in such a control group have earlier received the prevailing routine treatment

without any special attention and extra examinations, or they have simply not been given any specific medical treatment at all. Some studies claim to show that the placebo effect in historical control groups is on average 30% lower than in comparable prospective control groups. This is the reason why studies using historical groups are not regarded as good as ordinary RCTs.

Surprisingly for many people, the effect in the control group (CG) is seldom negligible, and since there is no reason to believe that the same effect does not also occur in the experimental group (EG), one may *informally* (i.e., deleting some important problems of statistics) say that in RCTs the biomedical effect is defined as the difference between the effects in two groups:

Biomedical treatment effect (B) =
Total effect in EG (T) – Non-treatment effect in EG (N)
(N might be regarded as being approximately equal to the total effect in CG).
That is: B = T – N.

The real statistical procedure looks like this. The point of departure for the test is a *research hypothesis*, e.g., a hypothesis to the effect that a new treatment is more effective than the older ones or at least equally effective, or a hypothesis and suspicion to the effect that a certain substance can cause a certain disease. Such research hypotheses, however, are tested only indirectly. They are related to another hypothesis, the so-called *null hypothesis*, which is the hypothesis that is directly tested. The word ‘null’ can here be given two semantic associations.

First, the null hypothesis says that there is nothing in the case at hand except chance phenomena; there is, so to speak, ‘null’ phenomena. That is, a null hypothesis says that the new treatment is *not* effective or that the suspected substance is *not* ‘guilty’.

Second, when a null hypothesis is put forward in medical science, it is put forward as a hypothesis that one has reason to think can be ‘nullified’, i.e., can be shown to be false. In clinical trials, the null hypothesis says that there is no statistically significant difference between the outcome measure in the control group and the experimental group. The opposite hypothesis, which says that there actually is a significant difference, is called the

counter-hypothesis. It is more general than the research hypothesis, which says that the difference has a certain specific direction. When the empirical data collecting procedure has come to an end, and all data concerning the groups have been assembled, a combination of mathematical-statistical methods and empirical considerations are used in order to come to a decision whether or not to reject the null hypothesis.

RCTs are as fallible as anything else in empirical science. Unhappily, we may:

- *reject* a null hypothesis that is *true* ('type 1 error') and, e.g., start to use a treatment that actually has no effect;
- *accept* a null hypothesis that is *false* ('type 2 error') and, e.g., abstain from using a treatment that actually has effect.

In order to indicate something about these risks in relation to a particular investigation, one calculates the *p-value* of the test. The smaller the p-value is, the more epistemically probable it is that the null hypothesis is false, i.e., a small p-value indicates that new treatment is effective. A *significance level* is a criterion used for rejecting the null hypothesis. It states that in the contexts at hand the p-value of the tests must be below a certain number, often called α. Three often used significance levels state that the p-value must be equal to or smaller than 0.05, 0.01, and 0.001 (i.e., 5%, 1%, and 0.1%), respectively.

From the point of view of ordinary language this terminology is a bit confusing. What has been said means that tests that pass *lower significance levels* have an outcome that is *statistically more significant*, i.e., they indicate more strongly that the treatment works. In other words, the lower the p-value is the higher the significance is.

There is also another confusing thing that brings in difficult topics that we discussed in relation to the question of how to interpret singular-objective probability statements (Chapter 4.7). Now the question becomes: 'How to interpret the claim that one singular test has a certain p-value and level of significance (α)?' Does this p-value represent a real feature of the test or is it merely a way of talking about many exactly similar tests? In the latter case, we may say as follows. A significance level of 0.05 means that if we perform 100 randomized controlled trials regarding the specific

treatment under discussion, we are prepared to accept that mere chance produces the outcomes in five of these trials. That is, we take the risk of wrongly rejecting the null hypothesis in five percent of the trials. To accept a significance level of 0.001 means that we are prepared to wrongly reject a null hypothesis in one out of a thousand trials.

There are many different ways of dividing statistical hypothesis testing into stages, be these chronological or merely logical. From an epistemological point of view, five stages might profitably be discerned. They make visible the complex interplay that exists between empirical (inductive) and logical (deductive) factors in statistical hypothesis testing.

Stage 1. One makes a *specific* null hypothesis that tells what pure chance is assumed to look like. From a mathematical perspective, one chooses a certain mathematical probability function as representing the null hypothesis. One may choose the binomial probability distribution, the standard normal distribution (the famous bell curve), the Poisson distribution, the chi-square distribution (common in medicine, at least when results are categorized in terms of alive or died, yes or no etc), or some other distribution that one has reason to think in a good way represents chance in the case at hand. From an epistemological perspective, *hereby, one makes an empirical assumption.* It is often said that the null hypothesis is 'statistically based', but this does not mean that it is based only on mathematico-statistical considerations, it is based on empirico-statistical material, too. This material can even be identical with the sample that is to be evaluated. In simple cases, symmetry considerations are enough.

(If it is to be tested whether the density of a certain 'suspected' die is *a*symmetrically distributed, the 'pure chance function' says that all sides have a probability of 1/6, and the null hypothesis is that this function represents the truth, i.e., that the die has its density symmetrically distributed.)

Stage 2: One chooses a determinate number, α, as being the significance level for the test. This choice has to be determined by *empirical considerations* about the case at hand as well as about what kind of practical applications the investigations are related to. Broadly speaking,

significance levels for tests of mechanical devices are one thing, significance levels for medical tests another. What α-value one accepts is a matter of convention; this fact has to be stressed. There is no property 'being statistically significant' that can be defined by purely statistical-mathematical methods; from a mathematical point of view there is only a continuum of p-values between zero and one. In analogy with the talk about 'inference to the best explanation', one may in relation to statistical tests talk of an 'inference to the best significance level'.

Stage 3: One makes *the empirical investigation*. That is, one selects experimental and control groups, and does what is needed in order to assemble the statistical observational data (the sample) searched for.

Stage 4: One compares the statistical empirical material collected with the null hypothesis. On the *empirical assumption* that the specified null hypothesis is true (and the *empirical assumption* of independence mentioned in Chapter 4.7), one can *deduce* (more or less exactly) what the probability is that the sample has been produced by pure chance. This probability value (or value interval) is the p-value of the test. Since it is deduced from premises that contain empirical assumptions, *to ascribe a p-value to a test is to make an empirical hypothesis.*

(When the null hypothesis is represented by a chi-square distribution, one first calculates a so-called chi-square value, which can be regarded as being a measure of how well the sample fits the chi-distribution and, therefore, is in accordance with the null hypothesis. Then, in a second step, one *deduces* the p-value from the chi-square value. In practice, one just looks in a table, or the results from a statistical software program, where the results of such already made calculations are written down.)

Stage 5: One makes a *logical comparison* between the obtained p-value and the α-value that has been chosen as significance level. Depending on the result, one decides whether or not to regard the null hypothesis as refuted.

Unhappily, we cannot always simply choose high significance levels (= small values of α), say $\alpha = 0.001$. We may then wrongly (if $p > 0.001$)

accept a false null hypothesis (commit a type 2 error), i.e., reject a treatment with usable effect. With a high significance level (small α) the risk for type 2 error is high, but with a low significance level (high α) the risk for type 1 error (acceptance of a useless treatment) is high. Since we want to avoid both errors, there is no general way out. Some choices in research, as in life in general, are simply hard to make.

Going back to our little formula, B = T – N, the null hypothesis always says that there is no difference between T and N, i.e., no difference between the result in the group that receives treatment (or the new treatment) and the one that recieves no treatment (or the old treatment). If we are able to reject the null hypothesis, then we claim that there is inductive support for the existence of an effect, B. This biomedical effect is often called just 'the *real* effect', but of course the non-treatment effect is also a real effect. However, it is only the biomedical effect that has real clinical relevance.

Our very abstract formula leaves skips over one important problem. The variable for non-treatment (N) does not distinguish between a placebo curing and *spontaneous* or *natural* bodily healing processes, be the latter depending on the immune system or something else. The spontaneous course of a disease is the course this will follow if no treatment is provided and no other kind of intervention is made. Since some diseases in some individuals are spontaneously cured, it may be the case that parts both of the estimated total effect (T) and the estimated non-treatment effect (N) are due to the body's own healing capacity. This capacity may in turn vary with factors such as genetic background, age, and life style, and this makes it hard to observe the capacity in real life even though it is easy to talk about such a capacity theoretically. We will return to this issue in the next chapter.

Even if we disregard the placebo effect and spontaneous bodily curing, there remains a kaleidoscope of other factors complicating clinical research. Pure biological variables, known as well as unknown, might interfere with the treatment under test. This means that even if there is no non-treatment effect in the sense of a mind-to-body influence, it is necessary to randomize clinical trials.

Most sciences have now and then to work with simplifications. They are justified when they help us to capture at least some aspects of some real

phenomena. In the case of RCTs, it has to be remembered that the patients we can select in order to make a trial may not allow us later to make inferences to the whole relevant population. RCTs can often be conducted only in hospital settings; old and seriously ill patients have often to be excluded; and the same goes for patients with multiple diseases and patients receiving other treatments. Furthermore, some patients are simply uncooperative or display low compliance for social or other reasons. Accordingly, it is often the case that neither the experimental group nor the control group can be selected in such a way that they become representative of all the patients with the actual diagnosis. Nonetheless, accepting fallibilism, they can despite the simplification involved give us truthlike information about the world.

In their abstract form, the RCTs are complex hypothetico-deductive arguments (Chapter 4.4). First one assumes hypothetically the null hypothesis, and then one tries to falsify this hypothesis by comparing the actual empirical data with what follows deductively from the specific null hypothesis at hand. At last, one takes a stand on the research hypothesis. As stated above, the same empirical data that allow us to reject the null hypothesis also allow us to regard the research hypothesis as having inductive support. The rule that one should attempt to reject null hypotheses has sometimes been called ‘quasi falsificationism’. (This procedure is not an application of Popper’s request for falsification instead of verification; see Chapter 4.4. Popper would surely accept RCTs, but in his writings he argues that one should try to falsify *already accepted research hypotheses*, not null hypotheses.)

Within the clinical medical paradigm, various simple and pedagogical RCTs constitute the exemplars or the prototypes of normal scientific work. The assessment of a new medical technology, e.g., a pharmaceutical product, is a normal-scientific activity in Kuhn’s sense. The articulation of the problem, the genesis of the hypothesis, the empirical design used, the inclusion and exclusion criteria used, the significance level chosen, the analysis made, and the interpretation made, all take place within the framework of the biomedical paradigm and its theoretical and methodological presuppositions. This might appear trivial, and researchers might be proficient researchers without being aware of all these theoretical preconditions, but they are necessary to see clearly when alternative

medical technologies are discussed and assessed. The reason is that alternative medical technologies may have theoretical presuppositions that are in conflict with those of the biomedical paradigm. This is the topic of the next section, 6.4, but first some words on RCTs and illusions created by *publication bias*.

The concept of 'bias' has its origin in shipping terminology. A vessel is biased when it is incorrectly loaded. It might then begin to lean, and in the worse case scenario sink. A die becomes biased if a weight is put into one of its sides. A RCT is biased if either the randomization or the blinding procedure is not conducted properly – resulting in e.g., *selection bias*. Such bias distorts the result of the particular RCT in question. Publication bias (of RCTs) occur when RCT-based knowledge is distorted because the results of all trials are not published; it is also called 'the drawer effect'.

It is well known that negative results of RCTs (i.e., trials where the null hypothesis is not rejected) are not published to the same extent as positive results, i.e., trials indicating that the medical technology in question has a usable effect. One reason is of course that it is more interesting to report that a new proposed medical technology is effective than to report that it fails (the null hypothesis could not be rejected). Another casual reason is related to the sponsoring of the study. If it is publicly known that a pharmaceutical company expects to obtain support from an on-going study, and the results turn out to be negative, then the company might try to delay or create obstacles for publication. Now, if all RCTs concerned with same treatment would always give the same result, such publishing policies would be no great problem. But, unhappily, they do not. Therefore, we might end up in situations where only studies reporting positive results have been published despite the fact that several (unpublished) trials seem to show that the treatment does not work. This problem becomes obvious when Cochrane collaboration groups are conducting meta-analyses (see Chapter 4.1). In meta-analyses one tries to take account of all RCTs made within a certain research field. If negative results are not reported the meta-analyses give biased results; at the very worst, they allow a new treatment to be introduced despite strong indications that the opposite decision should have been taken.

Despite all problems in connection with the practical conductions of RCTs and their theoretical evaluations, they represent a huge step forward

in comparison to the older empirical methods of casual observations, personal experience, and historical comparisons.

6.4 Alternative medicine

What is alternative medicine? In the 1970s and earlier, one could delineate it by saying: ‘presumed medical treatments that are not taught at the medical faculties at the universities in the Western world, and which are not used by people so educated’. This answer is no longer accurate. Nowadays some traditionally educated physicians have also learnt one or a couple of alternative medical methods, and even at some universities some such methods are taught. This situation has given rise to the term ‘complementary medicine’, i.e., alternative medicine used in conjunction with traditional medicine. However, there is a characterization of alternative medicine that fits both yesterday’s and today’s practical and educational situations:

- Alternative medical treatments are treatments based on theoretical preconditions that diverge from the biomedical paradigm.

We would like to distinguish between two kinds of alternative medicine: somatic and psychosomatic. The first kind consists of therapies such as acupuncture, homeopathy, and chiropractic (or osteopathic manipulation in a more general sense); and these have just as much a somatic approach to diseases, illnesses, and disabilities as the biomedical paradigm. Remember that this paradigm allows causality directed from body to mind; for instance, it allows somatic curing of mental illnesses and psychic disorders. In the psychosomatic kind of alternative medicine, activities such as yoga, meditation, and prayers are used as medically preventive measures or as direct therapies. We will mention this kind only in passing in the next chapter when discussing placebo effects. Now we will, so to speak, *discuss how to discuss* somatic alternative medical therapies; we will use acupuncture and homeopathy as our examples. Here, the distinction between correlation and mechanism knowledge will show itself to be of crucial importance.

From most patients' point of view, acupuncture is a black box theory. Inputs are needles stuck into the body on various places, outputs are (when it works) relief or getting rid of illnesses, diseases, or symptoms. However, classic Chinese acupuncture (in contrast to some modern versions) lays claim to have mechanism knowledge. According to it, a certain kind of energy (chi) can be found and dispersed along a number of lines or so-called 'meridians' in the body (Figure 5). All acupuncture points are to be found somewhere on these meridians, and all causal relationships are propagated along them. An acupuncture needle in an acupuncture point can only affect something (e.g., pain) that is in some way connected to the same meridian as the needle in question. One may well think of these meridians in analogy with old telephone cables. As before the mobiles were invented, it was impossible to call someone to whom one was not linked to with a number of connected telephone cables, an acupuncture needle needs a 'chi energy cable' in order to come in contact with the illness in question.

The problem with the meridian mechanism theory of acupuncture is that we cannot find any such lines or 'chi energy cables' in the body. Once they could with good reasons be regarded as unobservables in the way many entities in physics have been regarded as unobservable entities, but today, when modern surgery makes it possible to make direct observations in the interior of living bodies such a view cannot be defended. Neither anatomically, nor microscopically, nor with X-ray technology, nor with functional magnetic resonance imaging (fMRI) or positron emission tomography (PET) scanning has it been possible to find these channels of energy – although fMRI actually shows that acupuncture causes an increased activation (blood flow) in the brain. We have to draw the conclusion that the presumed mechanism knowledge of acupuncture is false. But this does not imply that there is no useful correlation knowledge to retain. Let us explain.

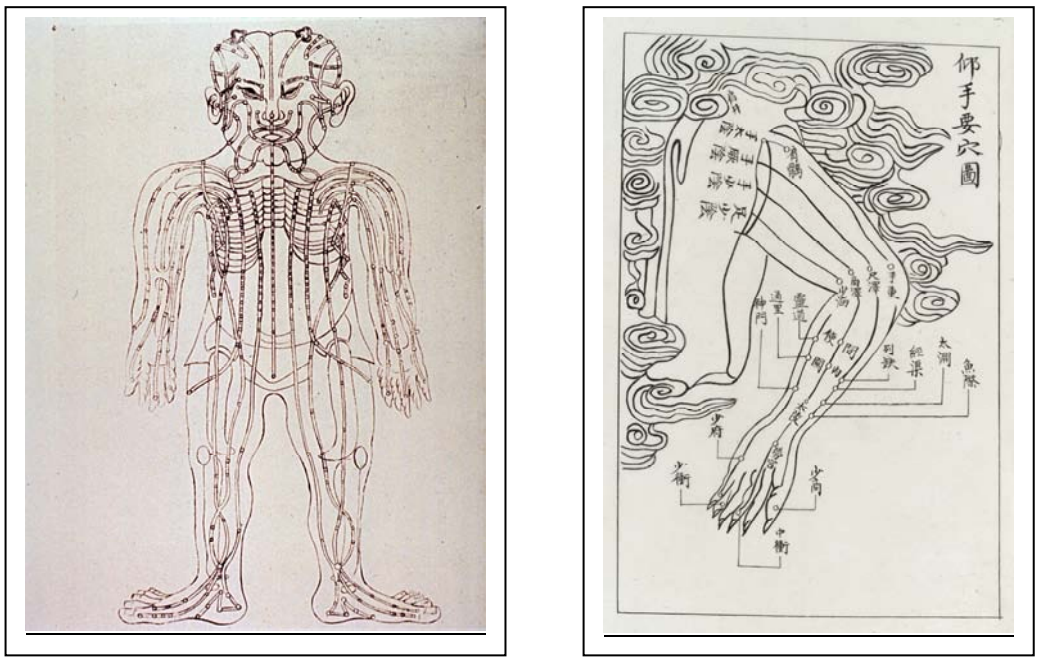


Figure 5: *Acupuncture meridians in old Chinese drawings.*

Acupuncture can be assessed as a black box theory based medical technology. We simply so to speak *de-interpret* the classical theory. That is, we put the chi energy and meridian mechanism in a black box and start to think about acupuncture in the same way as skeptical patients do: ‘the acupuncturist puts needles in my body, I hope this will make my pain go away’. And then we can make statistics on how often the treatment is successful and not. Such investigations can be made where the acupuncturists themselves make the treatments. If the acupuncturist allow it, even placebo controlled RCTs are possible. The placebo treatment of the control group would then consist in putting in needles in such a way that these, according to acupuncture teachings, ought not to have any influence. If such tests are made and no correlation knowledge can be established, then the particular acupuncture therapy in question ought to be rejected, but if there is a statistically significant association, one should use the technology. From a theoretical perspective, one should then start to speculate about what mechanisms there can be that may explain the correlation knowledge that one has obtained. In fact, the development of modern neuroscience has turned the black box theory discussed a little grey; the acupunctural effects may be due to mechanisms that rely on the

creation and displacements of neuropeptides such as endorphins. The procedure we have described consists of the following three general steps:

1. Turn the implausible mechanism theory into a black box theory
2. Test the black box theory; accept or reject it as providing correlation knowledge
3. Turn the good black box theory into a plausible mechanism theory.

In short: de-interpret – test – re-interpret. This means that the assessment of alternative medical technologies is not something that follows either *only* from a comparison of the alternative theoretical framework with the biomedical paradigm or *only* the rules of the RCT. Rational acceptances and rejections of alternative medical technologies are determined by at least two factors: the outcome of relevant RCTs *and* the plausibility of proposed underlying mechanisms. These two factors might mutually impose or weaken each other in the manner shown and simplified in the following four-fold matrix.

		The statistical effect of the treatment is:	
		High	Low
The underlying mechanism is:	Plausible	1	2
	Implausible	3	4

Medical technologies, conventional as well as alternative, can be put in one of these four squares. In squares one and four we have the best and the worst-case scenarios, respectively. In the first we find medical technologies based on mechanism knowledge within the biomedical paradigm, which by RCTs have been proven to be statistically significant, e.g., insulin in the treatment of diabetes.

In the second square we find technologies that can be given a good theoretical explanation but which nonetheless do not work practically. They should of course not be used, but one may well from a theoretical

point of view suspect that the RCTs have not been perfect or that one has put forward the theoretical mechanism too easily. That is, something ought here to be done from a research point of view.

In the third square we find treatments that should be accepted despite the fact that an underlying mechanism explanation is lacking; one contemporary example might be (research is going on) the injection of gold solution as a treatment of rheumatoid arthritis.

In the fourth square, lastly, we find technologies which have been proven ineffective by means of RCTs, and where the proposed mechanisms appear incomprehensible, absurd or unreasonable.

Let us now make some reflections on the theoretical framework of homeopathy. The effect of homeopathic treatment on some allergic conditions has previously been reported as statistically significant (Reilly 1994), but recently other researchers (Lewith 2002) using more participants (power) in their trials have come to the opposite conclusion. Similarly, a meta-analysis in Lancet (2005) suggests that the weak effect of homeopathy is rather to be understood as a placebo effect.

The essence of homeopathy can be stated thus: If some amount of the substance Sub can cause the symptoms Sym in healthy persons, then much smaller doses of Sub can cure persons that suffer from Sym. We are consciously speaking only of *symptoms*, not of diseases, since, according to homeopathy, behind the symptoms there is nothing that can be called a disease. The homeopathic treatments were introduced by the German physician Samuel Hahneman (1755-1843), and rests upon the following principles:

1) The law of simila. In Latin it can be called ‘Simila Similibus Curantur’, which literally means ‘like cures like’. It says that a substance that is like the one that gave rise to a symptom pattern can take away these symptoms. A substance is assumed to work as a treatment if it provokes the same symptoms in a healthy individual as those symptoms the present patient is suffering from.
2) The principle of self-curing forces. Homeopathic treatments are supposed to activate and stimulate self-curing forces of the body; such forces are assumed to exist in all people apart from dying or terminally ill patients.

3) The principle of trial and error. Every new proposed homeopathic remedy needs to be tested, but a test on one single healthy person is enough.
4) The principle of symptom enforcement. An indication that the right homeopathic remedy has been chosen as treatment is that the symptoms of the patient at first become slightly worse.
5) The principle of complete cure. If the right homeopathic treatment is applied, the patient will recover completely – homeopathic treatments are supposed not only to reduce symptoms or symptom pictures, but to cure completely.
6) The principle of mono-therapy. There is one and only one specific homeopathic treatment for a specific symptom picture. If the symptom picture changes, the treatment should be changed, too.
7) The principle of uniqueness. The individual patient and his symptom picture are unique. This assumption makes the performance of randomized controlled trials difficult to accept for homeopaths.
8) The law of minimum. Also called 'the principle of dilution and potentiation'. The smaller the dose of the homeopathic substance is, the stronger the curing effect is assumed to be. Homeopaths have special procedures for diluting the substances they use.
9) The principle of not harming. Hahneman was anxious not to harm any patients. He had seen severe negative effects of bloodletting, and he was in a kind of responsibility crisis when he started to develop homeopathy.

These principles have not been changed over the last 150 years. Many homeopaths use this as an argument that homeopathy is a stable and valid theory in contradistinction to the conventional medicine, where both basic assumptions and treatment principles have been altered numerous times during the same period. Fallibilism seems not so far to have ranked high among homeopaths. Be this as it may; how can we look at these principles?

Principle eight is of special interest. The claim made in 'the law of minimum' contradicts basic assumptions concerning the dose-response relationship in pharmacology, and is thus controversial according to the clinical medical paradigm. Bradford Hill's fifth criterion (see above, section 2) says that more of a dose should lead to more of the response.

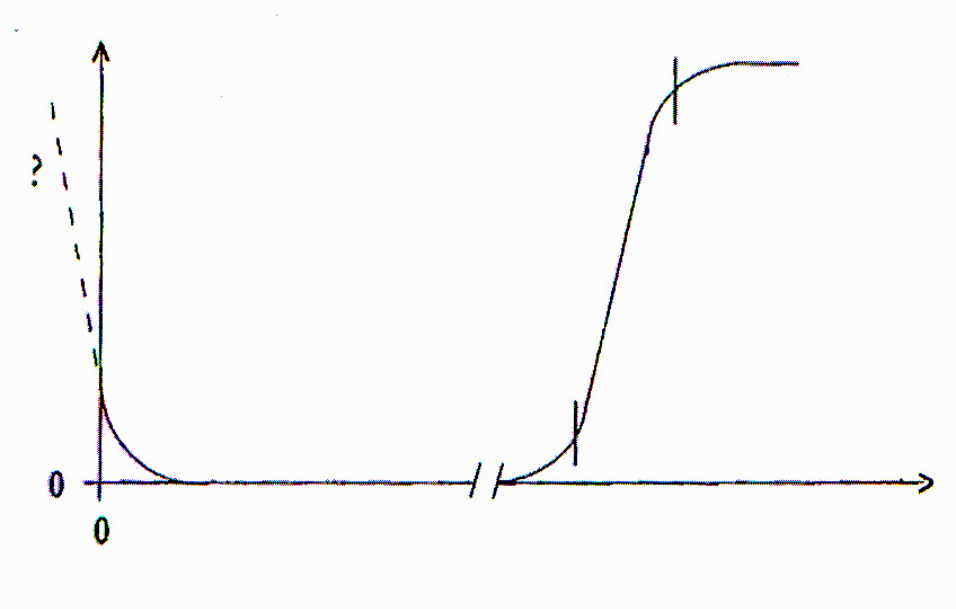


Figure 6: *To the right is a normal dose-response curve; doses on the x-axis and responses on the y-axis. Within a certain range there is a rather linear relationship where a higher dose gives a higher response. To the left is an attempt of ours to illustrate homeopathic theory. Only small doses give any response at all, and within a certain range there is an inverse dose-response relationship, i.e., the lesser the dose the higher the response. The question mark in the area of 'negative doses' is meant to highlight the fact that homeopaths even talk of 'traces of homeopathic substances'.*

But there is even more to be said. Homeopaths seem to wrongly think of substances as completely homogeneous. That is, they seem to think that substances can be divided an infinite number of times without losing their identity as a certain kind of substance. This is in flat contradiction with modern molecular chemistry. If a single molecule of a certain substance is divided further, the molecule and the chemical substance in question disappears. Homeopaths do really accept as effective dilutions where there cannot possibly be any molecules left of the homeopathic substance. This is simply incredible. Now, homeopaths could very well have introduced limit values for their dilutions that had made their 'law of minimum' consistent with modern chemistry, but they have not. As an auxiliary ad hoc hypothesis homeopaths have introduced the idea that water have a kind of memory, and that it is the trace or the shadow of the homeopathic

substance, and not the homeopathic substance in itself, that produces the effect. As we have said, fallibilism seem not to rank high in homeopathy.

We have argued that it is possible to de-interpret somatic alternative medical technologies and assess them by means of RCTs. Homeopaths may mistakenly use their principle of uniqueness (7) to contest this. In an RCT the same treatment is supposed to be given to many patients, and the homeopaths may claim that there are no two cures that are exactly alike. But this evades the issue. Approximate similarity is enough for the statistical purposes at hand. Furthermore, the uniqueness claimed cannot be absolute since homeopathy can be taught and applied to new cases and as stated above RCTs have actually been conducted regarding certain homeopathic treatments.

Now back to our fourfold matrix, especially to square number three (high effect but implausible mechanism). The fact that some effective pharmaceutical products have (unhappily) been rejected because the mechanism appears implausible has been called ‘the tomato effect’. Tomatoes were originally cultivated solely in South America; they came to Europe in the fifteenth century. Up until the eighteenth century tomatoes were grown both in northern Europe and North America only as ornamental plants and not as food. The reason was partly that many people took it for granted that tomatoes were poisonous. Tomato plants belong to the family Solanaceae, which includes plants such as belladonna and mandrake, whose leaves and fruits are very toxic; in sufficient doses they are even lethal. That is, tomatoes were rejected as food because people assumed that they contained an unacceptable causal mechanism. Since the tomato effect (rejecting a useful treatment because of mistaken mechanism knowledge) has similarities with the statistical ‘type 2 error’ (rejecting a useful treatment because of mistaken correlation knowledge), we think it had better be called ‘type T error’; ‘T’ can here symbolize the first letter in both ‘Tomato’ and ‘Theory’. In the 1950s and 60s many Western physicians committed the type T error vis-à-vis acupuncture.

Our views on alternative medicine are equally applicable to old European medical technologies such as bloodletting. The fact that many physicians actually found a certain clinical effect of bloodletting – apart from side effects and death – might be due to the following mechanism. Bacteria need iron in order to proliferate. Hemoglobin contains iron. When

letting blood, the amount of free iron in the circulating blood is reduced; and parts of the remaining free iron is used by the blood-producing tissues in their attempt to compensate for the loss of blood. Therefore, in a bacteria-based sepsis, bloodletting might theoretically have a bacteriostatic effect and thus be effective in the treatment of bacterial infections.

According to our experience, many proponents of the clinical medical paradigm as well as many proponents of alternative medicines claim that the conflict between them is of such a character that it cannot be settled by rational and empirical means. We think this is wrong.

Reference list

Boorse C. Health as a Theoretical Concept. *Philosophy of Science* 1977; 44: 542-73.

Brain P. In Defence of Ancient Bloodletting. *South African Medical Journal* 1979; 56: 149-54.

Bunge M. Phenomenological Theories. In Bunge (ed.) *The Critical Approach to Science and Philosophy*. Free Press of Glencoe. New York 1964.

Cameron MH, Vulcan AP, Finch CF, Newstead SV. Mandatory bicycle helmet use following a decade of helmet promotion in Victoria; Australian evaluation. *Accident Analysis and Prevention* 2001; 26: 325-37.

Churchland PM. *Matter and Consciousness*. The MIT Press. Cambridge Mass. 1988.

Churchland PS. *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. The MIT Press. Cambridge Mass. 1989.

Guess HA, Kleinman A, Kusek JW, Engel L. *The Science of the Placebo. Toward an Interdisciplinary Research Agenda*. BMJ Books. London 2002.

Hill AB. The Environment and Disease: Associant or Causation? *Proceedings of the Royal Society of Medicine* 1965; 58: 295-300.

Hrobjartsson A, Gotzsche P. Is the Placebo Powerless? An Analysis of Clinical Trials Comparing Placebo with No Treatment. *New England Journal of Medicine* 2001; 344: 1594-1602

Höfler M. The Bradford Hill Considerations on Causality: a Counterfactual Perspective. *Emerging Themes in Epidemiology* 2005; 2: 1-11.

Johansson I. The Unnoticed Regional Ontology of Mechanisms. *Axiomathes* 1997; 8: 411 28.

La Mettrie de JO. Man A Machine. In Thomson A. *Machine Man and other Writings*. Cambridge University Press. Cambridge 1996.

Lewith GT et al. Use of ultramolecular potencies of allergen to treat asthmatic people allergic to house dust mite: double blind randomised controlled trial. *British Medical Journal* 2002; 324: 520.

Lindahl O, Lindwall L. Is all Therapy just a Placebo Effect? *Metamedicine* 1982; 3: 255-9.

Lynöe N; Svensson T . Doctors' Attitudes Towards Empirical Data - A Comparative Study. *Scandinavian Journal of Social Medicine* 1997. 25; 3: 210-6.

Mackie JL. *The Cement of the Universe. A Study of Causation*. Oxford University Press. Oxford 1974.

Morabia A. *A History of Epidemiologic Methods and Concepts*. Birkhäuser. Basel, Boston, Berlin 2004.

Nordenfelt L. *On the Nature of Health. An Action-Theoretic Approach*. Kluwer Academic Publishers. Dordrecht 1995.

Petrovic P, Kalso E, Petersson KM, Ingvar M. Placebo and Opioid Analgesia – Imaging a Shared Neuronal Network. *Science* 2002; 295: 1737-40.

Reilly D et al. Is Evidence for Homoeopathy Reproducible? *The Lancet* 1994; 344: 1601-6.

Rothman KJ, Greenland S. Causation and Causal Inference in Epidemiology. *American Journal of Public Health* 2005; 95: Supplement 1:144-50.

Rothman KJ, Greenland S. *Modern Epidemiology*. Lippincott Williams & Wilkins. Philadelphia 1998.

Senn SJ. Falsificationism and Clinical Trials. *Statistics in Medicine* 1991; 10: 1679-92.

Shang A, Huwiler-Muntener K, Nartey L, et al. Are the Clinical Effects of Homoeopathy Placebo Effects? Comparative Study of Placebo-Controlled Trials of Homoeopathy and Allopathy. *The Lancet* 2005; 366: 726-32.

Silverman WA. *Human Experimentation. A Guided Step into the Unknown*. Oxford Medical Publications. Oxford 1985.

Sharpe VA. Faden AI. *Medical Harm: Historical, Conceptual, and Ethical Dimension of Iatrogenic Illness*. Cambridge University Press. Cambridge 1998,

Smith CM. Origin and Uses of Primum Non Nocere – Above All, Do No Harm! *Journal of Clinical Pharmacology* 2005; 45: 371-7.

Starr P. *The Social Transformation of American Medicine*. Basic Books. New York 1982.

White L, Tursky B, Schwartz GE. (Eds.) *Placebo – Theory, Research and Mechanism*. Guildford Press. New York 1985.

Wulff H, Pedersen SA, Rosenberg R. *The Philosophy of Medicine – an Introduction*. Blackwell Scientific Press. London 1990.

Wu MT, Sheen JM, Chuang KH et al. Neuronal specificity of acupuncture response: a fMRI study with electroacupuncture. *Neuroimage* 2002; 16: 1028-37.

Philosophy,
medical science,
medical informatics, and
medical ethics
are overlapping disciplines.

7. Placebo and Nocebo Phenomena

The Latin word ‘placebo’ means ‘I shall please’, and the opposite term ‘nocebo’ means ‘I shall harm’. In the twelfth century a placebo was an evening service made in order to please a deceased person. In the fourteenth century the content of the term had changed a bit. It came to refer to the simulated tears shed by a professional mourner in connection with deaths and memorial services. The term appears in medical literature in the 1780s, and it then meant that a doctor was more ready to please and to follow the wishes of the patient than to be truly medically useful. In 1906, the prominent American physician, Richard Cabot (1868-1939) stated that placebo giving was nothing but quackery.

7.1 What is the placebo problem?

The modern expression ‘the placebo effect’ evolved during the 1950s, when Randomized Controlled Trials (RCT) became more commonly used. It means that a patient’s symptoms or a disease disappears merely because the patient expects that they will disappear as an effect of a certain biomedical treatment. From a purely conceptual point of view, this biomedical placebo effect had better be placed in a wider conceptual framework, where it is merely one of several different kinds of placebo effects, and where, in turn, placebo effects are merely one of several conceptually possible kinds of psychosomatic curing.

All placebo effects are constituted by self-fulfilling expectations; the patient is cured simply because he believes and expects to become cured by some treatment. In order for a placebo effect to arise, the applied treatment must, for the patient, have the *meaning* or *symbolic significance* that it will cure him. The patient shall consciously and sincerely (but falsely) believe that he is being given a real treatment of some sort for his somatic problems. The existence of placebo effects can be investigated in relation to all kinds of health treatments that are claimed to be effective. Here is a list of four different kinds of possible placebo effects:

- the placebo effect in relation to biomedical treatments
- the placebo effect in relation to psychotherapeutic treatments
- the placebo effect in relation to healing (see Figure 1)
- the placebo effect in relation to self-healing (e.g., meditation regarded as a spiritual activity, and made in the belief that through this activity one will be cured).

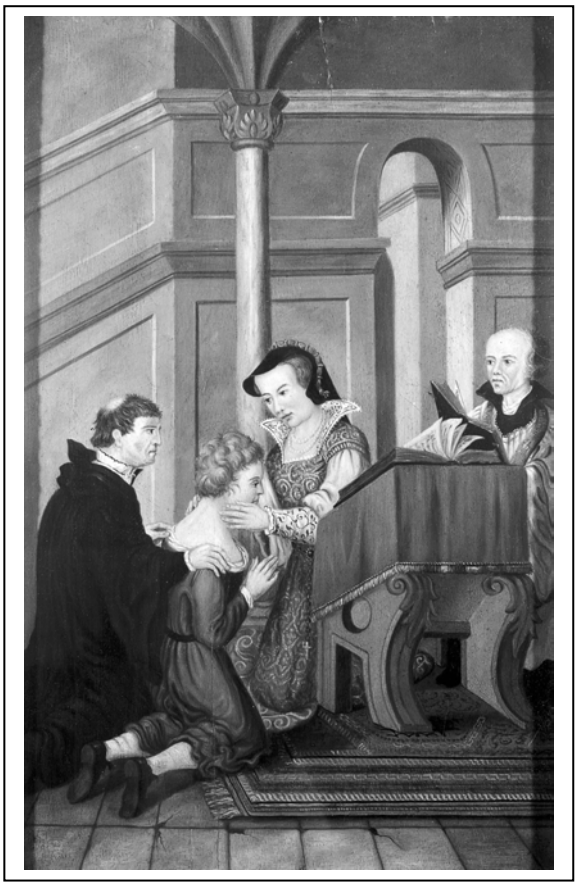


Figure 1: *"A man with a skin disease came to him and begged him on his knees, 'If you are willing, you can make me clean.' Filled with compassion, Jesus reached out his hand and touched the man. 'I am willing' he said, 'Be clean!' Immediately the leprosy left him and he was cured."(Mark 1:40-42, Matthew 8:2-3, Luke 5:12-13) It is not only Jesus who is believed to have cured patients in this way. For more than 700 years the appointed kings and queens of France and England were regarded as possessing supernatural abilities. Just like Jesus the king had only to touch the patient for him to be cured. In the present painting Queen Mary (1516-1558) provides the 'Royal touch' to a patient, probably in the year 1553.*

All the four listed kinds of effects are by definition psychosomatic effects, since expectations are psychic (mental) entities. But they are not the only psychosomatic effects that might be in need of more careful philosophical-scientific investigations. Behind claims such as ‘his strong will saved his life’, there is a belief that a psychic will-to-live can be a causal factor in curing an otherwise deadly disease. Behind sayings such as ‘his promotion/falling-in-love/winning-the-prize made his problems disappear’, there is a belief that some somatic illnesses and diseases can become better only because the patient for some accidental reason enters a more positive psychological mood. Therefore, let us distinguish between at least the following three kinds of possible psychosomatic curing or (to cover a broader range) health improvement:

- psychosomatic health improvement due to *expectations* that one will be better
- psychosomatic health improvement due to a *will* to become better
- psychosomatic health improvement due to positive psychological *moods*.

We will mainly present the problem of psychosomatic curing as it appears in relation to the biomedical placebo effect that occurs in RCTs. To study the placebo effect in relation to healings and psychotherapies has problems of its own; for instance, the placebo treatments must be construed in another way. However, from general theoretical considerations, we find the following two theses plausible:

(a) if the biomedical placebo effect exists, then this is indirect evidence for the existence of placebo effects in psychotherapies and healings
(b) if there is psychosomatic health improvement by means of self-fulfilling expectations, then this is indirect evidence for the existence of psychosomatic health improvement through will power and psychological moods.

When the persons in the control group of a normal RCT are given dummy pills or some other kind of fake treatment, this is not done in order to study and elucidate the placebo effect. On the contrary, the intention is

to use what happens in the control group as a means to isolate the biomedical effect in the experimental group. Mostly, the control group of an RCT is simply called ‘the control group’, but sometimes when the RCT is placebo controlled the group is also called ‘the placebo group’. In our discussion below, it is important that the label ‘control group’ is chosen. Why? Because the outcome in this group might be due also to other factors than the literal placebo effect. Let us explain by means of our earlier introduced *informal* formula B = T – N; or T = B + N, meaning that the **T**otal effect in the experimental group contains both the real **B**iomedical effect of the treatment as well as an effect that is similar to the **N**on-treatment that obtains in the control group. However, the last effect has to be regarded as, in turn, containing three different possible factors:

(1) a literal placebo effect (P)
(2) an effect caused by spontaneous bodily curing (S)
(3) an ‘effect’ that is merely a statistical artifact (A).

That is, we can informally write: N = P + S + A.

When one is only interested in testing whether there is a biomedical effect, one can rest content with the variable N and look only at B = T – N, but if one is interested in the placebo effect as such, then one has to discuss all the factors that may contribute to the value of N. The *general* question is whether or not P should always be ascribed the value zero. If it should not, then many *specific* questions regarding the placebo effect arise. Then, of course, it ought to be investigated when, why, and how biomedical placebo effects occur; how their strength can vary; and what different kinds of biomedical placebo effects there are.

In a rather recent meta-analysis (Hrobjartsson and Gotzsche 2001), it has been claimed that (some minor pain reductions apart) the effects of clinical placebo treatments do not differ from mere non-treatment. To put their – and many others – view bluntly, the placebo effect is a myth. Apart from the biomedical effect there is, these researchers claim, only spontaneous curing (in the patients) and unnoticed (by some researchers) statistical artifacts. That is, N = S + A, since P is always zero; and B = T – (S + A). Good researchers try to take A into account before they talk about the real effect; some not so good researchers sometimes forget it.

The spontaneous course of a disease is the course the disease will follow if no treatment is provided, and no other kind of intervention is made. What in this sense is spontaneous may differ between different kinds of individuals. Therefore, one had better talk about the spontaneous cours*es* of a disease. Be this as it may. To investigate the spontaneous course as such has its own problems, and we will not touch upon these. For the philosophical purposes now at hand, it is enough to have made the variable S clearly visible. So, what about the variable A?

When telescopes and microscopes were first used in science, optical illusions created by the instruments were quite a problem (see Chapter 2). It was hard to say whether some seen phenomena were caused by the instrument or whether they were real features of the investigated objects. Analogously, when statistics was first used in empirical science, statistical illusions created by the mathematical machinery appeared. It was hard to say whether some data were mere statistical side effects or representations of real empirical phenomena. Famous is the story about how the English polymath and statistician Francis Galton (1822-1911) first thought that he empirically had found a causal relationship in a statistical material; called its statistical measure ‘coefficient of reversion’; realized that it was in fact a result of his mathematical-statistical presuppositions; and then re-named into ‘coefficient of regression’. He was reflecting on an inheritance experiment with seeds, where he compared the size of the ‘parents’ with the size of their offspring. The kind of fallacious reasoning he discovered is nowadays called ‘the regression fallacy’.

In *pure mathematical statistics*, there is a theorem called ‘regression towards the mean’ that is concerned with a comparison between a pre-given primary sample and a secondary sample. Both samples belong to the same population, which has a given probability distribution on a certain variable. If the mean value (X_1) of this variable in the primary sample is far away from the mean value (μ) of the whole population, then the mean value of the randomly chosen secondary sample (X_2) will probably be closer to μ than X_1 is. ‘Going’ from the primary sample to the secondary sample means ‘going’ from X_1 to X_2, and thereby ‘regressing’ towards μ. When the two samples represent empirical data, the pure mathematical probability statements have been exchanged for some kind of objective probability statements (Chapter 4.7). The purely mathematical regress

towards the mean may then in the empirical material falsely be taken as an indication that there is some kind of cause that makes it the case that X_1 in the primary sample turns into X_2 in the secondary one.

When the regression fallacy is presented in this verbal fashion, it may seem odd that any modern researcher should ever take a regress towards the mean to represent causality, but in statistical investigations with many dimensions the regression can be hard to discover. Consequently, the converse mistake – to find only a mathematical regression to the mean where there in fact is a causal relation – is also quite possible. In our opinion, much discussion around statistical artifacts seems to be permeated with the problems of the interpretation of probability statements that we highlighted in Chapter 4.7. There are also reasons to think that the kind of discussion of the causal relation that we presented in Chapter 6.2 might have repercussions on how to view presumed regression fallacies.

We regard the general problem of the existence of a biomedical placebo effect as not yet finally solved. That is, this book does not try to answer questions such as ‘Is there anywhere a biomedical placebo effect?’ and ‘Can there be some kind of psychosomatic curing?’ We try to make these issues clearer in the hope that such a feat might improve future medico-philosophical reflections and medico-scientifical investigations. We will, though, in the next section write in the traditional RCT-way as if there were a biomedical placebo effect. If there is none, then some of the things we say have in some way to be re-interpreted as being comments on spontaneous curing.

On the assumption that there is a biomedical placebo effect, some people have been classified as being high placebo responders and others as being low responders. A couple of recent investigations by means of position emission tomography (Petrovic et al 2002, Lieberman et al 2003, Zubieta et al 2005) and functional magnetic resonance scan (Eisenberger et al 2003, Petrovic et al 2005) lay claims to show that there is a specific area in the brain that can be associated with the placebo and the nocebo effects. In this area there is a significant difference in brain activity between high and low placebo responders (see Figure 2).

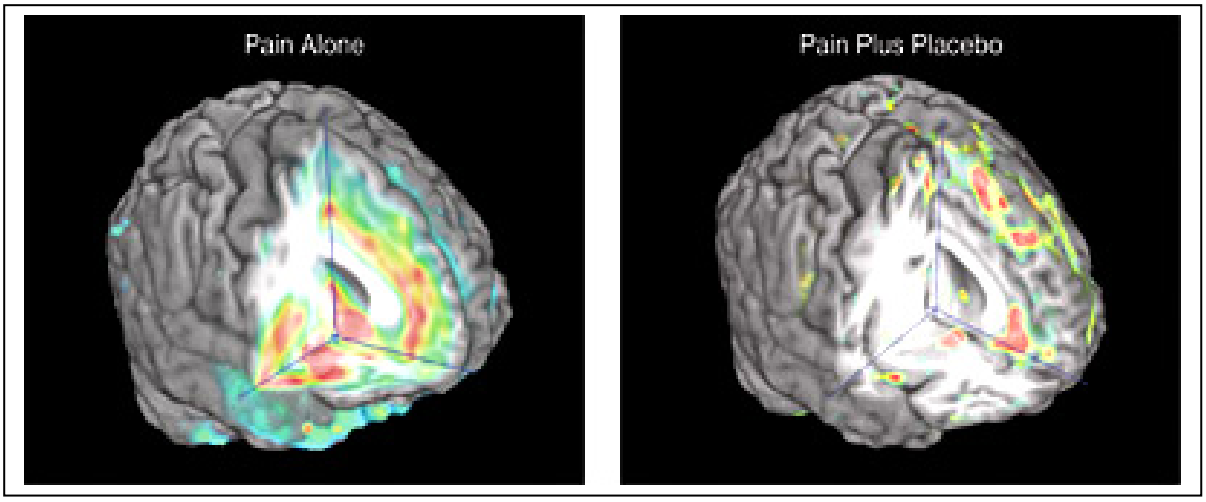


Figure 2: *The pictures above are so-called molecular PET (positron emission tomography) images. They show the activity of the endogenous opioid system on μ-opioid receptors in humans in sustained pain without (to the left) and with (to the right) the administration of a placebo. The difference is much more visible in color pictures. (Credit: Jon-Kar Zubieta.)*

7.2 Variables behind the placebo effect

There are studies around the biomedical placebo effect as such. It is by no means a constant effect, and it varies with many variables. In some RCTs the placebo effect is close to zero and in others close to 100 percent. Some relevant variables – which may mutually interact – are: a) characteristics of the patients; b) characteristics of the doctors/providers; c) the seriousness of the diseases; d) the nature of the medical treatments; and e) the context and setting of the treatments. In a sense, all the factors (b) to (e) have to do with how the individual patient perceives the corresponding factors.

(a) *The patient*. In relation to the biomedical placebo effect, there seems to be two broad categories of patients: ‘high placebo responders’ and ‘low placebo responders’. However, there has been a tendency to over-interpret these labels and make the corresponding features too basic properties of the persons in question. How an individual patient reacts depends on the cultural setting; how the patient feels; how ill the patient is etc. In one setting a patient may be a high responder and in another a low one.

Nonetheless, there is a difference. Moerman has shown that high responders improve more than low responders in both the experimental group and the placebo group. In one investigation the mortality rate (after five years) among those who took three real tablets three times per day (of a cholesterol-lowering treatment) against myocardial infarction was 15% for high responders and 24.6% for low responders. Among those who received dummy tablets the mortality rate was 15.1% and 28.2%, respectively. That is, in this study there was no difference between the experimental and the placebo groups with respect to the outcome, but there was such a difference between the group of high responders and the group of low responders.

(b) *The doctor*. What the patient-provider relationship looks like affects the degree of the biomedical placebo effect. Confidence in the skill and positive attitude of the doctor makes the placebo effect higher. However, what actually creates confidence and trust in the patient may also vary. An authoritarian attitude might create high confidence in some patients and low in others; and the same goes for a non-authoritarian attitude. Listening to the patient, capacity of being emphatic, respecting the patient's autonomy, etc., contribute to high placebo effect and vice versa. A study (Olsson et al 1989) shows that patients suffering from tonsillitis, who were treated both medically correctly and with empathy (including giving explanations of various sorts to the patients) were cured significantly more promptly compared to patient who were treated only medically correctly.

(c) *The nature of the disease*. In general, it seems to be the case that the more pain a disease is associated with, the higher is the placebo effect. For instance, in angina pectoris the placebo effect is higher during winter than in summertime when the conditions causing angina pectoris are more significant. On the other hand, some studies show remarkably uniform values. In a compilation of 22 RCTs concerning the treatment of pain with different drugs (weakest, aspirin; medium, codeine; and strongest, morphine), the placebo effect was on average within the 54-56 % interval in all trials. Diseases where the placebo effect has been shown to be rather high include angina pectoris, peptic ulcer, claudicatio intermitens, and allergic conditions; but even some types of cancer have been reported to

show a non-negligible placebo effect. One might well speculate about whether the prestige of certain diseases (both in view of the patient and in view of the doctor) might influence the magnitude of the placebo effect, but this has not yet been studied. Some diseases such as acute myocardial infarct and endocrine dysfunctions have been estimated by doctors and medical students to be more prestigious than others such as schizophrenia and fibro-myalgia (Album 1991).

(d) *The nature and meaning of the treatment*. Dummy injections seem to be more effective than dummy pills; and when administrating dummy pills or capsules, the color, size, taste and form of the pills are of importance. The number of pills seems also to have impact on the placebo effect. This has been shown in a meta-analysis of 71 RCTs for the treatment of peptic ulcer; four placebos per days is significantly more effective than two. Capsules (of chlordiazepoxide) seem to be more placebo effective against anxiety than pills – ceteris paribus; but here even cultural aspect may influence the effect. As stated by Moerman, Italian females associate the blue color with the dress of the Virgin Mary, who is seen as a protective figure and thus blue pills or capsules might be used as sedatives or sleeping pills. In contrast to this, Italian men associate the blue color with the national Italian Soccer team, the Azzurrii, which means success, strength, and excitement. Blue sleeping pills might accordingly be effective to Italian females and work less well for Italian men. Moerman notes in passing that Viagra is marketed in a blue pill. Sham operations have also proven to give rise to a certain placebo effect, e.g., sham chest surgery (Cobb et al. 1959), sham operations (masteoidectomies) behind the ear for a disease (called Mb. Menniére) in the inner ear (Thomsen et al. 1983), and arthroscopic sham operations in osteoarthritis in the knee (Moseley et al 2002).

(e) *The situation and the cultural setting.* Treatments provided in a highly specialized and famous university hospital may give rise to a stronger placebo effect than an ordinary country side treatment by a general practitioner. Even the surrounding – a room with a view (compared to looking at a brick wall) – seems to influence recovery after gall bladder surgery. It seems also to minimize complaints to the nurse as well as need

of painkillers and even scores for post-surgical complications. In a meta-analysis of 117 trials concerning the treatment of peptic ulcer with acid inhibitors, there was a large variation between the different countries involved. For instance, the trials from Brazil had a placebo effect that was significantly lower than in other countries; seven percent versus thirty-six. Part of this difference might be due to the fact that the Helicobacter pylori bacteria might be more common in Brazil compared to the rest of the world, but it does not explain why (six) studies from Germany (part of the same meta-analysis) show that the average of the placebo effect was 59%. In Germany's neighboring countries, Denmark and the Netherlands, which in most respects are quite similar to Germany, the average placebo effect was 22% in the above mentioned six trials. The placebo effect is, however, not always high in a German context; in comparative placebo controlled trials concerning hypertension, the German studies showed the least improvement in the placebo treatment group. The natural culture as well as the setting and the nature of the treatment seem to influence the placebo effect.

If there is a biomedical placebo effect, then many factors that work in concert influence it. If there is no such placebo effect, then the conclusion is that there are many factors that influence what is called the spontaneous curing or natural course of an illness and disease. That is, there are no absolutely spontaneous or natural processes.

7.3 Specific and unspecific treatment

Many medical treatments are said to be *unspecific*. That is, one does not know exactly what is to be regarded as, so to speak, the real treatment in the treatment. That is, one has not (yet) been able to isolate or point out the specific substance supposedly functioning. The notion of specific treatment emerged in early biochemistry (1880s) when the function of enzymes was described in terms of a key-lock model. The model implies that to a certain lock there is only one key. A treatment is regarded as specific when we are able to point out the functioning mechanism (on a biomechanical or molecular level) of the treatment. There has also been talk about specific causes and effects in relation to other mechanism theories and disease processes. Specificity, etiology and pathogenesis have been central concepts in medicine. Etiology refers mainly to causes that are responsible

for bringing about or initiating a disease; pathogenesis is concerned with the disease process. Intrusions of micro-organisms as causes of infections are examples of etiology, but the immunological reaction and consequences of the manifestation of the infection are examples of pathogenesis. The clinical medical paradigm has said that – ideally – medical treatments should consist in specific treatment of specific diseases.

We have made this brief remark here in order to stress that placebo treatment and its effects are usually classified as being unspecific. Although it has been possible to understand some aspects of the placebo effect in terms of modern neurophysiology (neuro-peptides like endorphins) as well as in terms of Pavlovian conditioning, the effect is in general unspecific.

Research in modern biogenetics (including genomics and proteomics) is constantly developing, and if these efforts are rewarded, then we might in the future have pharmaceutical treatments which are specific not only in relation to general etiology and/or pathogenesis, but specific also in relation to the individual patient. Genetic knowledge about the patient's enzyme system, pharmacological kinematics, immunological system, weight, and so forth, might be of importance when choosing the right medicine and the right dosage. Probably, it will be more appropriate to call such treatments 'individuated' or 'tailored treatments' than call them 'specific treatments'.

7.4 The nocebo effect and psychosomatic diseases

In Section 1, we distinguished between three kinds of psychosomatic curing and four kinds of placebo effects. All these conceptual possibilities have a negative counterpart:

- psychosomatic health impairment due to *expectations* that one will become worse
- psychosomatic health impairment due to a *will* to become worse
- psychosomatic health impairment due to negative psychological *moods*.

The first of these conceptual slots represents nocebo effects. That there is a nocebo effect means that there is a somatically healthy person that becomes somatically ill or worse merely because he expects to become ill or worse. Nocebo effects are constituted by negative self-fulfilling expectations. The second possibility relates to self-destructive persons as well as to persons that want to become ill or sick in order to benefit from this negative state of affairs in some other and life-quality increasing respect.

The third possibility represents that possibility that there might be psychosomatic illnesses and diseases. Traditional but nowadays debated examples are asthma, some allergies, and migraines. To folk wisdom belongs the view that one can become sick by sorrow and lost love. A somewhat spectacular example is the report of two hundred cases of so-called psychogenetic blindness among Cambodian refugee women who were forced to witness the torture and slaughter of their family members, a procedure conducted by the Khmer Rouge. The women themselves say that they have seen the unbearable and cried until they could not see, and it has not been possible to identify any organic changes or defects in their visual system (Harrington 2002). We will return to and discuss psychosomatic diseases in more detail in Section 7 below.

The nocebo effect comes in at least the following varieties:

(i) the nocebo effect in relation to biomedical conditions and treatments
(ii) the nocebo effect in relation to psychotherapeutic treatments
(iii) the nocebo effect in relation to anti-healing
(iv) the nocebo effect in relation to self-punishment.

(i) Here are some possible examples of the biomedical nocebo effect. It has been reported (Jarvinen 1955) that when the chief of a cardiac intensive unit conducted the rounds, the number of those who had another myocardial infarction doubled compared to ward rounds where the assistant doctor was in charge. Among geriatricians, the expression ‘a peripety’ is sometimes used to describe sudden unexpected deaths by elderly patients who receive a cancer diagnosis. The expression is derived from the Greek word ‘peripeteia’, which means a sudden reversal of

something to its opposite. In many ancient plays such a change of scenery takes place when an unexpected messenger appears. Transferred to the geriatric context it means that elderly cancer patients, who according to the medical prognosis might continue to live for a year or more, die quickly only because they think that there is a somatic cause that will make them die quickly. This phenomenon has also been described in connection with Nazi concentration camps. If a watchman told a new prisoner that he should not expect to leave the place alive, the latter sometimes isolated totally, wandered around with an empty gaze, and eventually died.

(ii) The existence of self-fulfilling expectations that one will become worse from psychotherapy, can be discussed in relation to people who feel forced to enter such a therapy.

(iii) What we call ‘the nocebo effect in relation to anti-healing’ is what is traditionally called ‘voodoo’. It has been claimed that persons can become sick merely because some other but special person tell them that he will, by means of his spiritual powers, make them sick. The most extreme example is the so-called ‘voodoo death’. This is a phenomenon reported by anthropologists to exist in communities where the bonds to particular norms, values, and taboos are very strong (Cannon 1942). The medicine man in such cultures is both a judge and executioner, though not executioner in the traditional sense, he simply pronounces the sentence and then psychological mechanisms in ‘the sinner’ makes the rest. The phenomenon has been claimed to exist even in modern societies. According to Harrington (2002), a mother who one and the same day became aware of the fact that her son was both homosexual and suffering from AIDS, reacted by making a prayer in which she expressed the wish that her son should die because of the shame he had caused her. The son heard the prayer, and one hour later he died. The physician was surprised because the patient was not terminally ill.

(iv) Self-destructive people who believe in voodoo can of course try to curse themselves with some somatic disease. This would be an attempt at what we call ‘the nocebo effect in relation to self-punishment’.

We will return to psychosomatic diseases, but not to the other conceptually possible psychosomatic phenomena mentioned in this brief taxonomic section.

7.5 The ghost in the machine

The question whether psychosomatic health improvements and health impairments really exist touches a core issue in both the biomedical paradigm itself and in its relation to the sub-paradigm that we have labeled the clinical medical paradigm. There is a tension between the general paradigm and the sub-paradigm that we think deserves more attention than it has received so far.

The expression that constitutes the title of this section comes from the English philosopher Gilbert Ryle (1900-1976), who used it in his famous book, *The Concept of Mind* (1949). He maintains that the view that there is a mind, a psyche, or a mental substance is a ghost created by our language. If he is right, there can be no placebo and nocebo effects since there is no mind that can produce any somatic effects.

We have previously claimed that, with respect to the medical realm, the ontology of the biomedical paradigm is an *epiphenomenalist materialism* (Chapter 6.1). In relation to the patients' normal daily life, the ontology is the common sense one where agency (Chapter 2.1) on part of persons is simply taken for granted. The biomedical paradigm has never said that patients should be treated simply as machines, even though it was not until the second half of the twentieth century that it became an official norm that physicians have to respect the integrity of their patients (and, where this is not possible, respect the views of some close relatives or friends). In the biomedical paradigm, to repeat:

- there is no complete denial of the existence of mental phenomena
- it is taken for granted that brain states can cause mental phenomena
- there is a denial that something mental can cause a bodily medical change
- mental phenomena are regarded as being phenomena within the spatiotemporal world.

From a philosophical point of view, this list immediately raises two questions: 'What characterizes mental phenomena?' and 'What characterizes here causality?' Crucial to the contemporary philosophical characterization of the mental are two concepts: 'qualia' and

‘intentionality’. Our presentation of the causal problem will bring in the concepts of ‘mental causation’ and ‘agency’.

7.5.1 Qualia and intentionality

The following kind of situation is quite common. I look at a red thing from a certain perspective and you from another; since the light is reflected a bit differently in the two directions, I see a certain hue of red but you see another. These two hues of red belong to our different perceptions as mental conscious phenomena, and they are examples of *qualia*. To suffer from tinnitus is to suffer from the existence of a certain kind of qualia. Whatever the head and the brain looks like, a person that does not hear anything cannot have tinnitus. There are qualia in relation to all the classical sensory systems. A person born blind cannot really know what it is like to have visual qualia; a person born deaf cannot really know what it is like to have auditory qualia. Pains are another kind of qualia. They exist only as mental apprehensions. In a mind-independent world there are no pains, only pure pain *behavior*. In order to find out whether or not a certain anesthesia works, one has to find out if those who receive it still *feels* pain or not. Corpses, and persons in coma or dreamless sleep, have no qualia. They cannot feel any pain, but, on the other hand, neither can they feel anything pleasant. A quale is a mental phenomenon. The existence of qualia is no anomaly to the biomedical paradigm.

Intentionality means ‘directedness’ and ‘aboutness’ in a sense now to be explained. When I think *of* past and future events, I am *directed towards* and think *about* them; when I have a desire *towards* something, my desire can be said to be *about* this something; when I am angry *at* somebody, my anger is *directed towards* and is *about* this person; and so on. Most mental states and acts contain intentionality, but having the feature of intentionality should not be considered a necessary condition for something to be mental. For instance, some experiences of pure qualia lack intentionality. Whether having the feature of intentionality is in general a sufficient condition for there to be a mental phenomenon, is a question that we will not consider, but, definitely, we do not today ascribe intentionality to dead matter and plants. To be a quale is in itself to be a mental phenomenon, but there can be mental states and acts even if there are no qualia; thinking, for instance, can exist unaccompanied by qualia.

The aboutness characteristic of intentionality comes out vividly if one focuses on spatial and temporal relations. At each moment of time, dead matter and plants are *completely confined* within their spatiotemporal region. Not so with us. Our bodies are, but we can nonetheless think and talk of things that are far away from the spatiotemporal region that our bodies happen to occupy. Even perception is an intentional phenomenon. For instance, when we perceive another person, we are directed at something that exists *at another place* in space than our perceiving bodies do. Our perception is in a sense *about* the other person. In ordinary perceptions, intentionality and qualia are intimately fused.

Perhaps the most peculiar feature of the directedness and aboutness of intentionality is that it can be directed towards, and be about, entities that do not exist at all. We talk more or less every day about fictional figures from novels and cartoons, and these figures do neither exist in space and time nor in some other kind of man-independent realm (such as the realm of ideas postulated by Plato or the realm of mathematical numbers as postulated by many mathematicians). Nonetheless, we can identify and re-identify these figures, be it Hamlet or Superman, in such a way that a conversation about them is possible. The same goes for what is false in false assertions. If such assertions were not about anything, they could not be false. If fictional literature and cartoons were not about anything, we would not read them. False assertions and fictional assertions are similar in that none of them refers to anything in reality that corresponds to them exactly, but they differ in that false empirical assertions only as a matter of fact lack truthmakers, whereas fictional assertions cannot have any. No inorganic matter and no plants can have this kind of directedness and aboutness.

In summary, individual intentional states and acts can, although anchored in our body (especially our brain) at a certain time, be directed towards entities that:

1. are spatially distinct from the body
2. are both in the past, in the present, and in the future
3. do not in the ordinary sense exist at all.

The third point may seem remarkable. Does it not imply the logical contradiction ‘there are non-existent things’. No, it doesn’t. It implies only: ‘there are intentional states and acts that are directed towards non-existent things’. The fact that an intentional state or act is about something does not imply that this something can exist independently of acts of apprehension of it. Now, since in everyday language we speak as if fictional figures really exist (‘Have you read the last book *about* Harry Potter?’), one might perhaps have better say that falsities and fictional figures exist, but that they have a special *mode* of existence. They can only exist in and through intentional states and acts of human beings, but they can nonetheless be the same (be re-identified) in many different such intentional acts.

Having made clear that in one sense fictions do not exist, but in another they do exist, we would like to point out that most measurement scales in medicine and the natural sciences (blood pressure, mass, electric charge, etc.) are constructed without any constraint that every magnitude on the scale has to refer to a really existing quantity in the world. Many magnitudes must be taken as having fictions as referents; it seems odd to think that all the infinitely many magnitudes of continuous scales have referents. Similarly, it makes good sense to speak of entities such as purely hypothetical kinds of genomes. In this sense there are fictions in science as well as in novels and cartoons, but this does not imply that fictions exist in some mind-independent realm of their own. Often, in both novels and science, the fictional is mixed with the real (compare the comments on ‘fictionality content’ in Chapter 3.5).

Intentional states and acts seem to be able to inhere in at least humans and some other higher animals, but not in pure matter and plants. What then about texts and pictures? Are they not pure matter, and mustn’t they be said to contain intentionality? No, a further distinction is here needed. Texts and pictures have only a *derived*, not an *intrinsic*, form of intentionality. They can *cause* specific intrinsic intentional states and acts in beings with consciousness, but they do not in themselves contain the kind of directedness and aboutness that we have when are reading the texts and are seeing the pictures.

We call texts and pictures ‘representations’, as if they in and of themselves were directed towards and were about (represent) something distinct from themselves. But this way of speaking is probably due to the

fact that in everyday life we take our own presence so much for granted, that a distinction between intrinsic and derived intentionality is of no pragmatic value. But in the ‘ghostly’ context now at hand, such a distinction is strongly needed.

It is in relation to derived intentionality, especially words and sentences, that talk of *meaning* and *symbolic significance* is truly adequate. Nouns, verbs, adjectives, and adverbs have meaning because (i) they can cause intrinsic intentional states, and we can by analytical thinking (ii) divide such signs/terms into two parts: (iia) a purely graphical sign and (iib) what contains the directedness in question. The latter *is* the meaning (symbolic significance), and the former, the graphical sign, *has* meaning (symbolic significance). The same meaning can be connected to different graphical signs (e.g., the German word ‘Hund’ and the English word ‘dog’ have the same meaning), and two different meanings can be connected to the same graphical sign (e.g., ‘blade’ as meaning the cutting part of things such as knives and machines and as meaning the leaf of plants).

7.5.2 Intentional states and brain states

Only when something with *derived* intentionality interacts with a brain can the corresponding *intrinsic* intentionality come into being. In other words, only when a representation of X (entity with derived intentionality) interacts with a brain can there be a real representation of X, i.e., an intentional state or act that really is about X. The sentence ‘Clouds consist of water’ is a representation of the fact that clouds consist of water only because it can cause people to be directed towards this fact and have thoughts about it.

Obviously, the feature of intentionality is not to be found among any of the scalar properties of physics and chemistry (length, mass, energy, etc.), but neither is it to be found among vector properties such as velocity, acceleration, and field strength. The directedness and aboutness of intentionality must not be confused with the directionality of motions and forces.

Intentionality is not even to be found in what quantum physicists call ‘physical information’ or what molecular biologists call ‘genetic information’. When an intentional state contains information there is either a true assertion or a true belief about some state of affairs. But even though

in fact true, the same kind of intentional states might have been false. Assertions and beliefs lay claim to be about something, and if this something does not obtain they are false. Assertions and beliefs have a true-falsity dimension built into their very nature. Nothing of the sorts is to be found in what is called information in the natural sciences. A distinction between two kinds of information is needed. Assertions and beliefs can contain ‘intentional(-ity) information’, whereas physical information and genetic information exemplify ‘non-intentional(-ity) information’

According to (some interpretations of) quantum physics, there is ‘physical information’ that can move between ‘entangled states’ faster than light and ‘inform’ one of the entangled states about what has happened in the other. But such states contain neither intrinsic nor derived intentional directedness towards the other state; they completely lack a true-falsity dimension.

Genetic information exists in the double helix two-molecule combination DNA, and it can be represented by so-called DNA sequences consisting of a number of individual letters chosen (for human beings) out four different types of letters (A, C, G, T), each of which represents a certain nucleotide. Such information can be transferred from DNA to other molecules, in particular to ‘messenger-RNA molecules’, which, in turn, can transfer the information to places where the protein syntheses that create new cells can take place. In the sense of information used here, things such as vinyl records, tapes, and cd’s can be said to contain information about melodies. And bar codes on commodities can be said to contain information about prizes. In all these cases, to speak about ‘information’ is a way of speaking about complicated causal processes where the internal structures and patterns of the causes and effects are necessary to take into account; here, the causes and effects are not simple events such as a person turning on a switch or a bacterium entering the body (as in Chapter 6.2). Biological information that resides in chemicals is *not* like the derived intentionality that resides in texts. The information contained in DNAs consists of *patterns* constituted by four different kinds of nucleotides that play a certain role in some purely non-intentional processes (taken for granted that no superhuman deity has created DNA the way humans create cd’s).

In the natural-scientific use of ‘information’ now made clear, i.e., ‘non-intentional information’, our perceptual systems can be said to receive perceptual information about the environment even when there are no associated intentional perceptual states (or acts). Perceptual psychology has made the expression ‘to perceive’ ambiguous. Today, it can mean both ‘to receive perceptual information’ and ‘to be in a perceptual intentional state’. What makes the problem of perceptual intentionality even more confusing is that in order for us as persons to have veridical perceptual intentional (mental) states, our brains have to receive some corresponding perceptual information. However, this fact cannot possibly cancel the conceptual distinction between intentional and non-intentional information. Therefore, nor can it in itself show that brain states (and/or processes) containing perceptual non-intentional information are identical with the corresponding intentional states. *Put in another way, the fact that brain states can (even without external stimuli) cause intentional states (even dreams are intentional states) does not show that intentional states are brain states*; at least not in the way the latter are conceived of in today’s physics, chemistry, and molecular biology.

When thinking about philosophical identity problems such as those concerned with brain states and intentional states, one should be acutely aware of the fact that ordinary language often relies heavily on the context and the speakers’ background knowledge. For instance, to say in our culture ‘Sean Connery *is* James Bond’ is to say that Sean Connery *is (playing)* James Bond, not that SC and JB are identical. Similarly, nothing can be wrong in saying ‘intentional states *are* brain states’ as long as one means that intentional states *are (caused by)* brain states.

That there is an oddity in a complete identification of intentional states with brain states can be illustrated as follows. Touch your head with your hands. You have now a tactual percept of the outside of your head. Assume, next, that this percept is *completely identical* with some of your brain states. If so, what you perceive as happening on the outside of your head must in fact be happening inside your head. And the same must be true of all your veridical perceptions of events in the external world; they seem to exist outside your head, but (on the assumptions stated) they only exist inside of it. If one thinks that all intentional states are completely identical with one’s brain states, then one implicitly places one’s percepts

of the ordinary world in one's brain. On such an analysis, veridical perceptions differ from dreams only in the way they are caused: dreams only have proximate causes, but veridical perceptions have distal causes too.

To accept that there are mental phenomena (qualia and intentional states) that in some way or other are connected to or inhere in the body is to accept *property dualism*. This dualism differs from Descartes' *substance dualism* in that it is not assumed that what is mental can exist apart from what is material. According to property dualism, mental phenomena can inhere in matter even though they are not like properties such as shape and volume. Property dualism is compatible with a naturalist outlook. Qualia and intentional phenomena exist in the spatiotemporal world, but they differ in their very nature from everything that so far has been postulated in physics, chemistry, and molecular biology.

7.5.3 Psyche-to-soma causation

Without intentional states there can by definition be no placebo effects; these effects are by definition caused by self-fulfilling mental expectations, and such expectations are intentional states. If there are neither intentional states nor qualia there are no mental phenomena at all and, consequently, no psychosomatic effects whatsoever. In order for there to be any psychosomatic effects there have to be mental phenomena, but, equally trivially, there also has to be a causal relation that goes from the mental to the bodily, from psyche to soma. In contemporary philosophy, the possibility or impossibility of such a relation is discussed under the label 'mental causation', but we will call it 'psyche-to-soma causation'. Although causal talk is ubiquitous in both everyday life and scientific life, the notion of causality is philosophically elusive (cf. Chapter 6.2). The special problem that pops up in the context now at hand is that the causes and the effects are of radically different kinds.

Conspicuous cases of causality are those where one material body affects another: a stone breaks a window, a saw cuts a piece of wood, a billiard ball pushes another billiard ball, etc. Here, some philosophical reflection may even find a metaphysical explanation: since two material bodies cannot be in the same place at the same time, something simply has to happen when two material bodies 'compete' for occupying in the same

place. Such a kind of reasoning cannot be used when it comes to psyche-to-soma causation. Mental phenomena do not compete with matter about spatial regions. What then about an electrically charged particle in an electromagnetic field? The field causes the particle to move even though they occupy the same place and they are not of exactly the same ontological kind. Instead of soma-to-soma causation there is field-to-soma causation. This is closer to psyche-to-soma causation, but the difference between the mental and the bodily seems to be much greater than that between electromagnetic fields and electrically charged particles. However, if the epiphenomenalistic materialism of the biomedical paradigm is already taken for granted, one might argue as below. The form of the argument to be presented is the indirect form that is used in RCTs: assume the opposite of what you hope to prove (namely that the null hypothesis is false), and then prove that your assumption (the null hypothesis) cannot be true.

Assume that psyche-to-soma causation is impossible. For reasons of symmetry, it then follows that even soma-to-psyche causation is impossible. Surely, this must be wrong. This means that all our experiences that being hit hard creates pain must be illusory, and that all our knowledge that alcohol and drugs can influence mental states is only presumed knowledge. The fact that somatic changes may cause mental changes is not a fact related only to the biomedical paradigm; it is a fact that is generally accepted. That bodily events can cause pain, is in common sense as obvious as the fact that one billiard ball can cause another such ball to move. Therefore, for reasons of symmetry psyche-to-soma causation is just as possible as soma-to-psyche causation, which, in turn, according to everyday perception, is just as possible as soma-to-soma causation.

7.5.4 Agency

So far, we have spoken of psyche-to-soma causation the way we spoke of causal relations between purely material events, soma-to-soma causation. Even agency (Chapter 2.1) is, if it exists, a kind of psyche-to-soma causation; one which brings in free will and human freedom. It shall explain why soft (and not hard) determinism is true. This issue, let it be noted, is of no relevance for the question of the existence of placebo and nocebo effects and other passive psychosomatic processes. But since it

belongs to the general issue of psyche-to-soma causation, and is a necessary condition for the existence of what we termed *active* psychosomatic curing, we will take the opportunity to make some remarks on it here; especially, since experiments by a physiologist, Benjamin Libet (b. 1916), has become prominent in the discussion.

Agency contains a special kind of intentionality: intentions. Obviously, some unconscious electric processes in the brain are necessary preconditions for our actions; without a brain no agency. In Libet's experiments, the actions were very simple ones such as pressing a button, which we know are correlated with neuron activity in the supplementary cortex. If his results are generalized to all kinds of actions, one can state his conclusions as follows. The neurons in the brain that are responsible for a certain physical movement in a certain body part start firing about 500 milliseconds before this movement takes place, but conscious intentions or urges to make such a movement/action arise about 150 ms before the action starts. That is, seemingly free actions are triggered about (500–150=) 350 ms before the presumably free urge or free decision to act appears. It might be tempting to conclude that the experienced decision to act is merely an epiphenomenon to the first 350 ms of the relevant brain processes, and that we should look upon agency as a complete illusion. Libet's own conclusion, however, is not that radical. He says that we are still free to inhibit actions that are on their way; there is at least 150 ms left for intervention by a free will. On his view, true agency can only be controlling, stopping some actions and letting others through.

As we have said, science and philosophy overlap. One kind of criticism leveled at Libet from a neurologist-philosopher pair says that he has neglected the fact that voluntary actions need not be preceded by any felt urge or decision to act (Bennet and Hacker, 8.2). Let us put it as follows: sometimes we have in the mind's eye a specific intention to act *before* we act, but often we become aware of our intention only *in* our very acting. That is, there are two kinds of intentions, reflective (or prior) intentions and non-reflective intentions. Libet seems to think that all free actions require reflective intentions.

The distinction between reflective and non-reflective intentions is quite in conformance with common sense and judicial praxis. We talk of children as performing actions long before they are able to have any

reflective intentions; they are said to *act* spontaneously, not to be mere stimulus-response machines. We even take it for granted that we often directly in a movement can see whether or not it is an action or ‘mere movement’; and we cannot see reflective intentions, only be told about them. For instance, in soccer, where playing with the hands is forbidden, the referees are expected to be able to see whether or not an arm that touches the ball is *intentionally* touching it. If there is an intention, they give a free-kick; otherwise not. Usually, players have no time to form prior intentions before they are acting. In most laws, ever since ancient times, there is some distinction between law-breakings that are done reflectively (e.g., murder) and the same ones done un-reflectively (manslaughter). In judicial Latin, they are called ‘dolus’ and ‘culpa’, respectively.

7.6 Biomedical anomalies

A scientific anomaly is some empirical datum or unified collection of empirical data that cannot – at a certain time – be accommodated by the current paradigm. The future will tell whether an anomaly is in fact reality’s way of falsifying the paradigm, or if it is merely a symptom of a need to develop the paradigm further. Anomalies should be kept conceptually distinct from paradigm conflicts, even though such conflicts can contain mutual accusations that the other paradigm is confronted with falsifying anomalies. In this vein we have already discussed alternative medicine. Before discussing what might be regarded as present-day anomalies to the biomedical paradigm, we will illustrate the concept of anomaly a little more than we have done so far (see Chapter 2.4). History of astronomy provides an illustrative example of how similar kinds of anomalies can be either falsifications or merely cries for improvement. The same two possibilities exist in relation to biomedical anomalies too.

During ancient and medieval times, scientists claimed that not only the Moon, but also the Sun, the planets and the stars move around the Earth in certain orbits (the geocentric worldview). Since the Renaissance, we believe the opposite: the planets, now including the Earth, orbit around the Sun (the heliocentric worldview). Today, we also know that our solar system moves in relation to other solar systems. After the heliocentric theory had become accepted, Isaac Newton proposed his famous mechanics. This theory is meant to explain movements of all kinds of

material bodies. We shall highlight some of its anomalies in relation to our solar system.

Only the orbits of a few planets were in precise accordance with the calculations based on Newtonian mechanics. Especially significant were the deviations for Mercury and Uranus (the latter planet discovered after Newton's death). Newton's law of gravitation says that the force between two interacting bodies is determined by the masses of the bodies and the distance between them. This does not mean, however, that a planet's orbit around the Sun is only dependent on the mass of the planet, the mass of the Sun, and the distance between these bodies. The planets also influence each other mutually. Therefore, scientists could easily attempt to rescue Newton's theory from the anomaly of Uranus's orbit by proposing an auxiliary hypothesis (see Chapter 4.4.), namely that there is another planet – not yet discovered – that also influences the movement of Uranus. Scientists calculated what the mass and the orbit of the hypothetical planet would have to look like if the planet should be able to explain Uranus' actual deviation from the earlier calculated orbit. After having calculated also an actual position of the still hypothetical planet, they directed their telescopes towards this place in the heavenly sphere – and, what a wonder – they discovered the planet; it was baptized 'Neptune'. However, when the actual orbit of Neptune was plotted, one discovered that it did not conform to the calculated one. That is, the solution to one anomaly gave rise to another anomaly. What to do? Well, the strategy was clear. One started to search for another planet and found Pluto. This is a wonderful research story. (Let it be noted, that in August 2006 it was decided that Pluto should no longer be classified as a 'planet' but as a 'dwarf planet'; some words about this in Chapter 8.2.)

What then about the anomaly caused by the funny orbit (a rotating ellipse) of Mercury? Even in this case the astronomers tried to explain the anomaly by means of an auxiliary hypothesis that posited an unknown planet. This planet was assumed to be situated between Mercury and the Sun, and it was given the name 'Vulcan'. One searched and searched for Vulcan, but it could never be found. Later, the orbit of Mercury became explained by the theory of general relativity. In retrospect we can say that the auxiliary hypothesis failed to explain the anomaly.

Let us now return to the biomedical paradigm and its anomalies. We can now ask if contemporary medical researchers should only search for ‘unknown planets’ or for new kinds of ‘gravitational fields’ too. Should they look only for new physiological, biochemical, and microbiological mechanisms or for something else as well? Psychosomatic diseases and the placebo effect are anomalies that bring in psyche-to-soma causation, but there are other anomalies that do not necessarily involve this problem. We will comment on the phenomenon of psychosomatic diseases in the next section but on the placebo effect at once.

The placebo effect is an anomaly in the biomedical paradigm, but it is nonetheless a normal part of the clinical medical paradigm since it is an explicit part of many RCTs. How come? Shouldn’t the clinical medical paradigm be considered a sub-paradigm to the biomedical paradigm? The explanation is that the placebo effect is in RCTs used only as a means to find what is regarded as a biomedical effect, which is based in a soma-to-soma causality. There is so to speak only a *negative acceptance* of the placebo effect, and this, we guess, is the reason why the placebo effect only creates a bit of a tension, a cognitive dissonance, between the clinical medical paradigm and the biomedical paradigm. But this is an unhappy state of affairs. The placebo effect should be considered an anomaly and be treated as such. That is, one should either try to show that there simply is no placebo effect or try, positively, to investigate the placebo effect as such. In the former case, the anomaly is explained away because the biomedical paradigm becomes more developed in relation to regression fallacies and studies of the natural courses of diseases and illnesses; in the latter case, today it is impossible to say what would happen to the biomedical paradigm if its soma-to-soma and soma-to-psyche causal relations would have to be truly complemented with psyche-to-soma relations.

But there is also another kind of anomalies, well known to e.g., physicians. Concepts such as ‘specific symptoms’, ‘disease’, ‘etiology’, ‘pathogenesis’, and ‘specific treatment’ belong to the normal everyday discourse of clinicians and are central to the biomedical paradigm. To claim that a *symptom* is specific is to claim that it is the sign of a specific underlying pathological anatomical state/structure or a pathological physiological process. Prototypical examples are: dysfunction of the

endocrine system, deficiency diseases, chromosome aberrations, genetic diseases, infectious diseases, and cancer diseases. As previously stated, a *treatment* is specific when we have knowledge about the underlying pathological mechanism and where and how the treatment interferes in the mechanism.

The ideal of the biomedical paradigm is to discover and describe specific symptoms and diseases, and on such a basis develop specific treatments. By means of symptoms physicians should be able to diagnose an underlying disease and, knowing the disease, they should know what treatment to give. However, this ideal is at present, at least for GPs, impossible to attain. Several symptoms cannot be connected to any known underlying pathological structure or mechanism, and many symptoms simply fade away without leaving any clues as to what the patient was specifically suffering from. Sometimes the underlying disease is not wholly understood or only partly known, and then GPs speak of ‘symptoms diagnoses’; and sometimes the symptoms constitute such a complex pattern that they speak of ‘syndrome diagnoses’.

Physicians have given some of these unspecific symptoms and diagnoses nicknames such as ‘address-less symptoms’, ‘diagnoses of exclusion’, and even ‘kitchen sink diagnoses’. Unspecific symptoms of the back, the chest, the head, and the urethra are examples of address-less symptoms, and the Irritable Bowl Syndrome (IBS) is an example of a diagnosis of exclusion. Often, unspecific symptoms are also called ‘functional inconveniences’, and the patients suffering from them are sometimes described as tricky, difficult, heavy, hysterical, or hypochondriac; sometimes even as malingerers.

One might wonder why physicians so readily speak disparagingly about patients for whom they are not able to make a specific diagnosis. Arguably, there is a general tendency in human nature to blame the things one cannot cope with rather than to blame one’s own capacity or the knowledge of the community to which one belongs. But perhaps a better understanding of the philosophy-of-science phenomena of anomalies can alter this behavior. If anomalies are normal, then neither the doctors nor the patients need to be blamed when problematic symptoms and disease patterns occur. The moral of the story is that some major discoveries are made when researchers concentrate on anomalies.

7.7 Helicobacter pylori in the machine

According to the views of science and philosophy that we have presented, there is no simple litmus test that can tell what specific theory is the most truthlike, or what paradigm one should work with in the future. The positivist dream is gone. But, we also said, paradigms and theories have to be evaluated; anything does not go. Reality constrains our conceptual schemas. To regard the human body as a social construction is psychologically possible only for people who do not suffer from any serious somatic disease. Fallibilist epistemological and ontological realism is the golden mean between positivism and social constructivism. Let us now take a look at the case of peptic ulcer and Helicobacter pylori in the light of what has been said about correlations, black boxes, grey boxes, and shining mechanisms.

Peptic ulcer was clearly identified in the nineteenth century thanks to the increased use of autopsies. It was first discovered in the stomach and later on in the duodenum. Up until the 1930s, the cause of it was regarded to be purely biomedical; hyperacidity was regarded as the proximate cause and (in some corners) spicy food as the distal cause. At that time surgery was a common treatment, cutting off nerves which were supposed to stimulate the production of acids in the stomach. But then one started to question if such factors as personality traits, psychological trauma, and psychological stress could cause peptic ulcer. That there is a correlation between psychological stress and peptic ulcer seemed obvious to many people from mere everyday observations. And the biomedical paradigm didn’t have a grey box yet, i.e., an outline of some mechanism. At this time, it was more or less a scientific fact that bacteria cannot live in strongly acidic environments in the stomach. In many scientific disciplines such as social medicine and stress research, it was during the second half of the twentieth century a commonplace that peptic ulcer is wholly a psychosomatic disease. This view also influenced textbooks and medical education programs.

In the early 1980s, two Australians, the pathologist Robin Warren (b. 1937) and the physician Barry Marshall (b. 1951), published a couple of papers claiming that there is a spiral-shaped bacterium (now named Helicobacter pylori) that causes peptic ulcer and gastritis. That is, there is a

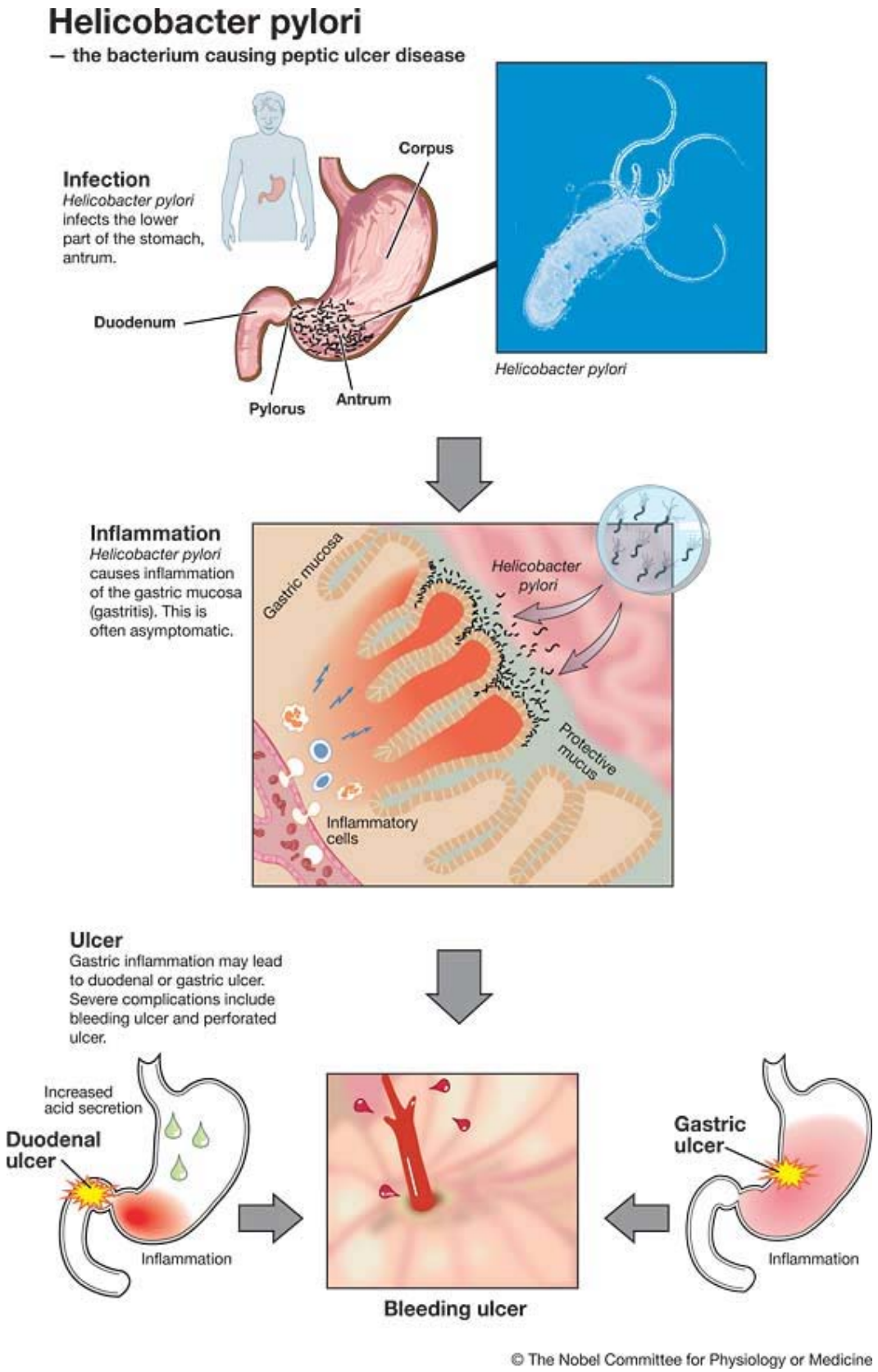


Figure 3: *From the press release of the 2005 Nobel Prize in Physiology or Medicine (from the Nobel Assembly at Karolinska Institutet).*

bacterium that can thrive in acidic milieus. In fact, this bacterium had been observed by German scientists in the nineteenth century, but these observations were forgotten since no one was able to isolate it and grow it in a culture (compare Koch's second postulate; Chapter 2.4). Warren and Marshall, however, could isolate the bacterium and show that it was present in most cases of peptic ulcer and gastritis (cf. Koch's first postulate). Marshall belongs to the list of courageous medical researchers who made tests on themselves. He drank a Petri dish of H. pylori and developed gastritis (cf. Koch's third postulate), whereupon the bacteria could be found in and re-isolated from his stomach lining (cf. Koch's fourth postulate). Also, they showed that H. pylori can be eradicated by means of antibiotics. Patients with chronic peptic ulcer that previously had to suffer from the disease more or less chronically can now recover completely. Later, beginning in the mid 1990s, the official government health authorities of many countries turned Warren and Marshall's views into an official doctrine. In 2005, W&M received the Nobel Prize in medicine. Before their discovery, peptic ulcer and gastritis were mostly treated with medicines that neutralized stomach acid or decreased its production.

What now to say about the H. pylori discovery from a philosophico-scientific point of view? First, it refutes all those who thought that no microbiological mechanism that explains peptic ulcer could ever be found. Second, because of the problem of induction (Chapter 4.2), the discovery cannot without further ado be taken as having shown that all talk about psychosomatic diseases must be illusory talk. Third, and more interestingly, it does not completely eliminate the peptic ulcer anomaly. Something more has to be said. It is estimated that:

(a) as much as 85-90% of all persons that are infected by H. pylori are asymptomatic and do not develop any ulcer (50% of all humans have this microbe), and
(b) H. pylori is not a causal factor in all ulcer cases; it accounts for 75% of gastric ulcer and 90% of duodenal ulcer.

Let us now make use of the concept of 'component cause' ('INUS-condition') that we presented in Chapter 6.2. Points (a) and (b) imply that

H. pylori is a component cause of peptic ulcer. It is not in itself sufficient to produce peptic ulcer, but it is a necessary part of some kinds of causal mechanisms that are sufficient to produce peptic ulcer; however, peptic ulcer can also be produced by mechanisms that do not contain H. pylori at all. The abstract structure of the causal problem can now be illustrated as follows.

Assume that there are three kinds of mechanisms (M_1, M_2, M_3) that need to have H. pylori (H) as a part in order to produce gastric ulcer, and that there is only one kind (M_4) that does not need this bacterium. Furthermore, assume that all these mechanisms can be divided into three main parts (A, B, C, etc). That is, there are four different situations:

1. M_1 (= A & B & **H**) produces gastric ulcer
2. M_2 (= C & D & **H**) produces gastric ulcer
3. M_3 (= E & F & **H**) produces gastric ulcer
4. M_4 (= J & K & L) produces gastric ulcer

According to these assumptions, H. pylori is a cause of gastric ulcer in three cases out of four (75%). In relation to the fourth case, we can still ask the traditional question whether anything purely mental can constitute the whole of J&K&L and be sufficient to produce an ulcer; psychological stress need only be one of the parts in J&K&L. In relation to the first three cases, we can ask if any of the cofactors to H. pylori, A to F, can be identical with psychological stress. If the answer is ‘yes’, there is still a problem of how to understand a psyche-to-soma relation, but this time it needs perhaps not be a strictly causal relation.

Assume that ‘E’ represents psychological stress; and that E, the physiological conditions F and H. pylori together produce gastric ulcer. We do not then necessarily have to interpret the &-signs in ‘E & F & H’ as representing casual interaction. Our assumptions only force us to say that there is a (psyche+soma)-to-soma causality, i.e., (E&F&H)-to-ulcer causality, not that there is a pure psyche-to-soma causality. However, instead we arrive at another problem: how can something mental and something somatic be non-causally connected? How can a bacterium (H), a physiological condition (F), and a psychological state (E) make up a unity that can function as a cause?

Let us leave our speculations at that. As we have already said once in this chapter, we do not consider the problems of the existence of psychosomatic diseases and placebo effects to be solved. This book is meant to introduce the readers to such problems and to ways one might think and speculate about them, not to tell the readers what the true answers are. Science and philosophy will never come to an end. We will conclude this chapter with historical remarks.

In our opinion, the problem of the psyche-to-soma and the soma-to-psyche causal relations look neither worse nor better than some causality problems modern physics faces. For instance, Newtonian mechanics has a problem with action-at-a-distance. According to many intuitions both inside and outside philosophy, things that cause and things that are caused have to be in contact with each other, i.e., causality implies contact. But this is not the case in Newton's theory. (The theory of special relativity does not, by the way, solve this problem; rather the contrary, since it says that no form of energy can move faster than light.) According to the law of gravitation, any two material particles in the whole universe, whatever the distance between them, affect each other *instantly*. Think of the Sun and the Earth. How can they affect each other instantly when they are eight light minutes away from each other? The problem was clearly perceived by Newton himself, and he simply declared the law to be a 'phenomenological law'; and he explicitly said that he did not make any hypotheses about the causal mechanism.

Field theories lack this problem; in such theories interacting particles have their interaction mediated by a field that connects the particles. One electrically charged particle can influence another if there is an electric field emanating from the first particle that affects the other. In the direct field-to-particle interaction there is then action by contact. However, there still seems to be no really good theory available for gravitational fields.

In modern quantum mechanics, the problem of action-at-a-distance reappears in a completely different kind of mathematical formalism and under a new name: 'entanglement'. Two entangled particles (at any spatial distance from each other) are assumed to be able to exchange physical information instantly (but not to transmit energy signals faster than light). The equations allow physicists to make predictions, and so did Newton's laws, but a causal mechanism seems wanting. The problems of action-at-a-

distance and entanglement can be said to be problems of how to conceive what is spatially disconnected as being nonetheless connected. The problems of psyche-to-soma and soma-to-psyche relations can be said to be problems of how to conceive that entities that seemingly have radically distinct ontological natures can nonetheless have something in common.

Reference list

Album D. Sykdommers og Medisinske Spesialiteters Prestisje [The prestige of illnesses and medical specialities]. *Nordisk Medicin* 1991; 106: 232-6.

Balint M. *The Doctor, his Patient and the Illness*. Churchill Livingstone. London 1964.

Bloch M. *The Ryoal Touch: Monarchy and Miracles in France and England*. Dorset Press. New York 1961.

Bootzin RR, Caspi O. Explanatory Mechanism for Placebo Effects: Cognition, Personality and Social Learning. In Guess HA, Kleinman A, Kusek JW, Engel LW (eds.). *The Science of the Placebo. Toward an Interdisciplinary Research Agenda*. BMJ Books. London 2002.

Cabot RC. The Physician's Responsibilty for the Nostrum Evil. *Journal of the American Medical Association* 1906; 57: 982-3.

Cannon WB. Voodoo Death. *American Anthropologist* 1942; 44: 169-81.

Cobb L, Thomas GI, Dillard DH, et al. An evaluation of internal-mammary artery ligation by a double blind technic. *New England Journal of Medicine* 1959; 260: 1115-8.

Cohen S. Voodoo Death, the Stress Response, and AIDS. In Bridge TP, Mirsky AF, Goodwin FK (eds.). *Psychological, Neuropsychiatric, and Substance Abuse Aspects of AIDS*. Raven Press. New York 1988.

Davies CE. Regression to the Mean or Placebo Effect? In Guess HA, Kleinman A, Kusek JW, Engel LW (eds.). *The Science of the Placebo. Toward an Interdisciplinary Research Agenda*. BMJ Books. London 2002.

Dorland Medical Dictionary. 30th edition 2003.

Eisenberger NI, Lieberman MD, Williams KD. Does Rejection Hurt? A fMRI Study of Social Exclusion. *Science* 2003; 302: 290-2.

Guess HA, Kleinman A, Kusek JW, Engel LW (eds.). *The Science of the Placebo. Toward an Interdisciplinary Research Agenda*. BMJ Books. London 2002.

Harrington A. Seeing the Placebo Effect: Historical Legacies and Present Opportunities. In Guess HA, Kleinman A, Kusek JW, Engel LW (eds.). *The Science of the Placebo. Toward an Interdisciplinary Research Agenda*. BMJ Books. London 2002.

Hrobjartsson A, Gotzsche P. Is the Placebo Powerless? An Analysis of Clinical Trials Comparing Placebo with No Treatment. *New England Journal of Medicine* 2001; 344: 1594-1602.

Jarvinen KA. Can ward rounds be a danger to patients with myocardial infarction? *British Medical Journal* 1955; 4909: 318-20.

Kleinman A, Guess HA, Wilentz JS. An overview. In Guess HA, Kleinman A, Kusek JW, Engel LW (eds.). *The Science of the Placebo. Toward an Interdisciplinary Research Agenda*. BMJ Books. London 2002.

Leibovici L. Effects of remote, retroactive intercessory prayer on outcomes in patients with bloodstream infection: randomised controlled trial. *British Medical Journal* 2001; 323: 1450-1.

Lieberman MD, Jarcho JM, Berman S, et al. The Neural Correlates of Placebo Effect: a Disruption Account. *NeuroImage* 2004; 22: 447-55.

Moerman, DE. Meaning, Medicine and the "Placebo Effect". Cambridge University Press 2002.

Moerman DE. Explanatory mechanism for placebo effects: cultural influences and the meaning response. In Guess HA, Kleinman A, Kusek JW, Engel LW (eds.). *The Science of the Placebo. Toward an Interdisciplinary Research Agenda*. BMJ Books. London 2002.

Moseley JB, O'Malley K, Petersen NJ, et al. A Controlled Trial of Arthroscopic Surgery for Osteoarthritis of the Knee. *New England Journal of Medicine* 2002; 347: 81-8.

Olsson B, Olsson B, Tibblin G. Effect of Patients' Expectations on Recovery from Acute Tonsillitis. *Family Practice* 1989; 6: 188-92.

Petrovic P, Kalso E, Petersson KM, Ingvar M. Placebo and opoid anagesia – imaging a shared neuronal network. *Science* 2002; 295: 1737-40.

Petrovic P, Dietrich T, Fransson P, Andersson J, Carlsson K, Ingvar M. Placebo in emotional processing – induced expectations of anxiety relief activate a generalized modulatory network. *Neuron* 2005; 46: 957-69.

Ryle G. *The Concept of Mind*. Penguin Books. Harmondsworth. Middlesex 1966.

Searle J. *Intentionality*. Cambridge University Press. Cambridge 1983.

Silverman WA. *Human Experimentation. A Guided Step into the Unknown*. Oxford Medical Publications. Oxford 1985.

Thomsen J, Bretlau P, Tos M, et al. Placebo effect in surgery fot Meniere's disease: three years follow up. *Archives of Otolaryngology—Head & Neck Surgery* 1983; 91: 183-6.

White L, Tursky B, Schwartz GE (eds.). *Placebo – Theory, Research and Mechanism*. Guildford Press. New York 1985.

Zubieta JK, Bueller JA, Jackson LR, Scott DJ, Xu Y, Koeppe RA, Stohler CS. Placebo effects mediated by endogenous opioid neurotransmission and μ-opioid receptors. *Journal of Neuroscience* 2005; 25: 7754-7762.

8. Pluralism and Medical Science

Cultural monism and religious fundamentalism are nowadays often without much thought contrasted with cultural and religious pluralism. But let us pause for a moment and ask what is and can be meant by pluralism. Are there different kinds of pluralism? If so, are all of them positive? Might it not be the case that pluralism can be desirable within certain areas but undesirable in others? Is pluralism within society at large one thing, and pluralism within the sciences another? In this chapter we will reflect on these topics.

8.1 What is pluralism?

Primarily, ‘pluralism’ means multiplicity. Where there is pluralism there is a multiplicity of something. But the word ‘pluralism’ has also received a secondary meaning: where there is pluralism there should be freedom for each individual to choose what part of the multiplicity he wants to have or to belong to. That is, pluralism in a certain respect should be kept distinct from a corresponding multiplicity that is based only on power balance. Today, most countries are pluralistic in several respects; they contain many different cultures, religions, and political parties – between which the citizens are free to choose.

Democracy is political pluralism, and freedom of religion is religious pluralism. In a democratic society, different political parties identify each other as political parties, and the citizens have the right to vote for and be members of the party they prefer. In a totalitarian society, people opposing the party or ideology in power are not identified as being *political* opponents, but as opponents that represent pure evil, a betrayal of absolute values, or as being simply mentally sick. Therefore, they are not allowed to speak and organize freely and to appear as political opponents. Similarly, in a religiously fundamentalist society, other religions are not really regarded as *religions*. They are merely conceived as heresies or paganisms. According to the fundamentalist, there is only one religion worth the epithet. Therefore, to him, the very notion of ‘religions freedom’ is at bottom nonsensical.

What is involved in a shift from a non-pluralistic attitude to a pluralistic one can be exemplified by the case of language. Our pluralistic attitude towards language is so well entrenched that we do not even think of it as pluralistic. No one thinks that his language is the one and only true language and fights other languages just because they deviate from the true language. But this has seemingly not always been the case. The Greek word for foreigner, ‘barbarian’, originally meant a person who babbles, i.e., a person who talks or makes noises but does not say anything that is meaningful. Probably, when the Greeks first encountered foreign cultures, they perceived these as lacking a language in the way we perceive even pet animals as lacking a real language. This is not strange; this is the way a very foreign language appears when it is first met with. Since nowadays we take it for granted that all human cultures have a proper language, we reflectively know that, in fact, what appears to us as mere babbling is much more than that, but once this could not have been equally obvious.

Something crucial happens when people begin to perceive what has so far appeared as babbling as meaningful language acts. The distance from the ‘barbarians’ diminishes; the latter are no longer as dissimilar as was first thought. Correspondingly, there is a difference between regarding our own political ideology or religion (atheism included) as the only ideology or religion worth the label ‘political view’ and ‘religious view’, respectively, and as regarding them as one among several possibilities. It is one thing to regard a whole population as consisting of heretics, and quite another to regard it as a people with another religion.

A shift from a totalitarian or fundamentalist perspective to a pluralistic one can, but need not, mean that all the old conflicts disappear. But it means at least that the conflicts are given a somewhat more peaceful form. When those who were formerly described as barbarians, betrayers, heathens, or representatives of evil become described as political opponents or members of another religion, they are naturally treated less ruthlessly. Pluralism always implies some tolerance.

Freedom of religion and democracy are the two grand pluralisms, but even other pluralisms have grown stronger. Nowadays, we are allowed to dress very much the way we want, and we are allowed to live in many different kinds of family-like relationships. Even the old opposition between (presumed normal) heterosexuality and (and presumed perverse)

homosexuality has very much become reduced to a distinction between two mere forms of sexuality, heterosexuality and homosexuality.

8.2 Pluralism in the sciences

During the last decades of the twentieth century, Western universities witnessed quite a change towards pluralism with respect to scientific theories and theorizing. It was most pronounced in the social sciences and the humanities. For instance, in psychology there are nowadays several rival theories such as cognitive psychology, behaviorism, humanistic psychology, various versions of psychoanalytic psychology, and so forth. In literary studies we find different research traditions such as hermeneutics, structuralism, semiotics and post-structuralism.

Within the medical sciences, pluralism is most conspicuous in psychiatry and in the acceptance of some alternative medicine. In psychiatry today, many types of psychotherapies and their underlying theories are used, but as late as during the 1960s they were in many countries classified as unscientific and being quackery. For a long period, electro-stimulating (ECT) and pharmacological treatment were the only psychiatric therapies sanctioned by official science.

In this subchapter we intend to remark only on pluralism within the scientific communities themselves. In most modern societies, scientific communities are allowed to be quite autonomous and self-contained, but they are not absolutely autonomous; and there is really much to discuss about how to obtain an optimal relationship between society and science, but we will not enter this discussion. We will only mention the two most obvious links: (a) the curriculum for primary schools and (b) laws and regulations for scientifically based professions.

(a) In modern societies, science is the public knowledge authority, but it is not allowed all by itself to decide what the primary schools shall teach as being simply facts, i.e., scientific facts. Here, political authorities have the last word, even though they normally respect what is the majority view within the scientific community. But not always, as is shown by the fight around creationism in the US; in several states there has been political quarrel about whether Big Bang cosmology and evolutionary biology shall be taught as facts, or whether they and Biblical creationism should be presented as merely two competing explanations.

(b) The discussion around alternative medicine and therapies such as chiropractics, acupuncture, reflexiology, homeopathy and magnetic therapy illustrates that pluralism sometimes need to be pluralism with explicitly stated limits. The medical sciences are related to the medical professions, which, in turn, are related to the encompassing society and its laws. Many countries have ‘quackery laws’, which means that in these countries there is not only an informal consensus among the medical scientists where to draw the line between medical science and non-science, but that there are also formal laws that state where parts of this line are to be found. The quack does not represent a competing scientific theory or paradigm; he represents a lack of science. Of course, fallibilism implies that one can be mistaken even in judgments about quackery.

Pluralism in university science has two areas: teaching and research. With respect to rival theories there is both the question ‘which of these rivals should be taught as being truthlike, possibly truthlike, and completely obsolete, respectively?’ and the question ‘which of these rival hypotheses should be given funding for research?’ We will mostly, however, make our remarks without bringing in the distinction between teaching and research; that is, many of our remarks are quite abstract. First, in order to understand the way pluralism functions in science, one has to realize that there are two different forms of pluralism:

- acceptive pluralism, i.e., pluralism *without* conflict and competition
- competitive pluralism, i.e., pluralism *with* conflict and competition.

Let us take a brief look at pluralism in dressing habits, pluralism in politics, and ecumenical movements. Pluralism in dressing often means ‘you can dress the way you want, and I will dress the way I want’; and then there is no debate about who wears the proper kind of clothes. All kinds are accepted; this is acceptive pluralism, and it is not the way a democracy works. Political pluralism is not meant to take away debates from the political arena. On the contrary, it is meant to encourage competition and discussions between the different political parties. Independently of whether one wants to say that democratic political parties compete about who is best suited to create a good society for all, or that they compete about who should be allowed to favor their pet social strata in society, it is

clear that democracy means competitive pluralism. Ecumenical movements, on the other hand, attempt to get rid of conflict and competition; both the Christian ecumenical movement and macro-ecumenical movements for all religions aim at turning the respective competitive religious pluralisms into acceptive pluralisms.

Now, what about the scientific community? Given its modern relative independence from the rest of society, should it be pluralistic? Trivially, it should contain and accept the multiplicity that comes with the division of labor between the different scientific disciplines. That is: 'you can work the way you want in your field, and I will work the way I want in mine'. There is no need to discuss this acceptive pluralism; let it only be noted that even here there may be competition about money. In such disputes, each discipline has to argue that it is more important to support its specific kind of knowledge than the others. In what follows we will be concerned only with the question of pluralism within one single discipline or specific domain of research.

Pluralism by definition implies a lack of consensus – against the background consensus of the pluralistic framework itself. But is it acceptable for a mature scientific discipline to lack consensus? Many academics have taken the picture of science that Kuhn paints in *The Structure of Scientific Revolutions* (Chapter 2) as implying, that a discipline that has a long-lasting pluralism cannot be a mature scientific discipline. We think that this is a false conclusion. It is not warranted by the history of science, and it seems not to take into account the fact that basic empirical theories as empirically underdetermined (Chapter 3). Therefore, our questions will be:

- should pluralism in a specific scientific domain be acceptive or competitive pluralism?
- if competitive, does it make more than metaphorical sense to appoint winners and losers in scientific contests?

From our fallibilist point of view, it is obvious that if there is a multiplicity of conflicting scientific hypotheses around a certain question ('how does the heart functions?', 'how do the planets move?', etc.) then there ought to be a competitive pluralism. Each scientist should try to show

that he has the most truthlike theory. To be a fallibilist is to have an epistemologically split vision on truth, and see that even if one as a researcher is wholeheartedly dedicated to the truth of one's own theory, one might in principle be wrong and some opponent right. If one is looking for truth but expects to find only truthlikeness, then there is no ground for claiming that one has monopoly on truth, and in general think that all one's opponents have not made real but merely perverted scientific inquiries. But since fallibilists aim at truth, a merely acceptive pluralism is of no use.

What does prototypical positivism and social constructivism, respectively, have to say about pluralism? True positivists aim at infallible knowledge. Now, if – but only if – a positivist thinks that he holds such a precious piece of knowledge, then there is for him no scientific reason (but perhaps personal) to be tolerant towards researchers with deviating views. In this sense, positivism contains in its theoretical kernel a view that can be used for defending a fundamentalist position in science; be then the actual positivists tolerant or not. True social constructivists, on the other hand, contain in their theoretical kernel a view that makes them tend heavily towards an *acceptive* pluralism. Since they do not believe in any notion of truth as correspondence between a theory (a social construction) and what the theory is about, competitions about truth and truthlikeness appear to them to be competitions around an illusion. They can, though, without contradicting their fundamental view, compete about who has made the most original or the most provocative construction.

In sum, we can write as follows about pluralism in scientific teaching and research:

- positivism might imply non-pluralism
- social constructivism naturally implies an acceptive pluralism
- fallibilism naturally implies a competitive pluralism

This being noted, we will in what follows stick to fallibilism and the form of pluralism it promotes: competitive pluralism. Our next question then is: does it make more than metaphorical sense to appoint winners and losers in scientific contests? How should they be chosen? For instance, does it make sense to try to solve truth-conflicts such as those listed below (when they are actual) by some kind of voting procedure among scientists?

- Galen's theory of the functions of the heart and the blood versus the modern biomedical theory
- The miasma theory of diseases versus the contact theory
- The homeopathic theory versus the biomedical paradigm
- The meridian theory of acupuncture versus the biomedical paradigm
- Psychosomatic theories versus the biomedical paradigm
- Copernicus' theory of the solar system (with circular orbits) versus Kepler's (with elliptical orbits)
- The phlogiston theory of combustion versus the oxygen theory
- The special theory of relativity versus non-relativity theories
- The Darwinian theory of evolution versus the creationist theory
- The flat earth theory versus the sphere theory.

In one sense it is possible to take vote on truth and truthlikeness, but in another sense it is completely impossible. First some words about the impossibility. Truth and degrees of truthlikeness are *relations* between a hypothesis/paradigm and what the hypothesis/paradigm is presumed to be about (Chapter 3.5); and these relations exist or do not exist independently of how any group of scientists vote. This notwithstanding, it is of course just as possible to take vote on *truth proposals* as it is to take vote on political proposals. It is, for instance, possible to vote about whether evolutionary biology or Judaeo-Christian-Muslim creationism should be taught as a truth in schools and universities, but *this voting does not settle the truth*. If everyone votes the same, then there is complete *consensus about correspondence* between theory and reality; if people vote differently there is only a majority view about this correspondence. But in neither case does the vote infallibly guarantee the truth.

In the history of science, there have been few formal votes, but this fact does not speak against the views we have ventured. Even though the development of science has in fact taken place by an innumerable number of informal decisions about what shall be counted as truth, the same changes could have taken place by means of formal voting. The development of early science can very well be described *as if* scientists had now and then voted about competing hypotheses and paradigms.

We have earlier (Chapter 3.5) pointed to the fact that applying the notion of truthlikeness to science allows one to think of scientific achievements the way engineers think of technological achievements. If a machine functions badly, engineers should try to improve it or invent a new and better machine; if a scientific theory has many theoretical problems and empirical anomalies, scientists should try to modify it or create a new and more truthlike theory. Now we want to add: it is possible to vote about what machine to use and what scientific theory to teach in schools, while waiting for better things to come in the future.

The kind of 'voting about truth' that we have now talked about has to be kept distinct from an epistemologically less problematic thing: 'voting about a classification schema *when all the facts are agreed upon*'. It might be said that on August 24, 2006, the General Assembly of the International Astronomical Union voted about whether our solar system contains nine or eight planets (is Pluto a planet or not?), that the '8-planets party' won, and that the 'Pluto-party' lost. But this is a very misleading description of what took place. What was really at stake was what classification schema for bodies orbiting the Sun that astronomers should use in the future. According to the schema that won, Pluto cannot be regarded as a planet, since it does not fulfill the third requirement in the new definition of a planet. A planet (1) orbits around the Sun, (2) because of its own mass it has an almost round shape, and (3) it has cleared the neighborhood around its orbit. Pluto is now classified as a 'dwarf planet'; there are also two further categories of objects orbiting the Sun: 'satellites' and 'small solar system bodies'. One kind of criticism waged at the new definition of 'planet' is that it will be very difficult to apply in other solar systems; this criticism might profitably be called methodology-pragmatic.

The voting had as its background new facts, the discovery of many very small bodies at the outskirts of our solar system. How should they be classified? Nonetheless, the voting in question was not due to any disagreement whatsoever about any facts about Pluto and the rest of the solar system. It was like a voting whether one should call people longer than 185 cm or longer than 190 cm 'tall', or whether one should measure lengths in yards or meters; and the history of science is replete with meetings where scientists have voted about what classification schemas

and measuring units to use, even though they cannot be given such a dramatic (mis-)description as that concerned with Pluto.

In medicine, some classification schemas are debated and decided upon in the same purely methodology-pragmatic way as in the case of the schema for solar system bodies, but sometimes votings about classification schemas are inextricably mixed with votings about truths (what the facts of the matter are). They can even be mixed with votings about norms. The latter complication is due to the close connection between disease-disorder concepts and conceptions of normality and how things ought to be. Especially when it comes to classifications of mental diseases-disorders, it can be hard to tell whether the scientists that are voting do agree on all the facts or not, i.e., whether the different opinions are only due to different methodology-pragmatic views or whether different truth-beliefs and different norm-acceptances are at work as well.

A long story could for instance be told about how, over the years, the American Psychiatric Association (APA) has voted about homosexuality in relation to different versions of the Diagnostic and Statistical Manual of Mental Disorders (DSM); sometimes with gay activist groups demonstrating at APA meetings. Until the end of 1973, homosexuality was classified as ‘sexual deviation’, but then DSM-II was modified and homosexuality called ‘sexual orientation disturbance’; it was now considered a mental disorder only if it was subjectively disturbing to the individual. DSM-III (1980) had at first a notion of ‘ego-dystonic homosexuality’, but in a revision (DSM-III-R, 1987) this notion was eliminated. The present DSM-IV-TR (2000) does not include homosexuality per se as a disorder, but it permits the diagnosis ‘sexual disorder not otherwise specified’ for persons who have a ‘persistent and marked distress about sexual orientation’.

Having now made clear that it makes as much good sense to vote about truth proposals as it does to vote about political parties and their proposals, let us present three problems connected to the paradigm form of competitive pluralism: democracy. All three problems will then be shown to have a counterpart within the pluralism of science.

Where there is competitive pluralism there must be some formal or informal way by means of which it is decided who the winners and the losers are. At bottom, in all formally democratic elections the voters

decide, but then there has to be (a) formal rules that tell who is to be allowed to vote; let us call this *the maturity problem*. Today, it is mostly seen as a matter merely of age, but during parts of the twentieth and the nineteenth centuries it was very much also a matter of profession (should unemployed and low-income people be allowed to vote?), sex (should women be allowed to vote?), and race (should black people be allowed to vote?). Apart from these kinds of rules, there also has to be (b) formal rules for how to transform all the votes to an appointment of winning parties or persons. In both cases, the rules can look different in different countries. In the (b)-case, we meet a special problem, *the proportionality problem*. Consider the following situation. There is a nation whose parliament has 100 members. From the very idea of representative democracy it might seem to follow that each political party that gets around one percent of the votes ought to have one member in the parliament. Why then do many countries have a lower limit (mostly between 3 and 6 percent), a fence, which says that parties that fall below this limit are not allowed to be represented in the parliament? The reason is that a democracy should not only meet the requirement of representing the wills of the voters, it has to meet the requirement of being a practically efficient decision procedure too. Democracy should be democracy in this world, not in an ideal supernatural world. And it is assumed that too many very small parties might often, with bad repercussions, block the parliamentary decision procedure.

A third problem well known in theories of democracy is called '*the paradox of democracy*'. It is a special case of 'the paradox of freedom', which says that universal freedom cannot for logical reasons always be defended, since if everyone is free to do whatever he wants, everyone is free to enslave others, and then the latter are not free. That is, as soon as someone tries to enslave someone else, one has to constrain the freedom of the constrainer. The paradox of democracy says: all political parties cannot for logical reasons be given democratic rights in all logically possible situations, since if parties that want to overthrow the democracy will win a majority, they will abolish democracy. That is, as soon as a totalitarian party is on the brink of winning a parliamentary election, democrats are faced with the dilemma of either in an undemocratic way forbidding this

party or in a democratic way let democracy come to an end. Absolute democracy is as much a myth as absolute freedom is.

We will next comment on ‘the paradox of democracy’, ‘the proportionality problem’, and ‘the maturity problem’ as they appear in science.

According to the view we have presented, defended, and think is implicitly adhered to by a majority of scientists, there are many kinds of reasoning and arguments in science that intertwine in the final decisions. Empirical sciences can often use some deductive inferences, but they have in the main to rely on much weaker forms of argumentation patterns. In particular, they have somewhere to bring in *argumentation from perceptions*. This means that at least empirical science cannot regard any holy book as completely authoritative. What totalitarian parties are to democracy, holy-book-scientists are to science. As the former are anti-democratic, the latter are anti-scientific. The *paradox of scientific pluralism* can be stated thus:

- For logical reasons, all kinds of scientists cannot be allowed a place in science in all logically possible situations, since if scientists who want to make a certain book overthrow the reliance on systematic observations will take the lead, they will abolish empirical science; in other words: absolute scientific pluralism is a myth.

We can imagine that a majority of voters vote for a party promoting the abolishment of democracy. But what would the corresponding thing look like in science? Well, one way it could take place is that a group of holy-book-scientists manage by ordinary scientific argumentation to convince a majority of ordinary scientists that a holy book of some sort contains all interesting truths. If such an incredible development would one day become reality, then ordinary scientist would face with the dilemma of either in an unscientific non-argumentative way force some scientists to leave the scientific community or let science become heavily restricted.

Usually, people in the creationist movement and in the flat earth movement are today regarded as the two most conspicuous cases of people who simply do not take science seriously. But there is an interesting difference between them. Most creationists can be said to rely completely

and only on a literal reading of the Bible, and these are the ones we are highlighting now, but the majority of the members of the flat earth movement have so far had no similar epistemological views; even if they nonetheless may look crazy. (It should be added and noted that some creationists have a view that was quite common among medieval theologians: truths about nature revealed by the Bible can be shown to be true *also* by means of man's natural reason. Such creationists try intensely to find anomalies in evolutionary biology; and if such anomalies there are, they ought of course to be seriously discussed.)

In many cultures long ago, Earth was taken to be as flat as it appears in most everyday perceptions, but already the thinkers of ancient Greece found out that this seems not to be the case. The latter noticed that the star constellations were different in Greece and Egypt, and that the shadow of the Earth on the moon at lunar eclipses indicates that Earth is a sphere. If Earth is a sphere, then some of our everyday observations must be illusory; e.g., water surfaces must be bent but look flat. On the other hand, if the flat earthers are right, then some other perceptual observations, such as the shape of the line that separates the dark and bright part of the moon, have to be given quite complicated explanations.

The modern flat-earth movement was launched in England with the publication in 1849 of a sixteen page pamphlet, *Zetetic Astronomy: A Description of Several Experiments which Prove that the Surface of the Sea Is a Perfect Plane and that the Earth Is Not a Globe!* Its author, Samuel Birley Rowbotham (1816-1884), repeatedly emphasized the importance of sticking to the facts. He called his system 'zetetic astronomy', because he sought only *facts*; 'zetetic' comes from the Greek verb *zetetikos*, which means to seek or inquire. He wrote: "Let the practise of theorising be abandoned as one oppressive to the reasoning powers, fatal to the full development of truth, and, in every sense, inimical to the solid progress of sound philosophy." Later, in a very empirically minded book, *Earth Not a Globe* (1881), Rowbotham claimed to have shown in a number of measurements that there is no curvature in the water of big lakes. The modern International Flat Earth Society was created in the 1950s, in the 1990s it had around 3000 members, but since then it has declined. The truly hard times for its members arose when the satellite pictures of the Earth were publicized. But the flat earthers did not say that their view is

beyond empirical control. Instead, they argued that all these photographs, as well as the whole trip to the moon in 1970, were mere movie fakes.

Whereas the holy-book-scientists of the creationist movement can be used to illustrate the *paradox* of scientific pluralism, the flat earth movement can illustrate the *proportionality problem* of scientific pluralism; not every idea that argues in a formally scientific way need to be allowed a place in science. However, we will not here try to find any default rules for where such fences for empirical theories ought to be installed. The third problem mentioned, the *maturity problem*, i.e., who is to be reckoned a full citizen in a scientific community, seems at first not be a hard problem. A default rule could be that at least all doctoral students and all more qualified persons should be such citizens. But then one forgets that they ought also to have reflected a little on fallibilism and the need for pluralism in science. Otherwise, they may be just as dogmatic as holy-book-scientists and scientists who think they have found an infallible truth can be; or they may be just as acceptive as social constructivists allow their adherents to be.

We would now once more like to highlight what we have labeled ‘type T error’ (Chapter 6.4). The type T error is the mistake to neglect a significant correlation and a useful treatment because its first protagonists claimed the data in question are caused by a causal mechanism that science cannot validate. Its importance for pluralism can be seen by the fact that many Western physicians by committing this type of error vis-à-vis acupuncture (in the 1950s and 60s) made the acceptance of this therapy unnecessary hard. According to our analysis, they should have interpreted the implausible meridian mechanism away, and then looked at the remaining pure correlation hypothesis. A mechanism theory can be looked upon in two ways, either in its entirety with its mechanism or without the mechanism. This makes it possible, as we will show, to widen the pluralism in the medical sciences. Look at the matrix in Table 1 (of course, both the dimension ‘high-low statistical effect’ and the dimension ‘plausible-implausible mechanism’ take degrees, even though these degrees are abstracted away:

	The statistical effect of the treatment is:	
	High	Low
The underlying mechanism is plausible:	1	2
No underlying mechanism is proposed:	3	4
The underlying mechanism is implausible:	5	6

Table1: *Six different kinds of epistemological situations for a hypothesis.*

Assume that we are medical scientists who, based on reflections around causal mechanisms that are well established in the medical-scientific community, have proposed a new kind of treatment; that we have tested it according to normal methodological standards; and that we have a high statistical effect. That is, we sit at the moment quite safely somewhere in box 1 in the matrix. From this perspective, the question then is: how tolerant ought we to be towards colleagues who at the moment are in the other boxes? Let us add that these colleagues are firmly convinced that sooner or later in the development of science they will enter box 1 too.

The mentioned proportionality principle seems to make it reasonable for us to create toleration fences somewhere on the various possible roads down to box 6. Let us for a moment forget that the dimensions in the matrix take degrees and assume that, in fact, there are only extremely high and low correlations and only extremely plausible and implausible mechanisms. The proportionality principle would then, reasonably, allow us to forbid all theories in the second column (boxes 2, 4, and 6), and seemingly also box 5, but what about the hypotheses in box 3? Answer: they should be allowed, since the statistical effect is high and they are not connected to any obscure mechanisms. But this acceptance has an interesting further consequence in relation to box 5. Since we can take a theory from box 5 and abstract away the presumed mechanism, i.e., turn the mechanism theory into a black-box theory, we can so to speak move a theory from box 5 to box 3. But this seems to imply that we ought also to be somewhat tolerant towards people working with theories in box 5. After all, it was probably their mechanism thinking that made them hit upon what we happen to regard as a good black-box theory.

We will end this subchapter with still another analogy between voting in the political sphere and voting in science. Assume that there is a nation where the two largest parties, P and Q, both present themselves as creating, *in the long run*, the best possible life for all the citizens. *In the short run*, however, from a social-economical point of view, the P-party explicitly favors stratum A and disfavors stratum B, whereas the Q-party takes it the other way round. From a common sense psychological point of view, one might then claim that it is easier for citizens in stratum A than for citizens in stratum B to reach the conclusion that the P-party is the best party for the whole nation; and vice versa for people in stratum B. But to make the whole picture clear, two more things should be noted. First, the socio-psychological fact noted states only that there is such a tendency; it does not mean that each and every member of stratum A is pre-determined to vote for the P-party. Second, the future may well tell (or at least indicate) whether or not the party who gets the power makes life better for all citizens or only for people in stratum A. This way of looking at political elections can be transferred to science.

Assume that in a scientific discipline two paradigms, P and Q, are competing about what paradigm is the most truthlike one. Paradigm P has been constructed by research groups in the so-called ‘A-universities’ and Q in the ‘B-universities’. Doctoral students at the A-universities will then have a tendency to think that the P-paradigm is more truthlike than the Q-paradigm; and vice versa for the doctoral students at the B-universities. But the doctoral students are not pre-determined to support the majority view at their home university. Furthermore, the future may well contain observations that make almost everyone think that one of the paradigms is much more truthlike than the other one. Let us repeat a sentence from Chapter 3.5: Fallibilism makes it possible to adhere simultaneously to the views that (i) science aims at truths, (ii) captures partial truths, and (iii) accepts theories partly because of the way these conform to dominant views in the surrounding society.

8.3 Methodological pluralism

What comes first, the hen or the egg? Similarly, one may ask: what comes first, the object of study (theory) or the methodology? Even if in the latter case, for sure, the two might in principle come simultaneously, there is a

kind of asymmetry in favor of theory. Think of the difference between molecular biology and studies in comparative literature. In molecular biology one can and has to do experiments with matter, and it makes good explanatory sense to try to find smaller units than molecules. In comparative literature, on the other hand, it is completely senseless to make laboratory experiments, and if one looks at too small pieces of texts the overarching meanings of the texts that should be compared are lost. Here, the choice of object of study comes before the choice of methodology. However, on a more concrete level where one works with very specific hypotheses against the background of an already chosen general theory, the concrete methodological rules might be said to come first. These two levels have to be kept apart if the theory-methodology interplay in science shall not be turned into a low-traffic one-way street.

As stated in Chapter 2.4, Kuhn first distinguishes between two parts of a paradigm, (i) a disciplinary matrix and (ii) exemplars/prototypes, and then between three components of the matrix: (ia) metaphysical assumptions, (ib) symbolic generalizations, and (ic) values. The values might equally well be called ‘general methodological norms’. Our comment in the last paragraph means that we regard the metaphysical assumptions and the symbolic generalizations as being logically prior in relation to the general methodological norms. And this is quite in conformance with Kuhn’s views. He thinks that even the basic empirical sciences have to contain assumptions that are not directly based on their experiments and observations, i.e., the methodologies of experiments and empirical data gathering are only possible provided some presuppositions. Applied to the clinical medical paradigm (as expounded in Chapter 6), these views imply that the methodology of the randomized controlled trial has to be regarded as being subordinated to the epiphenomenalist materialist ontology of the biomedical paradigm.

This fact, in turn, means that paradigm conflicts between, for instance, the biomedical paradigm and those of alternative medicine cannot be solved by a simple recourse to methodology. To the same extent that pluralism in science should reckon with pluralism between paradigms, it has to reckon with methodological pluralism too.

It is often said that science requires a certain critical mass of researchers in order to achieve anything of real interest. Therefore, it may be a good

thing to monopolize a methodology. If many researchers are working with the same methodology, the chances of getting skilled and methodologically creative researchers increase, and this is needed in order to show what the basic theory really is able to accomplish. If the researchers must learn many different methods, they may not learn all of them properly. But, on the other hand, it is equally true that such a monopoly might hamper the growth of knowledge. What to do? Answer: competitive pluralism.

A long standing topic in the philosophy of science has been the opposition between quantitative and qualitative methods. It dates back to the rise of modern science, which is the rise of mathematical physics. Galilei, one of its true heroes, claimed straightforwardly: *the book of nature is written in the language of mathematics*. He distinguished strictly between the *primary qualities*, which are properties of the material things in nature, and *secondary qualities*, which are properties belonging to entities in the mind such as (perceived) colors, sounds, smells, pleasures, and pains. The former, he claimed, are the object of physics, and the latter cannot be systematically studied. In the same era, great philosophers such as John Locke and Descartes made similar ontological bipartitions. During the second half of the twentieth century, this dichotomy was turned into a distinction between working with *hard data* and *soft data*, respectively.

In this latter debate, the notion of hard data fused the ontological feature of being able to be represented by means of numbers with the epistemological property of providing pretty secure knowledge, and the notion of soft data fused the feature of not being able to be quantified with the epistemological property of providing only subjective interpretations. When the ontological and the epistemological aspects are kept distinct, however, the remarks below can be made. They end with the conclusion that the opposition and perceived difference between the two methods have been considerably exaggerated.

First epistemology; from the fallibilistic perspective of this book no gap or clear-cut distinction between secure knowledge and subjective interpretations can be found. All empirical data include some kind of interpretation. This fact lessens in itself the importance of the epistemological part of the debate. But even more, nothing in principle says that quantitative measurements are always more certain than what one can obtain from, say, interviewing people.

Our second and ontological remark says: *every empirical science has to be concerned with some qualitative features*. Pure numbers exist only in mathematics; quantities have to be kept distinct from mathematical numbers. Let us explain.

Quantitative expressions such as ‘a mass of 4 kg’, ‘a length of 9 m’, ‘a temperature of 37 ^{o}C’ consist of three easily distinguishable linguistic parts:

1. a name of a physical dimension (‘mass’, ‘length’, and ‘temperature’, respectively)
2. a so-called numeral (4, 9, and 37, respectively); only *in abstraction from the rest of the quantitative expression* does the numeral refer to a mathematical number
3. a name of a conventionally chosen standard unit (‘kg’, ‘m’, and ‘^{o}C’, respectively); this unit cannot be chosen until the determinate properties of the physical dimension in question has already been ordered the way numbers are ordered on the number line.

Much can and has been explored about formal and other features of scales and quantities, but we will not go into such details in this book. Here we only want to point to the fact that each basic physical dimension is a purely qualitative feature that exists independently of whether or not it becomes quantified. That is, the quantities 4 kg, 9 m, and 37 ^{o}C are *quantifications of* the qualitative physical dimensions mass, length, and temperature, respectively. The ordinary opposition between ‘qualities’ and ‘quantities’ is at bottom an opposition between ‘*qualities* that cannot be quantified’ and ‘qualities that can be turned into *quantities*’. In everyday language, expressions often become condensed. Regardless of whether or not pain research will succeed in quantifying pain, pain will still remain a quality. If pain researchers succeed, it will enable us to rank and more easily compare different pains.

For some unclear reasons, philosophers of science have now and then written as if talk about physical dimensions would become linguistically superfluous as soon as the determinate quantities with the standard unit have entered language. But this cannot be true, since if we want to make

everything philosophically explicit we need to state trivial ‘laws’ such as these:

- no object can possibly at one and the same time take two values of the same physical dimension; e.g., no object can have a mass of both 4 kg and 2 kg
- only quantities of the same physical dimension can be added together and give rise to a meaningful sum; e.g., ‘4 kg + 9 m’ has no meaningful sum.

A good case can even be made against Galilei’s dictum that nature is written in the language of mathematics. Shape is just as much a basic property of material things as size and mass are, but so far no one has managed to construct for all determinate shapes something resembling even remotely the quantifications of size and mass. The physical dimension of shape contains not only good-behaving geometrical shapes, but each and every kind of curious and crazy-looking determinate shape there is; two-dimensional as well as three-dimensional.

In principle, it could have been the case that even though shape is a basic property of material things it is not important in science. However, this is not at all so. Two rhetorical questions ought to be enough to make this fact clear: (i) ‘what would anatomy be without shape talk?’ and (ii) ‘what would molecular biology be without shape talk?’ The discovery of the DNA molecule is partly a discovery of its shape.

The basic difference between so-called quantitative and qualitative *methods* can now be spelled out as follows: when we deal with quantitative methods, the qualities are taken for granted and we look for and focus on the quantitative aspect; when we apply pure qualitative methods, we are either not interested in quantitative aspects or the qualities are not amenable to quantitification. It should also be mentioned that the natural sciences might involve qualitative methods, and the humanities quantitative, even though the natural sciences are dominated by quantitative methods and the humanities by qualitative. If we may refer back to our case study of Harvey’s scientific practice (Chapter 4.8), it now illustrates our point that research can contain both qualitative and quantitative elements.

When quantification is possible, however, it makes comparisons in the domain at hand extremely easy and efficient. When you state that something has ‘a mass of 4 kg’, ‘a length of 9 m’, ‘a temperature of 37 ^{o}C’ you obtain in each case an infinite number of comparisons for free. A similar advantage has systematic classifications and taxonomies, even though they give rise only to a finite number of comparisons. To these topics we will return in Chapter 11.

8.4 Pluralism from the patient’s perspective

This little subchapter is inserted only to forestall the possible misunderstanding that patients naturally should have the same view on medical pluralism as the medical scientists and clinicians have. Let us take a quick look at the different perspectives.

The medico-scientific community needs some competitive pluralism in order to efficiently produce good diagnostic tools, treatments, and preventive measures. In each case the result is something that should be applicable to a whole population of some kind, i.e., the populations that have disease A, that have a risk of getting disease A, that have disease B, that have a risk of getting disease B, etc. As tax payers financing medical research, patients have the same interest in pluralism as the medical community has, but what about the patients *as patients*?

When one is sick, one primarily wants to become healthy. One is not especially interested in the whole population that happens to have the same disease. Existentialists used to say things such as ‘my death is mine’ and ‘my life is mine’, stressing that an individual can never be regarded as only and completely part of some community or other. In the same vein it is true for patients to say: ‘my disease is mine’ and ‘my possible curing is mine’. What follows from this observation? On an abstract level, mainly two things:

(a) black box theories count as much as mechanism theories
(b) lottery thinking (‘I may win even though the probability is very low’) can be rational

(a) From an existentialist point of view, the mechanisms (pathogenesis and etiology) behind the disease and its curing are of no special interest. If

something has cured me as an individual, then afterwards I am happy. I don't care much whether it is due to a well known biomedical mechanism, to a completely unknown mechanism, to a placebo effect, or even to an effect (e.g., homeopathic) that the medical community finds unbelievable.

(b) Much has been investigated about how to be rational in lotteries and situations in life that have a similar stochastic structure. To put one conclusion simply, it seems impossible to say that a certain behavior is rational in itself. To some extent it always depends on whether you are an extremely cautious person or a person who is fond of taking risks; and then there are all the in-betweens. Therefore, it is impossible to say that a patient who spends much money on a treatment that has a low statistical effect is irrational; especially if the gains are high, i.e., that the disease is very serious.

The last remark notwithstanding, there is in both case (a) and case (b) a kind of rationality at work, but one that is mostly adhered to by the patients themselves. It seems rational to start with treatments that have high statistical effects based on plausible mechanisms (slot 1 in Figure 1), but the interesting thing is what happens if such treatments fail. There are then, it seems, very good reasons to try black box theories, and if even these fail one may well opt for implausible theories. That is, we obtain the slots order 1, 3, and 5 in Figure 1. However, having accepted to play the medical lottery game, one may continue with the slots down the second column, i.e., 2, 4, and 6. It is in all probability often this kind of reasoning, rather than reasoning about the nature of medical science that makes it attractive for patients to consult quacks.

Reference list

Bunge M. Phenomenological Theories. In Bunge (ed.) *The Critical Approach to Science and Philosophy*. Free Press of Glencoe. New York 1964.

Johansson I. Pluralism and Rationality in the Social Sciences. *Philosophy of the Social Sciences* 1991; 21: 427-443.

Kvale S. To validate is to question. In Kvale S. *Issues of Validity in Qualitative Research*. Studentlitteratur. Lund 1989.

Lynoe N, Svensson T. Doctors' attitudes towards empirical data - a comparative study. *Scandinavian Journal of Social Medicine* 1997; 25: 210-216.

Mitchell S D. *Biological Complexity and Integrative Pluralism*. Cambridge University Press. Cambridge 2003.

Wulff H, Pedersen SA, Rosenberg R. *Philosophy of Medicine – an Introduction*. Blackwell Scientific Press. Cambridge 1990.

9. Medicine and Ethics

Outside of very well integrated groups, it is hard to legislate about word meanings. In everyday discourse, words tend to take on a life of their own independent of any stipulations. This is important to remember when it comes to the words ‘ethics’, ‘morals’, and ‘morality’. Even though some philosophers define these nouns in such a way that they receive distinct meanings (‘ethics’, for instance, is often defined as ‘the philosophy of morals’), in many everyday contexts they are synonymous. Etymologically, ‘ethics’ comes from the ancient Greek words ‘ethikos’ and ‘ethos’. The latter meant something like ‘the place of living’ and the former ‘arising from habit’. That is, both have to do with custom or practice. ‘Morals’ and ‘morality’ are derived from Latin words such as ‘moralis’, ‘mos’, and ‘mor-’ meaning ‘custom’. Today, what is moral or ethical is often contrasted with what comes about only by habit or custom.

The words ‘ethics’ and ‘morality’ are today sometimes used in a purely descriptive sense and sometimes in a normative. To say ‘their ethics (ethical system, morals, or morality) contains the following main rules: …’ is mostly only meant to describe in a neutral way what rules a certain group adheres to, whereas to say ‘this is what the laws dictate, but this is what ethics requires’ is to use ‘ethics’ in a normative sense. Normally, to call someone ‘a morally responsible person’ is to say that he wants to act in accordance with the right norms. But to call someone ‘a moralist’ is to say that he is a bit too keen on judging the moral behavior of others.

Where a moral system is accepted, it regulates the life of human beings. It even regulates how animals are treated. Through history, and still today, we find different cultures and subcultures with different and conflicting rules of how one should to act. Philosophers have tried to settle these conflicts by means of reasoning, but so far they have not succeeded in obtaining complete consensus even among themselves. Therefore, in some of the subchapters below, we will present a common tripartite classification of philosophers’ ethical positions:

- deontology (deontological ethics or duty ethics)
- consequentialism (one main form being utilitarianism or utilitarian ethics)
- virtue ethics.

A moral-philosophical system gives an elaborate presentation and defense of what should be regarded as morally right and wrong actions as well as morally good and bad persons, properties and states of affairs. During this presentation, we will argue in favor of a certain kind of modern virtue ethics, but we hope nonetheless that we make such a fair presentation of duty ethics and utilitarian ethics, that the readers are able to start to think on their own about the merits and demerits of these positions. Our views affect, let it be noted, how we look upon medical ethics. We will not deal with ethical nihilism and complete ethical subjectivism, i.e., the positions that end up saying that there are no super- or inter-personal moral constraints whatsoever on our actions. If some readers find all the ethical systems we present absurd, then there are reasons to think that they are either ethical nihilists or subjectivists.

Medical ethics is often, together with disciplines such as business ethics, environmental ethics, and computer ethics, called ‘applied ethics’. This label gives the impression that medical ethics is only a matter of applying to the medical field a set of pre-existing abstract moral principles. And so it is for deontological and consequentialist ethicists, but is not for virtue ethicists. The reason is that virtue ethics does not basically rely on verbally explicit moral rules. In Chapter 5 we argued that epistemology has to extend beyond ‘knowing that’ and take even the existence of ‘knowing how’ into account. Here, we will argue that, similarly, ethics has to take its kind of know-how into account too.

In *The Reflective Practitioner*, Donald Schön describes the relationship between theory and practice (or ‘knowing that’ and ‘knowing how’) as follows:

> In the varied topography of professional practice, there is a high, hard ground where practitioners can make effective use of research-based-theory and technique, and there is a swampy lowland where situations are confusing 'messes' incapable of technical solutions.

We regard this as being as true of medical ethics as of medical practice. Even medical ethics is influenced by the special topography of swampy lowlands; too seldom are there any straight ethical highways in this area.

Medical ethics may be defined as a multidisciplinary research and education discipline that historically, empirically, and philosophically scrutinizes moral and ethical aspects of health care in general, clinical activities, and medical research. It evaluates merits, risks, and social concerns of activities in the field of medicine. Often, new medical discoveries and clinical developments imply new ethical considerations. Just think of the development of DNA-technology, the cultivation of stem cells and cloning, prenatal diagnosis, fetus reduction, and xenotransplantation. Clinical ethics also bring in the physician-patient relationship, the physician's relations to relatives of the patient, relations between doctors and other health care professionals, and relations between doctors and the society at large. Medical ethics is part of biomedical ethics, which also includes things such as environmental ethics, animal rights, and the ethics of food production.

Moral philosophers have not only been discussing ethical systems and applied ethics. During the twentieth century, especially the latter half, Anglo-American philosophers were preoccupied with what they call 'meta-ethics', i.e., problems that are related to ethics but nonetheless are 'beyond' (= 'meta') ethics. Typical meta-ethical issues are (i) analyses of moral language (e.g., 'what does it *mean* to claim that something is morally good?'), and (ii) the questions whether there can be objectivity and/or rationality in the realm of morals. Some philosophers claim that such meta-ethical investigations are completely neutral with respect to substantive ethical positions, but others contest this and say that meta-ethics has repercussions on substantive ethics; we align with the latter ones. Meta-ethicists who believe either that moral statements are true or false or that (even if not true or false) their validity can be rationally discussed, are

called ‘cognitivists’; non-cognitivism in meta-ethics easily implies ethical nihilism in substantive ethics.

Since, in philosophy, the term ‘meta’ means ‘coming after’ or ‘beyond’, also mere presentations of moral and moral-philosophical systems can be regarded as a kind of meta-ethics. To such a meta-ethics we now turn.

9.1 Deontology

The Greek ‘deon’ means duty. Classical deontological ethics categorically prescribes that certain well defined actions are obligatory or are prohibited. A prototypical example is the Ten Commandments or the Decalogue from the Old Testament, which have played a dominant moral role in Christianity and (to a lesser extent) in Judaism. Here are some of these commandments (we do not give them number, since there is no culture independent way to do this):

- Honor your father and your mother!
- You shall not kill (human beings)!
- You shall not steal!
- You shall not bear false witness against your neighbor!
- You shall not covet your neighbor’s wife!
- You shall not covet your neighbor’s house or anything that belongs to your neighbor!

As they stand, these imperatives tell us what to do independently of both context and consequences. In no way can a deontologist say what Groucho Marx is reported to have said: ‘Those are my principles; if you don’t like them I have got others!’ The duties stated are meant to be *absolute* (i.e., they do not allow any exceptions), *categorical* (i.e., they are not made dependent on any consequences of the action), and *universal* (i.e., they are not, as nowadays presented, norms only in relation to a certain culture). In this vein, the commandment ‘You shall not kill!’ says that I am not under any circumstances allowed to kill anyone, even in a case of euthanasia, not even myself, whatever consequences of suffering I may then have to endure; and this is so irrespective of which culture I happen to belong to.

Two more things should be noted. First, it is harder to know exactly how to act in order to conform to the obligations (e.g., 'Honor your father and your mother') than to conform to the prohibitions (e.g., 'You shall *not* kill, steal, etc.'). Second, the last two commandments in our list do not prohibit actions, they *prohibit desires*. Such norms will be left aside. Like most modern moral philosophers, we will restrict our discussions to rules concerned with how to act.

In relation to every imperative of the form 'Do A!' it is formally possible to ask: 'Why should I do A?'. If the answer is 'It is your duty to do B, and doing B implies doing A', then it becomes formally possible to ask: 'But why is it my duty to do B?' If the next answer brings in C, one can ask 'Why C?', and so on. In order to justify a substantive moral norm, one has in some way to end this justificatory regress somewhere. If there are duties at all, then there has to be at least one duty that is self-justificatory.

To traditional Christian believers, lines of moral justifications end, first, with the Ten Commandments, and then – absolutely – with the answer: 'It is your duty to follow the Ten Commandments because God has said it is your duty', period. Most philosophers, however, find this answer question-begging. They want an answer also to the question 'Why should it be my duty to do what God says it is my duty to do?' The philosopher Immanuel Kant (1724-1804) is famous for having tried to exchange 'God' for 'a will enlightened by reason', a *rational will*. Although being himself a firm believer, Kant wanted to make the basic ethical norms independent of religion. In this undertaking, he claimed to have found a central principle, The Categorical Imperative (soon to be presented), by means of which other presumed categorical norms could be tested. At the end of the justificatory line that Kant proposes, we find: 'It is your duty to follow The Categorical Imperative and the commandments it gives rise to because every rational will wants this to be its duty'. In order to arrive at what we will label 'Kant's Four Commandments', Kant argued somewhat as in the brief reconstruction below; the words are for a while put directly into Kant's mouth.

Step 1: All basic moral norms have to be linguistically formulated as *categorical imperatives*: 'Do A!' or 'Don't do A!'. Hypothetical

imperatives such as '*If* you want A, *then* you have to do B!' and '*If* you want A, *then* you cannot do B!' can never in themselves state a basic moral norm. Why? Because they make the prescription dependent on a pre-given goal (A), and if this goal is a mere subjective whim, the prescription has nothing to do with morals. On the other hand, if there is a categorical norm that requires the doing of A, then this norm bears the moral burden, not the hypothetical imperative 'if A then B'. The imperative 'If you want to be a good doctor, then you have to develop your clinical and ethical skills' is a hypothetical imperative; it does not state that you have to try to become a good doctor.

Step 2: A basic moral norm cannot pick out anything that is spatiotemporally specific. Sensuous desires can be directed at one or several particular spatiotemporal objects ('I want to play with him and no one else!'), but reason and rational wills can directly be concerned only with a-temporal entities such as concepts, judgments, logical relations, and principles. Therefore, no *fundamental* moral rule can have the form '*This* person should do A!', '*I* should do A', or 'Do A at *place x* and *time t*'. Basic norms must be universal and have the form 'Do A!' or the form 'Don't do A!' Therefore, they cannot possibly favor or disfavor a particular person as such.

Step 3: If there is a moral norm, then it must in principle be possible for persons to *will* to act on it. There can be morals only where there is freedom; purely instinctual reactions and other forms of pre-determined behavior is neither moral nor immoral. Since the basic norms have to be stated in the form of categorical (step 1) and universal (step 2) imperatives, the last presupposition of a special 'causality of freedom' has the following consequence: in relation to a norm, it must not only be possible that I as a particular person want to conform to it, it must also be possible that the norm in question is *willed by everyone*. Therefore, when you are trying to find out whether you can will to act in conformance with a certain 'action maxim', you cannot think only of yourself, you have to ask yourself whether you think that the maxim can in principle be collectively willed. If your answer is 'yes', you can will that it becomes a law for everyone. That is:

- Act only according to that maxim (= categorical general imperative) by which you can also will that it would become a universal law.

This is *The Categorical Imperative*. It is not a substantial imperative on a par with the Ten Commandments, but a test that substantial imperatives have to pass in order to be able to count as duties.

Step 4: Since the requirement stated is very formal, it should be called 'the *form* of the categorical imperative'. Nonetheless, we can derive from it a more substantial formulation, 'the *matter* of the categorical imperative':

- Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end.

This imperative of Kant is similar to the Christian Golden Rule: 'Do unto others as you would have them do unto you' (Jesus, according to Matthew 7:12), but it is clearly distinct from the negative imperative 'Do *not* unto others as you would *not* have them do unto you', which exists in other religions too.

Kant's imperative does not say that we are never allowed to use another human being as means; only that we are never allowed to use another person *only* as a means. On the other hand, this absolute respect for the autonomy of persons means that we are not even allowed to reduce our own autonomy, which, for instance, we can do by drinking alcohol or taking drugs.

When Kant derives the second formulation of his categorical imperative from the first, he seems to rely heavily on what he puts into his notion of a 'will'. According to Kant, only *persons*, i.e., mature responsible ('mündige') human beings, can have the kind of will that is mentioned in the first and formal formulation of the categorical imperative. Such a free will can create an end for itself merely by willing it, and Kant seems to falsely think that since persons are able to freely create ends for themselves, persons are *thereby* also ends in themselves. In Kant's opinion, animals are not persons even though they can have conscious perceptions,

desires, and feelings of pleasure and pain. His reason is that they cannot be persons since they cannot self-consciously will anything.

(Out of the two imperatives stated, Kant then derives a third formulation, which has quite a political ring to it. It says that it is always one's duty to act as if one were 'a legislating member' in 'the universal kingdom of ends', i.e., in the formal or informal society constituted by all persons. This is Kant's *form-and-matter* formulation of his categorical imperative.)

Step 5: Kant gives four examples of categorical general imperatives (maxims) that he thinks pass the test of the form of the categorical imperative, and which, therefore, should be regarded as absolute duties. They are: 'Do not commit suicide!', 'Develop your talents!', 'Do not make false promises!', and 'Help others!' Unlike the duties of the Decalogue, these commandments are not 'duties for God's sake', they are 'duties for duty's sake'. To a true Kantian, the question 'Why is it my duty to do what my rational will tells me to do?' is a nonsensical question. All that can possibly be said has already been said. Moral justification has reached its end point.

In the second formulation of the categorical imperative, Kant uses the expression (with our italics) 'treat humanity, whether in *your own person* or in the person of *any other*'. Some of one's duties are concerned only with oneself, and some of them are concerned only with other persons. If this distinction is crossed with another distinction that Kant makes, namely one between *perfect* and *imperfect* duties (soon to be commented on), then his four substantial norms can be fitted into the following fourfold matrix:

	Duties to oneself	*Duties to others*
Perfect duties	Do not commit suicide!	Do not make false promises!
Imperfect duties	Develop your talents!	Help others!

Table 1: *Kant's Four Commandments.*

The two *perfect duties* are prohibitions, and the two *imperfect duties* are obligations. The distinction may have something to do with the fact that it is often easier to know how to conform to a prohibition than how to act in order to fulfill an obligation, but it may also have to do with the fact that – no doubt – one can conform to the perfect duties even when one is asleep or is taking a rest. Imperfect duties in Kant's sense must not be conflated with what in moral philosophy is called 'supererogatory actions', i.e., actions that are regarded as good or as having good consequences, but which are *not* regarded as *duties*. Donating money to charity is mostly regarded as a supererogatory action; there is no corresponding duty. One may also talk of actions that are 'juridically supererogatory'; they are not required by the law but are nonetheless in the spirit of the law. It might be noted that most patients who praise the job of their doctor, would probably rather like to hear him respond by saying 'I worked a bit extra hard because you were suffering so much', than simply 'I did it out of duty'. This is not odd. Many patients want a somewhat personal relationship, but duties are impersonal. To work, out of empathy, more and harder than duty requires is good, but it is a supererogatory goodness.

It is hard to see and understand exactly how Kant reasons when he tests his four commandments against The Categorical Imperative and lays claim to show that they have to be regarded as absolute duties. Nonetheless we will try to give at least the flavor of the way he argues; we will in our exposition take 'Do not make false promises!' as our example.

In Chapter 4.2, we presented the structure of 'reductio ad absurdum' arguments. In such arguments, a certain view or hypothesis is shown to imply a logical contradiction or to contain some other absurd consequence; whereupon it is claimed that, therefore, the negation of the original view or hypothesis has to be regarded as true. Kant tries to do something similar with imperatives.

Let us assume (contrary to what we want to prove) that no one has to take the imperative 'Do not make false promises!' as stating a duty. This means that no one would feel compelled to fulfill their promises; and, probably, also that many people would allow themselves to make false promises as soon as they would benefit from it. It may then happen, that people so seldom fulfill their promises that no one dares to take a promise seriously any more. In a community where this is the case, the citizens can

no longer make any real promises; they cannot even fool anyone by making a false promise, since the whole institution of promising has become extinct. (Another similar case is a society where each and every coin might be a fake, and every person knows this. In such a society, there would be no real money any more, and, because of this, nobody could fool anyone with fake money.) According to Kant, it is absurd for a rational person to will that his community should contain the possibility of having no institution of promising. Therefore, a rational will must will that the maxim ‘Do not make false promises!’ becomes a universal law.

In the modern world, bureaucrats and other appointed rule watchers can sometimes spontaneously vent a Kantian way of thinking. Often when they become confronted by a person who wants them to make an exception to the rules they are meant to apply, they ask rhetorically: ‘Now, look, what do you think would happen if I and my colleagues should always do like this?’ When asking so, they think or hope that the person in front of them shall immediately understand that this would mean – with very negative consequences – the death of the whole institution. ‘What do you think would happen with the parking system’, the parking man asks, ‘if I and my colleagues did not bother about the fees?’ The dummy way to understand Kant’s universalistic moral reasoning is to take him to ask rhetorically in relation to every person and every culture:

- What would your society, and thereby your own life, look like if everyone gave false promises all the time?
- What would your society, and thereby your own life, look like if everyone committed suicide?
- What would your society, and thereby your own life, look like if everyone neglected his talents all the time?
- What would your society, and thereby your own life, look like if no one ever helped any other person at all?

Modern physicians have to fulfill many roles, often also that of the bureaucrat and the rule watcher. Therefore, they may often say to patients: ‘What do you think would happen if I and my colleagues should always do the way you want me to do?’ A patient may, with some good reasons, request his doctor to give him sick leave even though he is not suffering

from any well-defined disease; well knowing that it is possible for this specific doctor to deceive the regional social insurance office. But if all physicians provide false certificates whenever a patient benefits from such a certificate, medical certificates would probably after some time become completely futile.

A contemporary American philosopher, Alan Gewirth (1912-2004), has tried to ground norms in a way that makes him similar to Kant. He claims that human agency is necessary evaluational, in the sense that all actions involve the agent's view of his purposes as being good; at least in the sense that he wants to see them realized. Now different agents have, of course, different purposes, but common to all agents, according to Gewirth, is that they must regard the freedom and well-being necessary to all successful action as necessary goods. Hence, since it would be logically inconsistent for an agent to accept that other agents may interfere with goods that he regards as necessary, he must claim rights to freedom and well-being. And since it is his being an agent that makes it necessary for him to make this claim, he must also accept the universalized claim that all agents equally have rights to freedom and well-being. Accordingly, a universalist theory of human rights is derived from the individual agent's necessary evaluative judgement that he must have freedom and well-being.

The main problem for classical deontology is its inflexibility. In relation to almost every categorical norm, there seems to be possible cases where one ought to make one's action an exception to the norm. For instance, most people believing in the Ten Commandments have with respect to 'You shall not kill!' made an exception for wars and war-like situations; and in relation to the norm 'Don't lie!' everyday language contains a distinction between 'lies' and 'white lies'. Especially in medicine, it is quite clear that physicians sometimes can help patients by means of (white) lies. A related problem is that it is easy to conceive of situations where deontological norms (say, 'Help others!' and 'Don't lie!') conflict; in such a situation one has to make an exception to at least one of the norms.

A famous modern attempt to leave this predicament behind but nonetheless remain a deontologist (i.e., not bring in the consequences of one's actions when evaluating them) has been made by W. D. Ross (1877-1971). His view is best understood by means of an analogy with the particle mechanics of classical physics. If, by means of this theoretical

framework, one wants to understand in principle or predict in detail the motion of a specific particle (call it ‘Alfa’), one obtains situations such as the following four. (i) Alfa is affected only by a gravitational force from one other particle (Beta). To follow this force (Newton’s law of gravitation) will then, metaphorically speaking, be the duty of the particle. However, (ii) if Alfa is similarly affected by the particle Gamma, then it is its duty to adapt to the gravitational forces from both Beta and Gamma. Furthermore, (iii) if Alfa, Beta, and Gamma are electrically charged, then it is the duty of Alfa to adapt to the corresponding electric forces (Coulomb’s law) too. In other situations, (iv) there can be many more particles and some further kinds of forces. All such partial forces, even those of different kinds, do automatically combine and give rise to a determinate motion of Alfa. In classical mechanics, this adding of *partial forces* is represented by *the law for the superposition of forces*. By means of this law, the physicists can add together all the partial forces into a *resultant force*, by means of which, in turn, the movement of Alfa can be directly calculated.

Let now Alfa be a person. According to Ross, he can then in a certain situation both be subject to many kinds of duties (partial forces), which Ross calls ‘prima facie duties’ (i.e., duties on ‘first appearance’), and have duties towards many persons (particles). When Alfa has taken the whole situation into account, there emerges an *actual duty* (‘resultant force’), which Alfa has to perform quite independently of the consequences both for him and for others. Without claiming completeness, Ross lists six classes of prima facie duties: (1) duties depending on previous acts of one’s own such as ‘Keep your promises!’ and ‘Repair your wrongdoings!’; (2) duties having the general form ‘Be grateful to those who help you!’; (3) duties of justice; (4) duties of beneficience; (5) duties of self-improvement; and (6) duties of not injuring others. A specific prima facie duty becomes an actual duty when it is not overridden by any other prima facie duty.

Ross’ move may at first look very reasonable, but it contains a great flaw. No one has so far been able to construe for this ethical system what the principle of superposition of forces is in classical mechanics. That is, in situations where more than one prima facie duty is relevant, Ross supplies no rule for how to weigh the relevant duties in order to find out what our actual duties look like; here he resorts to intuition. As discussions after Ross has shown, there are even good reasons to think that no

‘superposition principle for prima facie duties’ can ever be construed. We will present this case in Chapter 9.3.

*****

Even if Ross’ proposal to exchange absolute duties for prima facie duties should, contrary to fact, be completely without problems, it leaves a big moral problem untouched. It takes account only of how one man should find out how to act morally in a given particular situation, but modern states and institutions need verbally explicit general norms in order to function; states need written laws. How to look upon such necessary rules from a moral point of view? If singular actions can be morally right, wrong, or indifferent, the same must be true of institutional norms; and it seems incredible to think that all of them are morally indifferent. We now meet from a societal perspective the kind of justificatory regress that we have already pointed out in relation to the Ten Commandments and Kant’s categorical imperative. It now looks as follows.

Institutional laws might be regarded as justified as long as they stay within the confines of the state in which they are applied. Most modern societies then make a distinction between ordinary laws and some more fundamental and basic laws, often called ‘constitutional laws’. The latter not only regulates how ordinary laws should be made (and, often, also how the institutional laws themselves are allowed to be changed), they are also meant in some way to justify the procedures by means of which the ordinary laws come into being. But what justifies the constitutional laws themselves? Shouldn’t the power that enforces our laws be a morally legitimate power? Even if some of the constitutional laws can be regarded as being merely hypothetical imperatives, i.e., rules that function as means to a pre-given goal such as to find the will of the citizens, this cannot be true for all of them; at least not if everything is made explicit. Then there should (in the exemplified case) be a categorical norm that says that the citizens shall rule. When there is a bill of rights connected to a constitution, then the justificatory question appears at once: how are these rights justified? For instance: how is the Universal Declaration of Human Rights of the UN justified?

This problem is not necessarily a problem only in relation to the society at large. Several medical rules have a moral aspect to them, and this needs moral justification. Neither the rule to respect the autonomy of patients nor the rule that requires informed consent in clinical research are rules that are meant only to be efficient means for the work of doctors and researchers.

A way to solve this rule-justificatory problem in a non-religious and in a partly non-Kantian way has been proposed by the two German philosophers, Karl-Otto Apel (b. 1922) and Jürgen Habermas (b. 1929). It is called '*discourse ethics*', and they regard it as a modern defense of deontology. Instead of trying to ground the basic categorical norms in what a rational will has to will, they try to ground them in what a group of people who want to communicate with each other have to regard as an ideal form of behavior. Like most late-twentieth century philosophers, they think that language is necessarily a social phenomenon, and this creates in one respect a sharp contrast between their position and that of Kant. The free will and the reason that Kant speaks of can belong to a mature person alone, but the pragmatic language principles and the communicative reason that Apel and Habermas focus on are necessarily inter-personal.

We will now first state Habermas' version of the discourse ethic counterpart to Kant's first (formal) formulation of The Categorical Imperative, and then try to explain why discourse ethicists think that some pragmatic language principles are at one and the same time both necessary presuppositions for linguistic communication (discourses) and categorical imperatives. Like Kant's categorical imperative, Habermas' so-called '*principle of universalization*' is a formal principle against which all more substantial norms are meant to be measured. Whereas Kant says that 'a norm is valid if and only if: it can be rationally willed that it would become a universal law', Habermas says:

- a norm *is valid* if and only if: the consequences and side effects of its general observance for the satisfaction of each person's particular interests are acceptable (without domination) to all.

How is this deontological universal principle taken out of the wizard's hat? Let us look at it in the light of the norms (a) 'When talking, do this in order to be understood!', (b) 'Don't lie!', and (c) 'Give other people the

rights that you yourself claim!' Most individual language acts take something for granted. If someone asks you 'Have John stopped smoking yet?', he takes it for granted that John have been a smoker; otherwise his question makes no sense. Therefore, you might reason as follows: (1) there is this question about John; (2) *how is it possible* to put forward?; (3) answer: the speaker presupposes that John has been a smoker. In a similar way, Apel and Habermas make inferences from more general facts to presuppositions for the same facts. Their so-called 'transcendental inferences' can be schematized as follows:

(1) fact: there is communication by means of language
(2) question: how is such communication possible?
(3) answer: by means of (among other things) pragmatic language principles
(4) norm derivations: (a) if people never give the same words the same meaning, language stops functioning, therefore the pragmatic language principle 'You ought to give words the same meaning as your co-speakers!' is at the same time a valid norm; (b) if everybody lies all the time, language stops functioning, therefore the pragmatic language principle 'Don't lie!' is at the same time a valid norm.

These norms are said to be to be *implicitly present* in our language. What the discourse ethicists mean by this claim might be understood by means of an analogy (of ours). Assume that you are a zoologist that has discovered a completely new species and that after careful observations you have come to the conclusion that all the individuals you have observed have to be regarded as sick. One might then say that you have found an *ideal* (to be a healthy instance of the species) implicitly present in these really existing (sick) individuals. Discourse ethicists have come to the conclusion that human languages have an ideal built into their very functioning. In the ideal language community there are no lies. When people lie or consciously speak past each other, they do not kill language, but they make it deviate from its in-built ideal.

Here is another discourse ethical 'transcendental inference', and one whose conclusion comes closer to Habermas' principle of universalization:

1. fact: there are argumentative discourses
2. question: how are such discourses possible?
3. answer: by means of (among other things) pragmatic language principles that are specific for argumentative situations
4. norm derivation: (c) if *by arguments* you try to show that your opponent does not have the same rights in the discussion as you have, then you have to accept that there may be arguments that take away your own right to free argumentation; therefore, the pragmatic principle 'You should in discussions recognize the other participants as having rights equal to your own!' is at the same time a valid norm.

The last norm says that the ideal for discussions is that they are free from domination; they should only contain, to take the German phrase, 'Herrschaftsfreie Kommunikation'. If this norm is accepted, then it seems (for reasons of philosophical anthropology) natural to claim that all norms that can be communicatively validated have to take account of, as Habermas says: 'the satisfaction of each person's particular interests'. This means, to use the earlier analogy, that we cannot make our sick language healthy, and have an ideal communicative community, without a normative change in the whole community.

Habermas' principle of universalization is impossible to apply in real life situations (since it talks about *all* persons, consequences, and side effects), and he is aware of this fact. But he claims that this principle can function as a regulative idea that make us try to move in the right moral direction, and that we ought to approximate the principle when we put forward constitutional laws, bills of rights, or other kinds of basic moral norms. In order to come a bit closer to real life he has put forward another principle, which he calls 'the principle of discourse ethics':

- a norm can *lay claim to validity* if and only if it meets (or can meet) with the approval of all affected in their capacity as free and equal participants in a practical discourse.

In practice, we will never be able to finally validate a norm. Therefore, we have to rest content with a principle that tells us when we can at least lay claim to validity. According to the principle of discourse ethics, no number of thought experiments can be more validating than a real communicative exchange around the norm in question. For instance, without speaking with other people in your community you cannot find out whether promise breaking should be regarded as always forbidden or not. Note, though, that even if every discussion of such a norm takes place under certain given circumstances, it should aim (according to discourse ethics) at finding a moral norm whose validity is not culture bound, but is universal.

Discourse ethics is like Kant's duty ethics an ethical system that lays claim to be (i) deontological, (ii) in its centre formal, (iii) universalistic in the sense of not being culture bound, and (iv) cognitivist, since it thinks that the validity of moral statements can be rationally discussed. It differs from Kant in having exchanged a person-located reason for an inter-personal communicative reason.

At last, we want to point out a difference between Apel and Habermas. Apel is very much like Kant a philosopher who thinks he has found something that is beyond all doubt, whereas Habermas is a fallibilist. The problem for Apel is that he can *at most* lay claims to having shown that we are faced with an alternative: either to shut up or to accept the existence of norms. But then his norms are not true categorical imperatives, only hypothetical imperatives ('*if* you want to use language, *then* …'). Habermas is an outspoken fallibilist, who thinks that his principles may contain mistakes, but that, for the moment, they cohere quite well with most of our present moral particular judgments. In this opinion, he comes close to John Rawls' more elaborate views on what fallibilism in moral matters amount to. For our presentation of this view, and Rawls' famous notion of 'reflective equilibrium', the reader has to wait until Chapter 9.3; let it here only be noted that the very term 'moral fallibilism' is used neither by Habermas nor by Rawls.

9.2 Consequentialism

'Consequentialism' is an umbrella term for ethical systems that claim that it is only or mainly the *consequences* of a particular action (or kind of

action) that determines whether it should be regarded as being morally right, wrong, or neutral. Each specific consequentialist system has to state what kind of consequences should be regarded as being good, bad, or indifferent. In so far as good and bad consequences can be weighed against each other, the natural consequentialist counterpart to Kant's categorical imperative (the second formulation) can be stated thus:

- Among the actions possible for you, choose the one that maximizes the total amount of good consequences and minimizes the total amount of bad consequences that you think your action will (probably) produce.

In retrospect, one may come to the conclusion that what one thought was the objectively right action in fact was not so. One had estimated the consequences wrongly. Nonetheless, the basic imperative must have this form. It is impossible to act retrospectively, and one has to act on what one believes. The term 'probably' is inserted in order to make it explicitly clear that there is never in real life any complete and certain knowledge about the consequences spoken of. Rather, the epistemological problems that pop up when consequentialist imperatives shall be applied are huge. Because of these problems, there are in some contexts reasons to discuss whether or not one should stick to one of two more unspecific principles, one (the first below) states the thesis of 'positive consequentialism' and the other the thesis of 'negative consequentialism':

- Act so that you produce as much good consequences as possible.
- Act so that you produce as little bad consequences as possible.

The second principle is often given the formulation 'we demand the elimination of suffering rather than the promotion of happiness' (Karl Popper). It is claimed that at least politicians should stick to it. Why? Because, it is argued, if they make false predictions about the outcome of their policies, the consequences of acting on the reducing-suffering principle cannot be as disastrous as they can be when acting on the creating-happiness principle.

Most important and most well known among consequentialist systems are the utilitarian ones. We will briefly present four of them; they differ in how they define what is good and bad. They are (with their most famous protagonist within parenthesis):

1. simple hedonistic utilitarianism (Jeremy Bentham, 1748-1832)
2. qualitative hedonistic utilitarianism (John Stuart Mill, 1806-1873)
3. ideal utilitarianism (George Edward Moore, 1873-1958)
4. preference utilitarianism (Richard M. Hare, 1919-2002; Peter Singer, b. 1946).

According to simple hedonistic utilitarianism, pain and only pain is bad, and pleasure and only pleasure is good (or has utility). The term ‘utility’ is a bit remarkably made synonymous with ‘pleasure’; ‘hedone’ is the Greek word for pleasure. Normally, we distinguish between many different kinds of pleasures (e.g., in sex, in work, in music, etc.) and pains (aches, illnesses, anxieties, etc.), but according to Bentham this is at bottom only a difference in what causes the pleasures and pains, respectively. These can, though, differ with respect to intensity and duration. When the latter factors are taken into account, utilities can, he thinks, in principle be added (pains being negative magnitudes), and hedonic sums for amounts of pleasures and pains be calculated; the utility calculus is born. He even thinks that factors such as the certainty and the proximity of the expected pleasures and pains can be incorporated in the calculus. If we leave the latter out of account, his central thought on these matters can be mathematically represented as below.

Let us associate all pleasures with a certain positive number, p, and all pains with the negative number, -p. Furthermore, let us assume that each different degree of intensity of pleasure and pain can in a reasonable way be associated with a certain number, i_l. Let us then divide the life of a certain person during the temporal interval T into m number of smaller time intervals, called t_1 to t_m; these intervals shall be so small that they do not contain any changes of pleasures, pains, or intensities. All these things taken for granted, the actual total pleasure or the hedonic sum – $h_a(T)$ – for the life of the person *a* in the interval T can be mathematically represented

by formula 1 below. Formula 2 represents the aggregated pleasure for all sentient beings in the interval T:

$$(1)\quad h_a(T) = \sum_{n=1}^{n=m} t_n \cdot (i_l \cdot p) \qquad (2)\quad H(T) = \sum h_a(T)$$

(Formula 1 should be read: the total pleasure of person *a* in time interval T equals the sum of the pleasures in each time interval, 1 to m; the pleasure in each such interval being given by multiplying the length of the time interval with the intensity of the pleasure and the value of the pleasure. Formula 2 should be read: the total pleasure of all sentient beings in T equals the sum of all their individual pleasures.)

Bentham thinks it is our duty to try to maximize H(T) for a T that stretches from the moment of action as long into the future as it is possible to know about the consequences of the action. His categorical imperative is *the utility principle*:

- Among the actions possible for you, choose the one that maximizes the utility that your action produces.

Bentham makes it quite clear that he thinks that even animals can experience pleasure and pain, and that, therefore, even their experiences should be taken into account in the 'total amount of pleasure and pain' spoken of. Utilitarianism brought from the start animals into the moral realm. Nonetheless there is an ambiguity in the utility principle: should we relate our maximizing efforts only to the pleasures and pains of those sentient being that already exist (and probably will exist without extra effort on our part), or should we try to maximize pleasure absolutely, i.e., should we even try to create new sentient beings *only* for the purpose of getting as much pleasure in the world as is possible? In the latter case we might be forced to maximize the number of members of those species whose members normally feel more pleasure than pain during their lives, and to minimize the number (might even mean extermination) in species whose members normally feel more pain than pleasure (even if they are human beings). In what follows we will mainly write with the first

alternative (‘the prior existence view’) in mind, even though the second alternative (‘the total amount view’) is literally closer to the utility principle.

Despite utilitarianism’s explicit concern with animals, Bentham’s view is often regarded as being possible to condense into the so-called ‘greatest happiness principle’:

- One should always act so as to produce the greatest happiness for the greatest number of people.

As should be obvious from the formulas 1 and 2, the task of comparing for a certain action all possible alternative hedonic sums, H(T), or even a very small number of them, is an infinite task. For instance, how are we supposed to reason when applying the utility principle in clinical practice; especially in emergency situations which do not leave much room for reflection? Only extremely rough and intuitive utility estimates can be made – and questioned. Nonetheless, the utility principle immediately stirred the minds of social reformers. And it is easy to understand why: the principle makes no difference between different kinds of people, e.g., between noble men and ordinary men. Pleasures and pains are counted quite independently of what kind of person they reside in. The toothache of a beggar is regarded as being quite as bad as a similar toothache of a king. A famous dictum by Bentham (that leaves animals aside) is: ‘each to count for one, and none for more than one’.

John Stuart Mill was heavily influenced by Bentham and his utilitarian father, James Mill, who was a close friend of Bentham. But Mill the junior found the original utilitarianism too simple. Pleasures, he claimed, can differ not only in their causes and their intensity, but in *kind*, too. Broadly speaking, in Mill’s opinion, there are two qualitatively distinct realms of pleasures. One lower realm of physical or almost-physical kinds of pleasures, some of which we share with animals, and one higher realm. The latter contains more intellectual kinds of pleasures, pleasures that are typical of educated human beings. And Mill was not alone. Many of Bentham’s contemporaries regarded his utilitarianism as a doctrine worthy only of swine or of small children. Shouldn’t we distinguish between the sensibilities of a Socrates, an infant, or a pig? The pleasure of playing

push-pin, it was said, is always a lower kind of pleasure than the pleasure in reading good literature. Put in modern jargon, Mill's view is that the pleasures of sex, drugs, and rock-and-roll should count lower in a utility calculus than the pleasures of reading, writing, and listening to Mozart. The higher pleasures were said to give rise to happiness, the lower ones merely to contentment.

To illustrate further the difference between Bentham and Mill we might look at a thought experiment provided by Roger Crisp. He asks you to think that you are about to be born, and are given the choice between being born as an oyster or as a great composer such as Haydn. Your life as an oyster would, in all probability, be a safe, soft, and long life; you would be floating in a moderate sensitive pleasure without any pains for some two hundred years. Your life as a composer, on the other hand, would probably be very intense, complex, creative, and exciting, but also much shorter and contain phases of severe suffering. What would you choose? Bentham would perhaps choose to be born as an oyster, but Mill would surely prefer the shorter life of a Haydn. Even though the oyster-versus-Haydn example is extreme, it might nonetheless illustrate a problem in clinical medicine: is it better for terminally ill patients to stay alive in a long persistent vegetative state, or is it better for them to live a more ordinary life for only a brief period of time?

What we have termed *simple* hedonistic utilitarianism is sometimes called *quantitative* utilitarianism. We have avoided the latter label, since it gives the false impression that Mill's qualitative utilitarianism, which posits the existence of *qualitatively different kinds* of pleasures, cannot be given a mathematical representation. But it can. All we have to assume is that in a reasonable way we can associate with each and every *kind* of pleasure and pain a certain positive (pleasure) or negative (pain) number, p_k. Mill's higher pleasures should then be attributed large positive numbers and the lower pleasures small positive numbers. On the assumption of such a pleasure-and-pain (utility) scale, and on the assumptions earlier stated, the two utility formulas presented are transformed into the following ones (formula 1' for the total amount of pleasure of one person in the interval T, and formula 2' for the aggregated pleasure of all sentient beings):

$$(1') \quad h_a(T) = \sum_{n=1}^{n=m} t_n \cdot (i_l \cdot p_k) \qquad (2') \quad H(T) = \sum h_a(T)$$

(Formula 1’ should be read: the total pleasure of person *a* in time interval T equals the sum of the pleasures in each time interval, 1 to m; the pleasure in each such interval being given by multiplying the length of the time interval with the intensity of the pleasure and the value of the specific kind of pleasure in question. Formula 2’ should be read: the total pleasure of all sentient beings in T equals the sum of all their individual pleasures.)

Qualitative hedonistic utilitarianism differs from simple utilitarianism not only in its acceptance of many different kinds of pleasures, since this difference has at least two repercussions. First, the existence of different kinds of pleasures makes the original utility calculus much more complicated. Second, the difference affects what persons can be set to make utility calculations. Every person who constructs a utility scale for qualitative utilitarianism must be familiar with every kind of pleasure and pain that he ranks, which means that only persons who are familiar with both the higher and the lower kind of pleasures can make the calculations. Mill took it for granted that those who know the higher pleasures normally also know all the lower pleasures, and that he himself could rank all kinds of pleasures. This might be doubted.

Both Bentham and Mill were, like Kant, part of the Enlightenment movement, and all of them tried, each in his own way, to connect moral thinking with political thinking in order to promote the creation of good societies.

Mill’s line of thinking about utilitarianism was taken one step further by Moore, who argued that some ideal things are good without necessarily being connected to any pleasure. According to Moore, having knowledge and experiencing beauty are good independently of the pleasure they may cause. One consequence of this view is that since ‘ideal utilities’ are (normally considered to be) quantitatively incommensurable with ‘pleasure utilities’, to accept ideal utilitarianism is to drop the whole of classical monistic utilitarianism in favor of a pluralistic utilitarianism.

To claim that one is interested in (has a preference for) having knowledge even when it does not gives rise to pleasure, is to introduce a

distinction between *pleasure* and *preference satisfaction*. Such a distinction was foreign to the classical utilitarianism of Bentham and Mill. They thought that everything that human beings directly desire can adequately be called pleasure, and that every true desire satisfaction gives rise to some pleasure. But, with Moore, we may well question this view. Therefore, many utilitarian philosophers have tried to create forms of utilitarianism where utility is defined in terms of preference satisfaction instead of pleasure. Of course, these philosophers take it for granted that we often have pleasure from preference satisfactions and pain from preference dissatisfactions; the point is that a utility calculus *also* has to take account of preference satisfactions and dissatisfactions where this is not the case. For instance, if a physician has a preference for being regarded as a good physician but *unknowingly to himself* is regarded as bad, shouldn't this preference dissatisfaction of his count, even though there is no experienced displeasure?

The shift from pleasure utilitarianism to preference utilitarianism does not affect the general formulation of the utility principle; it only redefines utility. Unhappily, however, neither does the shift solve the two outstanding internal problems of utilitarianism:

- How to rank all the different kinds of preference satisfactions (utilities)?
- Given such a ranking, how to make a useful utility calculation in real life?

One obvious objection to preference utilitarianism is that our conscious preferences do not always mirror our real preferences. Sometimes, to our own surprise, we do not at all become pleased when a preference of ours becomes satisfied. For instance, one may be perfectly convinced that one wants to read a good book alone at home in the evening, and so one satisfies this conscious preference. But when one is sitting there reading the – no doubt – good book, one discovers that this was not at all what one really wanted to do. How should preference utilitarianism take such facts into account? It has been argued that only preferences based on informed desires that do not disappear after therapy should count (Brandt 1979). Such amendments, however, make the epistemological problems of

utilitarianism grow even more exponentially than they did with the introduction of Mill's qualitative utilitarianism.

A peculiar problem for utilitarians is that they have to try to take account also of the fact that there are many people who intensely dislikes utilitarian thinking. Since the latter become deeply dissatisfied when utilitarians act on a utility calculation, the utilitarians have to bring even this fact into their calculations. One famous utilitarian, Henry Sidgwick (1838-1900), has argued that from a utilitarian point of view it is hardly ever right to openly break the dominating norms of a society. In a weaker form, the same problem appears every time a utilitarian intellectual shocks his contemporaries with a proposal for a radically new norm. How does he know that the immediate amount of displeasure that his proposal gives rise to among traditionalists, is outweighed by future preference satisfactions that he thinks will come about if his proposal is accepted? Has he even tried to find out? If not, then he either is inconsistent or has implicitly some non-utilitarian norm up his sleeves.

However, reflections by utilitarians can have quite an impact in spite of the fact that most consequence estimations have better be called consequence conjectures or consequence speculations. This is so because a utilitarian's views on what should be regarded as having equal utility, or being equal amounts of preference satisfaction, can contradict the traditional moral outlook of his society. With the claim that pleasures and pains ought to be considered independently of who has them, classical utilitarianism questioned values that were central to feudal societies. Modern preference utilitarianism, particularly in the hands of Peter Singer, has continued a utilitarian tradition of re-evaluation of contemporary common sense morality. In Singer's case, the moral reform proposals arise from the way he connects certain anthropological views with a specific evaluation of the difference between self-conscious long-term preferences on the one hand and desires that want immediate satisfaction on the other. He thinks along the following lines.

Assume that rats like human beings, (a) can have conscious feelings of pleasure and pain, and that they instinctively in each particular moment seeks immediate pleasure and tries to avoid immediate pain; but that they unlike human beings, (b) are completely unable consciously to imagine future states where they have feelings of pleasure and pain or can have an

interest satisfied or dissatisfied. Now, given these assumptions, what is from a preference utilitarian point of view the difference between painlessly killing a rat and killing a human being? Answer: at the moment of the deaths, with the rat disappears *only* the desire for the immediate pleasure satisfactions, but with the human being disappears *also* the desire for the satisfactions of all his hopes, wishes, and plans for the future. With the death of a healthy human being more possible satisfactions of *existing* preferences disappears than with the death of a healthy rat.

Singer claims that both rats and human beings have a *right to physical integrity* because they are able to suffer, but that – with some qualifications – we human beings also have a *right to life* because we normally anticipate and plan our future; to human beings applies 'the journey model of life'. Now the qualifications indicated. According to Singer's (rather reasonable) anthropology, fetuses, very small infants, and severely disabled people lack the anticipatory ability in question. But this implies, he argues, that in certain special circumstances painless abortion and infanticide can be justified. One possible kind of case would be the killing of very small infants whose future life in all probability would be full of suffering both for themselves and for their parents.

To many people these views mean an unacceptable *degrading of some forms of human life*, but this is only one side of Singer's reform coin. On the other side there is an *upgrading of some forms of animal life*. The latter has made him a central thinker in the animal liberation movement. From the premises sketched, he argues that the United Nations should declare that great apes, like human beings, have a right to life, a right to the protection of individual liberty, and a right not to be tortured. Why? Answer: because even great apes can form long-term interests. Not to accept such rights is, Singer says, an act of *speciesism*, which is as morally wrong as racism and sexism. Note, though, that all of Singer's talk of 'rights' are made within a utilitarian, not a deontological, framework.

According to Singer, while animals show lower intelligence than the average human being, many severely retarded humans show *equally low* mental capacity; therefore, from a moral point of view, their preference satisfactions should be treated as equal too. That is, when it comes to experiments in the life sciences, whatever the rules ought to look like, monkeys and human infants should be regarded as being on a par. From his

premises, Singer has also questioned the general taboo on sexual intercourse with animals.

Singer thinks we have to consider equal interests (preferences) equally and unequal interests unequally; whatever species the individual who has the interests belong to. All sentient beings have an interest in avoiding pain, but not in cultivating their abilities. But there are even more differences to take into account. That two individuals of the same species want the same kind of thing does not necessarily mean that they have an equal interest in it. Singer regards interests as being subject to a law of diminishing marginal utility returns. If two persons, one poor and one rich, both want to win 100 000 euros in a lottery, then if the poor wins the preference satisfaction will be higher (and therefore his interest is higher) than if the rich one wins. Similarly, a starving person has a greater interest in food than someone who is just a little hungry. Singer relies on the following reasonable anthropological principle: the more one has of something, the less preference satisfaction one gets from yet another amount of the same thing. Therefore, according to Singer, anyone able to do so ought to donate part of his income to organizations or institutions that try to reduce the poverty in the world.

So far, we have mainly described utilitarianism as if it tries to say how we should think when we don't know what *individual action* is the morally right one. Such utilitarianism is called *act utilitarianism*, but there is also another brand: *rule utilitarianism*. According to the latter, we have only to estimate by means of what rules we can create the greatest possible happiness in the world; and then act on these rules. This means that these rules take on the *form* of categorical general imperatives; the main principle of rule utilitarianism can be stated thus:

- Act only according to those rules by which you think that the greatest happiness for all sentient beings can be achieved.

At first, it might seem odd to judge a particular action by means *not* of the consequences of the action itself, but by the consequences of a corresponding rule. But there is a good reason for the change: the tremendous epistemological and practical problems that surround all utility calculations. These problems seem to diminish if it is the consequences of

rules, rather than the consequences of actions that should be estimated. Singer is a rule utilitarian, and rule utilitarianism was suggested already by John Stuart Mill. The famous preference utilitarian, R. M. Hare (1919-2002), proposes a two-level utilitarianism. He says that 'archangels' can stay on *the critical level* where act utilitarian calculations are made, that 'proles' cannot raise above the *intuitive level* where situations are judged by means of already accepted rules, but that most human beings can now and then try to approximate the archangels.

These considerations apart, there is also a sociological reason for rule utilitarianism. As soon as an act utilitarian accepts that society at large and many of its institutions need explicit norms, he has to accept some rule utilitarian thinking.

It seems impossible that any rule utilitarian would come to endorse Kant's commandment 'Do not commit suicide!', but they may well end up with proposing Kant's three other commandments: 'Develop your talents!', 'Do not make false promises!', and 'Help others!' Similarly, the principle that doctors should treat patients as autonomous individuals might be regarded as justified by both deontologists and utilitarians. A Kantian may say that this principle follows from The Categorical Imperative, which says that we should always treat a person as an end in himself and never simply as a means; and a rule utilitarian may say the principle follows from The Utility Principle, since it obviously has good consequences such as an increase in the trust in the health care system, which, in turn, creates more happiness and health or at least prevent distrust and ill-health. The fact that people who subscribe to different ethical paradigms can make the same moral judgment of many individual actions is perhaps obvious, but the point now is that such people may even be able to reach consensus about moral rules. In everyday life, both these facts are important to remember when a moral-philosophical discussion comes to an end.

In Chapter 2.4 we said that paradigms and sub-paradigms are like natural languages and their dialects. It is as hard to find clear-cut boundaries between a paradigm and its sub-paradigms, as it is to find discontinuities between a language and its dialects; and sometimes the same holds true between two paradigms and between two languages. Nonetheless, we have to make distinctions between different paradigms and different languages. Now, even though it is easy to keep religious and Kantian deontology

distinct form all forms of consequentialism, it takes a real effort to see the difference between the deontology of discourse ethics and the consequentialism of preference rule utilitarianism. Their similarity can be brought out if Habermas' principle of universalization (U) and the preference rule utilitarian formulation of the utility principle (R) are reformulated a little; especially if utilitarianism is restricted to persons. Compare the following statements:

- (U) a rule *is morally valid* if and only if: the consequences and side effects of its general observance for the satisfaction of each person's particular preferences are acceptable to all

- (R) a rule *is morally valid* if and only if: the consequences and side effects of its general observance maximizes the preference satisfactions of all persons.

The principle (U) does not make moral validity directly dependent on consequences but on acceptability, but then, in turn, this acceptability depends on what consequences the rule has. Conversely, there seems to be something odd with the principle (R) if it is not acceptable to most people. As we will see in Chapter 9.3, utilitarianism has had problems with how to justify the utility principle.

Let it here be added that many discourse ethicists will, just like Singer, give rights even to animals. They then distinguish between the communicatively competent persons spoken of in their basic principles, which are both *moral subjects* and *moral objects*, and animals, which at most are moral objects. Persons are both norm validators and objects for norms, but animals can only be objects for norms. Human beings, however, can be moral advocates for beings that lack communicative competence. According to discourse ethics, between moral subjects there are real duties, but between persons and animals there can only be quasi-duties.

From a commonsensical point of view, rules such as 'Do not make false promises!', 'Help others!', 'Don't lie!', 'Don't steal!', and, in medicine, 'Do not make false certificates!' and 'Do not make transplants without consent!', have one positive side and one negative. The positive side is that

when such rules are enforced people can normally trust that others will help them when needed, that they will not be given false promises, and so on; and such trust characterizes good societies. The negative thing is the inflexibility. Indeed, it seems as if we often have good moral reasons to make exceptions to our norms. We will return to this issue in Chapter 9.3.

*****

No form of consequentialism can ever eliminate the problem of judging consequences, since consequentialism holds that it is *only or mainly* the consequences of an action that determines whether it is morally right or wrong. But judging consequences may be important also for other reasons. One often has from an egoistic point of view good reasons to try to find out what consequences one's actions may have; at least if one wants to be a bit prudent. Also, consequences may be given a *subordinate role* in some other kind of ethical system; a fact we have already noted in relation to discourse ethics, and we will return to it later in relation to virtue ethics. This being noted, we will here take the opportunity to present one way of trying to estimate and compare the values of various medical interventions. Such values, be they moral or not, simply have to be considered in clinical medicine and in health care administrations. We will present the method of evaluating consequences by means of so-called 'Quality-Adjusted Life Years' (QALYs). Some optimists hope that one day it will be possible – on a broad scale – to allocate healthcare resources by means of the costs per QALY of different interventions. If Bentham would have been living, he might have tried to estimate the cost per pleasure increase of different interventions.

A life year is of course one year of life, but what then is a *quality-adjusted* life year; and what magnitudes can it take? One year of perfect health for one person is assigned the value 1 QALY, the year of death has the value 0, and living one year with various illnesses, diseases, and disabilities obtain QALY-values between 0 and 1. The values between 0 and 1 are determined by various methods. People can be made to answer questions such as 'how many years in a certain bad state would count as equal to perfect health for some definite number of years?' If ten years in a diseased state is regarded as equal to two years with perfect health

(= 2 QALY), then each diseased year is attributed the value 0.2 QALY. If a terminally ill patient gets 10 years of 0.2 QALY/year with a certain medical intervention, and 3 years of 0.8 QALY/year without it, then, on the assumptions given, he should not be given treatment, since 2.4 QALY is better than 2.0 QALY. Different patients may evaluate different things differently, which means that when it comes to interventions in groups, health care administrations have to work with average QALYs.

Another way of determining what QALY number should in general be associated with a certain health state is to use standardized questionnaires; at the moment the EuroQol EQ-5D is ranked high. Here, the respondents are asked to tell whether they regard themselves as being good (= 1), medium (= 2), or bad (= 3) in the following five health dimensions:

- Mobility (M)
- Self-Care (S)
- Usual Activities (U)
- Pain/Discomfort (P)
- Anxiety/Depression (A).

Every answer is represented in the form of five numbers ordered as <M, S, U, P, A>, e.g., the value set <1, 3, 3, 2, 1>. There is then a ready-made table in which each such value set is associated with a single number between 0 and 1, called ‘the health state utility score’. One year with a utility score of 0.5 is reckoned as bringing about 0.5 QALY. If these measures are accepted, it is possible to compare a treatment A, which generates four years with a utility score of 0.75 (= 3 QALY), with another treatment B, which produces four years with a score of 0.5 (= 2 QALY), and immediately arrive at the conclusion that A is the better treatment; it is 1 QALY better.

QALY measurements of treatments can easily be combined with the costs of the same treatments. A ‘cost-utility ratio’ for a certain treatment is defined as the cost of the treatment divided by its number of QALYs. At the moment when this is being written, kidney transplantations (in the US) are estimated to cost 10.000 USD/QALY, whereas haemodialyses are estimated to cost 40.000 USD/QALY.

9.3 Knowing how, knowing that, and fallibilism in ethics

We can now chart the relationship between, on the one hand, utilitarianism and duty ethics, and, on the other hand, ethical systems that concentrate on evaluating rules or singular acts, respectively. It looks as follows:

	Utilitarianism	*Duty ethics*
Rule evaluating	J. S. Mill	I. Kant
Act evaluating	J. Bentham	W. D. Ross

Table 2: *Four classical moral-philosophical positions (Mill's position is debated).*

In Chapter 5, the distinction between knowing-that and knowing-how was presented from an epistemological point of view. We will now introduce the distinction in the present ethical context. To begin with, we will show its relevance for deontology and consequentialism. We will, when making our comments, move clockwise from Kant to Mill in Table 2.

Kant's categorical imperative and his four commandments lay claims to be instances of knowing-that of basic norms. But having such knowledge, i.e., *knowing-that about how to act* in contradistinction to knowing-that about states of affairs, does not assure that one is able to perform the actions in question. Sometimes there is no problem. Anyone who understands the imperative 'Do not make false promises!' will probably also thereby know how to act on it. But sometimes there can be quite a gap between having knowing-that of a rule and having knowing-how of the application of the same rule. It is easy to understand the rules 'Develop your talents!' and 'Help others!', but in many situations it is hard to know exactly how to act in order to conform to them. In the latter cases, trial-and-error, imitating, and some advice from others might be necessary before one is able to master the rules. Just hearing them, understanding them, and accepting them theoretically is not enough. The tacit dimension of knowledge is as important here as it is in craftsmanship. Its basis is the fact that perception, both in the sense of conscious perception and the

sense of informational uptake, mostly contains more than what a verbal description of the perceived situation contains. And, as we said in Chapter 5, there are four general ways in which know-how can be improved: (1) practicing on one's own, (2) imitation, (3) practicing with a tutor, and (4) creative proficiency.

A famous case of creative proficiency in the moral realm is the story of King Solomon. Two women claimed to be the mother of the same child, and Solomon had to decide which of them should be counted as being the real mother. He threatened to cut the child in two equal parts, and give the women one part each. One of the women then asked him to give the child to the other woman, whereupon Salomon instead gave it to her. He had suddenly realized that the real mother would prefer to save the life of her child more than anything else.

A similar creativity was showed by a surgeon who was refused to operate a young Jehovah Witness; the latter was suffering from an acute spleen rupture and internal bleeding, and needed blood transfusion, which this religion forbids. According to the rules, the surgeon would have to respect the young man's conscious decision not to be operated. Eventually, the patient became increasingly dizzy, and just as he was fading into unconsciousness, the surgeon whispered in his ear: 'Do you feel the wing-beats of death?' The young man's eyes opened wide in fright, and he requested the operation. Here, we might understand the surgeon's intervention as being based on rather paternalistic considerations, and not paying respect to the patient's autonomy. However, if the surgeon was in doubt whether the patient's stated view could be regarded as authentic, he might be seen as examining the patient's true emotional reaction. Even though we might question whether the emotional reaction should override the patient's ordinary views, the case illustrates ethical creativity proficiency.

The young man's eyes opened wide in fright, and he requested the operation. Although the surgeons intervention was based on rather paternalistic considerations – not paying respect to the patient's autonomy – the case illustrates surgical creativity proficiency.

Our remark to Kant's two imperfect duties applies with even more force to Ross' actual duties. Since he has put forward no knowing-that about how to combine different prima facie duties into one actual duty, it is only

by means of know-how that such a feat can be accomplished. And what is true of Ross' act duty ethics is equally true of the utility principle of act utilitarianism. Even though utilitarians can outline a mathematical representation (knowing-that) of what it means to calculate utility, only a god-like being, i.e., a person who is omniscient and has an almost infinite calculating capacity, could ever make a theoretical utility calculation that ends in a simple order: 'Now and here, you should do A!'. Act utilitarians have to make very crude estimations of consequences and utility values, and how to learn to make these in various situations must to a large extent be a matter of learning and improving a utilitarian kind of know-how.

What, then, to say about rule utilitarianism? Does it need know-how in some respect? Answer: yes, and for three reasons. It has one problem in common with act utilitarianism, one with act deontology (Ross), and one with rule deontology (Kant). First, rule utilitarians can no more than act utilitarians derive their duties by mere knowing-that; only people with a rather long experience can do the adequate estimations. Second, as soon as there are conflicts between rules, rule utilitarianism will encounter the same superposition problem as Ross. Third, some of the rules rule utilitarianism normally put forward have the same indeterminate character as Kant's imperfect duties have. In sum, rule utilitarianism is as much in need of knowing-how as the other moral positions mentioned are.

At least since the mid-1980s, many moral philosophers have stressed the need to make a very fine-tuned apprehension of the complexity of a situation before a moral judgment is passed. We see this as an implicit way of stressing the need for knowing-how even in matters of morals. One prominent such philosopher is Martha Nussbaum (b. 1947), who, among other things, claims that moral know-how can be gained by studying ancient literature. However, we will focus on a position whose most outspoken proponent is Jonathan Dancy (b. 1946). He defends what he calls '*moral particularism*' or 'ethics without principles'. According to him, situation determined knowledge can knock down every possible pre-given substantial moral principle such as the utility principle, Kant's commandments, and Ross' prima facie duties. This position, in turn, means that in principle moral knowing-how can always override moral knowing-that.

(Moral particularism in the sense mentioned might be seen as a *general* philosophical defense of what is sometimes called ‘case-based reasoning’ and ‘casuistic ethics’. It is not, however, identical with the Christian ethical theory that is called ‘situation ethics’; the latter allows for some principles.)

We will present the central thesis of moral particularism by means of an analogy with language. According to particularism, moral thinking is basically as little a matter of application of pre-given moral principles to singular cases as language understanding is basically a matter of applying words from a pre-given dictionary and rules from a pre-given grammar. Language and morals existed long before grammarians and law-makers entered the historical scene. Of course, language acts conform to some kind of pattern; therefore, it is always possible ex post facto to abstract word meanings and grammar, which may then be used when teaching a language. But dictionaries and grammar do not determine exactly or forever what sentence to use in a specific situation. Persons who speak a language fluently, and are able to find the proper words even in unusual and extraordinary situations, are not persons of principle; and neither is the morally good person. As dictionaries and grammar are at best crutches for the native speaker, moral principles are at best crutches for the morally sensitive person.

This does not mean that particularists are of the opinion that there are no moral reasons; their claim is that all reasons are context dependent. As a word can mean one thing in one sentence and quite another in another sentence, what constitutes in one case a reason for a certain action, can, it is claimed, in another case be no reason at all; or even be a reason for the opposite action. As the word ‘blade’ means one thing in ‘He sharpened the blade of his sword’ and another in ‘The blade of the plant had a peculiar green hue’, the fact that it knocks on your door is often a reason for you to open the door, but if you have decided to hide at home it is no reason at all; and if it is the big bad wolf that knocks, then the knocking is a reason to lock the door.

The opponents of particularism (i.e., *the generalists* of deontology and consequentialism) might want to object that the mistake is to think that ‘a mere knock’ is ever a reason to open; it is only the more specific

- ‘a knock by a friendly person’,
- or ‘a knock by a friendly person, when you don’t want to be alone’,

that is such a reason. To this the particularists retort: ‘But what if the big bad wolf stands just behind the friend who knocks when I have no special need to be alone?’ The generalists can then try an even more specific description such as:

- ‘a knock by a friendly person, when you don’t want to be alone, and you will not feel threatened when opening the door’.

This does not leave the particularists without an answer. Now they can say: ‘But what if the door has no handle since this was taken away when the door was painted ten minutes ago?’ Let us stop here. The general claim of the particularists is that they can falsify every proposal that a certain fact is – independent of the situation – a reason for a certain kind of action. In other words, particularists claim that there is no pre-given end to specification regresses of the kind exemplified above. When one speaks of reasons, one speaks in effect only of ‘default reasons’. (This exchange of ‘reasons’ for ‘default reasons’ is, by the way, quite parallel to the exchange of ‘causes’ for ‘component causes’ that we propose in Chapter 6.2.)

The particularists claim about specification regresses is, if correct, as devastating for an act duty ethics of Ross’ kind as it is for utilitarianism and classical deontology, since it implies that reasons cannot be added the way forces are added in classical mechanics. Therefore, there are no prima facie duties, i.e., there are no principles that surely *in each and every situation point in the same direction*. To take an example, according to moral particularism, there is neither an actual duty nor a prima facie duty that tells you: ‘Do not give make promises!’ Now, both Ross and Dancy accept that we might encounter situations in which it is morally right to make a false promise. The difference is this: Ross thinks that even in such situations there is a prima facie duty not to make false promises (which is overridden), but Dancy thinks there is not. When the situation is looked at in its very specificity, there is nothing pointing in the direction of not making a false promise; there is not even a rule saying ‘Do not give false promises, except when …!’; to the contrary, there is something in the situation that directly tells you to make a false promise.

Moral particularists believe, like the generalists, that a morally good person must be sensitive to moral reasons. What is at stake is the nature of moral reasons. The particularists thesis is that no verbally explicit moral rule can in its relative abstractness capture all the specificity that is characteristic of real life situations. Therefore, it is impossible to let linguistically formulated moral rules be the absolute end points of moral justifications. This does not imply that a person who wants to act morally is expected to simply gaze vacantly at the situation before him. He should rather look with an experienced eye, i.e., he should meet the situation with some know-how.

A last remark, the particularists' view that no linguistically formulated rule can capture everything that is of relevance for moral judgments is in no conflict with the following completely formal principle for moral consistency:

- all situations that are in morally relevant respects exactly alike should be judged in exactly the same way.

*****

The problem of how to apply the knowing-thats of deontological and consequentialist ethics is not the only problem that this kind of knowing-that has. Apart from *the application problem*, there is *the justification problem*, i.e., the problem of how to justify that the presumed basic norms really can count as valid; a problem that no substantive ethics can side-step. In relation to deontological ethics, we have already noted that it is hard to accept that God's commandments or Kant's categorical imperative justifies themselves. Now we will make some remarks on the same problem in relation to utilitarianism and the utility principle. Here comes a 'mnemonic doggerel' from Bentham:

> *Intense*, *long*, *certain*, *speedy*, *fruitful*, *pure*—
> Such marks in *pleasures* and in *pains* endure.
> Such pleasures seek if *private* be thy end:
> If it be *public*, wide let them *extend*

> Such *pains* avoid, whichever be thy view:
> If pains *must* come, let them *extend* to few.

The first three lines state a rather uncontroversial thesis. If your actions do not affect anyone at all apart from yourself, then feel free to seek your private pleasure and try to avoid pain. Only hardheaded ascetics who think there is something *intrinsically* wrong with pleasure can object; we will, just as Bentham, leave them out of account. The next three lines, however, are in much more need of justification. What makes Bentham able to move so swiftly from the first part of the verse, where only one's own pleasure seeking is at stake, to the second part, where one is also encouraged to help others to have pleasure? To the person who experiences them, pleasures come stamped as being in some sense positive, and pains as being negative. Therefore, if nothing else intervenes (e.g., deontological norms), the rule that I should seek pleasure and avoid pain for myself justifies itself. But why should a person care for the pleasures and the pains that he himself does not experience? In situations of compassion and empathy, one does in some sense experience even the sufferings of other people, but often the pleasures and pains of others are apprehended in a rather neutral way. Why care about them when this is the case? The utility principle puts one's own and all others' pleasures and pains on a par, but this is not the way they usually appear to us. A good argument in favor of the equalization is needed, but, a bit astonishingly, Bentham has none; and, even more astonishingly, neither has Mill, who in an oft-criticized sentence says:

> The sole evidence it is possible to produce that anything is desirable [= worthy of desire], is that people do actually desire it. If the end which the utilitarian doctrine proposes to itself were not, in theory and in practice, acknowledged to be an end, nothing could ever convince any person that it was so (*Utilitarianism*, Ch. 4).

But "the end which the utilitarian doctrine proposes" is to maximize the total amount of pleasure, and this is definitely not what most people "actually desire." Mill's move from each man's maximizing of his own

pleasure (egoistic hedonism) to each man's conformance to the utility principle (universalistic hedonism) is as logically unfounded as Kant's move from 'being capable of creating ends for oneself' to 'being an end in itself'. Mill can be accused of making two fallacies. First, he writes as if being 'subjectively desired' implies being 'objectively desirable' in the sense of worthy of desire. Second, he writes as if the fact that each man's happiness is objectively desirable should entail that it is always, even in cases of conflict with one's own happiness, objectively desirable to maximize the total amount of happiness. But what logically follows is only a trivial thing, namely that there is a kind of happiness of mankind that *can be defined* as the aggregated happiness of each person.

Justification is an enterprise not only in ethics. Let us illustrate the justificatory problem of utilitarianism by a detour back to Chapter 3.4 and the logical positivists' principle of verification. They claimed that only verifications (i.e., positive empirical observations) could justify one in regarding a non-formal scientific assertion as being true. But how then is the verification principle itself to be justified? Since it is meant to be both non-formal and justified, *it should be applicable to itself*. But this seems ridiculous. It would mean that we should regard the verification principle as true if and only if we can verify that it is true; but then we have nonetheless presupposed it, and there is no use in testing it. Similarly, if the utility principle is regarded as justified, it should be applicable to itself. But this seems ridiculous, too. It would mean that we should regard the utility principle as a basic norm if and only if we can show that following it would lead to a maximum of pleasure; but then we have nonetheless presupposed it, and there is no use in verifying it.

Within positivism, the verification principle is given no real justification, and neither is, normally, the utility principle within utilitarianism. The famous classical utilitarian Henry Sidgwick, however, has proposed a justificatory reform of utilitarianism. He claims that it is as possible in morals as in mathematics to have true intuitive non-empirical insights, and that it is by means of such an insight that the utility principle becomes justified. But to resort to intuitions has not ranked high in mainstream twentieth century philosophy, and Sidgwick did not have many followers. The justificatory problem of ethics is one reason why, as we mentioned at the beginning of this chapter, many twentieth century moral philosophers

have chosen to work only with presumed morally neutral meta-ethical problems. A new approach to the justificatory problems in moral and political philosophy was inaugurated by John Rawls (1921-2002) in his book *A Theory of Justice* (1971); here the notion of 'reflective equilibrium' is central.

Rawls does not use the term 'fallibilism', but his move does implicitly introduce fallibilism in ethical matters. He asks us to stop looking for self-justificatory moral principles and/or self-evident particular moral judgments. Instead, he claims, all we can reasonably strive for is that our considered principles and considered particular judgments cohere with each other. That is, they ought to balance each other or to be in *reflective equilibrium*. If they are not, then we have to change something. Let us take a simple example. Assume that someone who believed and defended for a long time the norm that it is absolutely forbidden to actively hasten a patient's death, ends up in a situation where he, after reflection, comes to the conclusion that he is – in this very special situation – allowed to do so in relation to an unbearably suffering patient's death. The patient intensely begs the physician to help him to die quickly, since all palliative treatments have failed. Then there is *dis*equilibrium between the physician's moral principle and his particular judgment, and if he is a reflective person he ought to reject or revise his principle, his particular judgment, or both in order to restore equilibrium.

Initially, in situations like the one above, we do not know whether to change a principle (or several), a particular judgment (or several), or perhaps both, which means that many conjectures may have to be tested before one can rest content and say that a new equilibrium has been found. Fallibilism enters the scene because no such equilibrium can be regarded as certainly stable. One day there may arise a completely un-thought of kind of situation, which makes us become very certain that, here, a particular moral judgment of ours have to contradict one of our earlier accepted principles. There is then disequilibrium. Our moral principle, or perhaps even our whole moral paradigm, contains an anomaly in much the same way as a scientific paradigm can contains anomalies caused by observations or measurements (see Chapter 3). Another day, a moral reformer (think, for instance, of Singer) may present a principle that so far we have not considered at all, and which contradicts many of our old

particular judgments. Then there would again be disequilibrium. Sometimes a whole new moral paradigm is proposed, e.g., utilitarianism at the end of the eighteenth century.

The method of reflective equilibrium consists in working back and forth among our considered judgments on both moral principles and particular moral judgments; revising any of these elements whenever necessary in order to achieve an acceptable coherence. Such a coherence means more than that our beliefs are merely logically consistent with each other. Our principles should be relevant for many of our particular judgments, and these judgments should be regarded as a kind of evidence for the principles. On Rawls' view of justification, one reflective equilibrium can be seen as being better than an older and superseded one, but nonetheless it can itself very well in the future be replaced by an even better reflective equilibrium. Even the 'principle-free' view of moral particularism can be evaluated in this way. When particularism is in reflective equilibrium, there is coherence between the view that there are no context independent moral reasons and all the considered particular judgements. Dancy is explicitly fallibilist and writes: "This [particularist] method is not infallible, I know; but then neither was the appeal to principle (Dancy 2005)."

Rawls' views on reflective equilibria are as possible to apply to rules and principles in the medical domain as to general moral principles. All those who work as health care professionals – doctors, nurses, and the rest of the staff – must try to find a reflective equilibrium between the rules of health care ethics and everyday medical practice.

After about 200 years of philosophical discussion of deontological ethics and utilitarian consequentialism, it is easy to summarize what is regarded as their main kind of anomalies, i.e., what kind of considered particular moral judgments each seems to be in conflict with. The Decalogue commandment 'You shall not kill!' is for many people contradicted by their particular judgments when in danger of being murdered or in case where someone begs for euthanasia. Kant's commandment 'Do not commit suicide!' is to many contradicted by what seems morally allowable when one is threatened by lifelong intense pains and sufferings. His commandment 'Do not make false promises!' seems in several situations not to be exactly fitting. Put more generally, what is wrong with

deontology is its inflexibility. And the same goes then of course for rule utilitarianism. Neither allows norms to take exceptions. Sometimes it does not seem morally wrong to break a rule, even though it would cause disaster if most people broke the rule most of the time.

Many find the utility principle of act utilitarianism contradicted by situations where they have to suffer for the happiness of all, even though they have done nothing wrong. Is it right, for instance, to force a person to donate one of his kidneys to another man if this would increase the total amount of pleasure? Examples of this kind can easily be multiplied. Also, when it comes to questions of justice, act utilitarianism is counter-intuitive. All act utilitarianism can say is that one is being treated justly if one is being treated as a means for the happiness of all. Since justice is normally regarded as justice towards (or between) persons, this view contains a complete redefinition of justice. It seems impossible for act utilitarians (but not for rule utilitarians) to bring personhood into their moral system. In short:

- Classical deontological ethics and rule utilitarianism cannot make sense of our considered particular judgments to the effect that, *sometimes*, we have to make exceptions to norms.
- Act utilitarianism cannot make sense of our judgments that, *in many situations*, we have some kind of moral rights as individual persons; rights which no utility principle can overrule.

Both these blocks of anomalies are highly relevant for medical ethics. Most physicians do now and then encounter situations that are quite exceptional from a moral point of view, and, as we will see in Chapter 10, many medical norms put the integrity of individual persons in the center. Normal patients shall be regarded as autonomous agents, and clinical research requires informed consent. But the utility principle does not automatically exclude even the killing of an innocent person. Here comes a common anti-utilitarian thought-experiment.

A patient suffering from injuries and several fractures arrives at an emergency room. With adequate treatment he is curable and able to be fully rehabilitated. This patient, however, happens to have the same tissue type as some other patients who are waiting for heart- and lung-transplants

as well as kidney transplants. On the assumption that it is for sure possible to keep it a secret, it would according to a utility calculation be right to take the organs from the injured patient and give them to the other patients. This would maximize happiness, but (rhetorical question) isn't it nonetheless wrong? It has to be added, however, that most act utilitarians are in fact against such interventions, since they regard it impossible to keep this a secret; accordingly trust in the health care system will erode and the overall negative consequences will become more significant than the positive.

The analogy between fallibilism in science and fallibillism in morals, which we noted in passing, contains yet another feature that it is good to be aware of. Neither in science nor in ethics is it possible to completely foresee the future. On the ruins of an old paradigm someone may one day be able to construct a wholly new and unforeseen version. Therefore, we are not trying to say that it is logically impossible for deontology and utilitarianism to recover from their present anomalies and stage a come back. Furthermore, wholly new moral paradigms might be created. Therefore, we are not claiming that it is logically impossible to create a moral paradigm that is neither a deontology, nor a consequentialism, and nor a virtue ethics.

*****

We would like to end this subchapter with some words on a minor school of thought called 'personalism'. It does not immediately fit into any of the three moral paradigms we are presenting. Mainly because it focuses on questions about *what has value and how to rank different kinds of values,* and does not bother to work out what the norms ought to look like. Like Kant, the personalists stress that persons have an absolute value, but unlike Kant they claim that to be a person means much more than to have a mind that has self-consciousness and can reason. All of them claim that to be a person essentially involves having an emotional life; some personalists also claim that personhood is necessarily social, i.e., that one cannot be a person without to some extent also caring for other persons. The basic personal relation is claimed to be, with an expression from Martin Buber (1878-1965), an 'I-Thou relation'. Such personalism contradicts

individualism, since it claims that true self-love involves love of others. Since many of the famous personalists have been religious thinkers, it should also be said that many personalists think there is an ‘I-Thou relation’ even between human beings and a higher being. Personalism has played quite a role in some discussions within nursing science and nursing ethics.

The most famous philosopher who has thought along personalist lines is probably the German Max Scheler (1874-1928). He claims that there are four distinct classes of values (and disvalues) that can be ranked bottom up as follows: (1) sensory values, i.e., sensory pleasures in a broad sense; (2) vital values, i.e., things such as health, well-being, and courage; (3) spiritual values, e.g., justice, truth, and beauty; (4) holy values such as being sacred. Scheler argues (but other personalists contest it) that in cases of conflict one has always to act so as to realize the higher value.

9.4 Virtue ethics

A virtue is a habit and disposition to act and feel in a morally good or excellent (virtuous) way. Etymologically, it might be noted, the term is quite sexist. It comes from the Latin ‘vir’, which means ‘man’ in the masculine sense, i.e., the virtuous man is a thoroughly manly man. The corresponding ancient Greek word, ‘aretē’, does not, however, have such an association. It means skill or excellence in a non-gendered sense.

In order to see what is specific to virtue ethics, one has to see what deontology and consequentialism have in common. As before, we will use Kantian ethics and utilitarian ethics as our examples. The contrast between these ethical views and virtue ethics contains some intertwined features. Whereas the central principle in both Kantianism and utilitarianism tells what a morally right *action* looks like (Kant’s categorical imperative and the utility principle, respectively), the central principle of virtue ethics tells us what kind of *man* we ought to be. It simply says:

- Be a morally virtuous man!

Virtue ethics was already part of Chinese Confucianism (Confucius 559-479BC), and it was the dominant ethics in Ancient Greece. Its

philosophical father in the Western world is Aristotle. During the nineteenth and the twentieth century, its influence in Western philosophy was low, but there were nonetheless critics of deontology and utilitarianism. Some famous late twentieth century Anglo-American virtue ethicists are Philippa Foot (b. 1920), Bernard Williams (1929-2003), and Alasdair MacIntyre (b. 1929).

The difference stated between virtue ethics and deontology/consequentialism does not mean that only virtue ethics can talk about virtuous persons. In Kantian ethics, a virtuous person can be defined as someone who always wants to act in conformity with Kant's categorical imperative; and in utilitarianism, a virtuous person can be defined as someone who always tries to act in conformity with the utility principle. Conversely, neither does the difference stated mean that virtue ethics cannot at all formulate an act-focusing principle. It becomes, however, a rather empty tautology that does not distinguish virtue ethics from deontology and consequentialism:

- Act in such a way that your action becomes morally right.

This virtue ethical imperative can neither fulfill the testing function that the first formulation of Kant's categorical imperative fulfills, nor can it have the regulative function that the utility principle has. Nonetheless, stating it makes a sometimes misunderstood feature of virtue ethics clear. A virtuous person is someone who out of character or habit chooses the morally right action, but his actions do not become the morally right ones *because* he is a virtuous person. It is the other way round. A person is virtuous because he performs the morally right actions. Compare craftsmanship. A good physician is good because he can cure people; the treatments do not cure because they are prescribed by a good physician. No virtuous person can stamp actions as being virtuous the way a baptizing priest can give a child a certain name. Aristotle writes:

> The agent [the virtuous man] also must be in a certain condition when he makes them [the virtuous actions]; in the first place he must have knowledge, secondly he must choose the acts, and

> choose them *for their own sakes* [italics inserted], and thirdly his action must proceed from a firm and unchangeable character (*Nicomachean Ethics* 1105a32-36).

For reasons of practical expediency, we must often take it for granted that virtuous-regarded persons have acted in the morally right way, but this does not mean that such persons act in the right way by definition. Another way to put this point is to say that for a virtuous person the primary target is to perform virtuous acts, but in order to become able to hit the target continuously, he should try to become a virtuous person. The fact that experiments in social psychology have shown that character traits are more situation dependent as was once widely thought, does not mean that they are completely insignificant. Rather, it means that virtuous persons should try to take even this situation dependency into consideration.

The misunderstanding that a virtuous person can *define*, not only find out and perform what is to be reckoned as the morally right action, may have many sources. Here is one possible: it is noted that a virtuous person often in everyday life functions a bit like a judge in a law court, but it is forgotten that juridical decisions can always be questioned. Let us expand. Court judges are meant to apply pre-written laws, but since the written law is often not detailed enough to take care of the cases under scrutiny, the judges (exercising their knowing-how) have often so to speak to define where the boundary between legal and illegal actions should be drawn. But his is not the absolute end of the procedure. If the court in question is not the Supreme Court, then the verdict can be appealed; and even if it should be the highest instance, there is nothing logically odd in thinking that the verdict is wrong.

In the last subchapter we described how know-how enters Kantianism and utilitarianism. In virtue ethics know-how becomes much more prominent because of the vacuity of its basic knowing-that, the norm: Act in such a way that your action becomes the morally right action! Kant’s categorical imperative, his four commandments, and all forms of the utility principle convey some substantial knowing-that, but the norm now stated does not.

We have earlier described the difference between hypothetical and categorical imperatives or rules. Know-how can exist in relation to both,

but now we are talking only about know-how in relation to categorical imperatives. In Chapter 5, on the other hand, we were concerned only with know-how in relation to goals that one is free to seek or not to seek. This distinction between knowing-how in relation to hypothetical and categorical imperatives can be found already in Aristotle. He calls the former kind of know-how 'techne' and the latter one 'phronesis'. In modern literature, they are often called 'technical skill' and 'practical wisdom', respectively. Technical skill is know-how in relation only to the choice of means, whereas practical wisdom is know-how in relation to questions of what one has categorically to do. (Theoretical wisdom, 'sophia', is simply having infallible knowing-that, 'epistéme'.)

If moral particularism (see the last subchapter) is right, then a further distinction is needed. There are two kinds of phronesis (moral knowing-how); we will call them '*principle-complementary*' and '*principle-contesting*' phronesis. The former kind is needed when a pre-given, but vague, moral principle (moral knowing-that) shall be applied; it is needed both in order to know if the principle is at all applicable, and in order to know how it should be applied. Such know-how can by definition never contest the principle, and this is the only kind of phronesis that deontological and consequentialist ethical systems require. The other kind of phronesis is the one required by particularism, i.e., a moral know-how that can contest every moral knowing-that. Virtue ethics contains both kinds of phronesis. A practically wise man can both apply rules in extraordinary situations and take account of particular situations where all existing moral principles have to be overridden. In this undertaking, it should be noted, consequences may sometimes have to be considered and sometimes not. Moral principles are neither necessary nor sufficient for becoming a virtuous person, but experience is necessary. In an oft-quoted passage Aristotle says:

> while young men become geometricians and mathematicians and wise in matters like these, it is thought that a young man of practical wisdom cannot be found. The cause is that such wisdom is concerned not only with universals but with particulars, which become familiar from experience, but a young man has no

> experience, for it is length of time that gives experience (*Nicomachean Ethics* 1142 a).

The fact that virtue ethics contains the principle-contesting kind of phronesis, does not imply that virtue ethicists find all substantive moral rules superfluous. It means only that they regard all such rules as rules of thumb, as default rules, or as rules for normal circumstances; as such, by the way, the rules can be in need of principle-complementary phronesis. As computers need default positions in order to function in general, human beings seem to need default norms in order to function in society. The question whether moral rules are useful in education and in practical moral thinking is distinct from the question whether moral justification ends with verbally stated substantial universal principles. It is only principles of the latter kind that the particularists' and the virtue ethicists oppose.

We have remarked that as soon as act utilitarians and act duty ethicists enter politics or institutional acting, even they have to accept discussions of rules and make room for some kind of rule utilitarianism and rule duty ethics, respectively. Exactly the same remark applies to virtue ethics. Virtue ethics cannot rest content with merely stressing the particularist aspect of individual acts; it has to say something about rule creation, too. As a person may ask what action he has categorically to do, a law-maker may ask what rules he has categorically to turn into laws. There are two kinds of phronesis: *act phronesis* and *rule phronesis*, respectively. At least modern societies require phronesis two times, first when laws and rules are made, and then when the laws and the rules are applied. The societies of Confucius and Aristotle did not contain as strict a division of labor between law-makers and law-appliers as our modern societies do. Rule phronesis is a virtue of politicians and legislators, and act phronesis is a virtue of judges, policemen, bureaucrats, and rule followers of all kinds.

For Aristotle, the rules followed in personal acting are not rules directly showing how to act, but rules for what kind of personal character traits one ought to develop in order to be a virtuous man. The character traits in question are connected both to certain ways of feeling and to certain ways of acting. Often, a virtuous character trait is contrasted with two contrary opposite and vicious character traits; one of these bad character traits is seen as expressing itself in an excess and the other in a deficiency of a

certain kind of action or feeling. The virtuous man is able to find the golden mean between two bad extremes. The classical example is the virtuous soldier. He is brave, which means that he has found a golden mean between cowardice (too much fear) and foolhardiness (too little fear). But what it means to be brave in a certain situation cannot be told beforehand, i.e., no one can be brave without know-how. Here comes an Aristotelian list, even though it lays no claims to be completely true to Aristotle's (sometimes hard-translated) examples in *Nicomachean Ethics*; our list is merely meant to convey Aristotle's general approach in a pedagogic way.

Vice (defect)	**Virtue (mean)**	**Vice (excess)**
Foolhardiness (too little fear)	Courage	Cowardice (too much fear)
Insensibility (caring too little about pleasure)	Temperance	Self-indulgence (caring too much about pleasure)
Stinginess (giving too little)	Generosity	Prodigality (giving too much)
Shamelessness (too little shame)	Modesty	Bashfulness (too much shame)
Humility (too little integrity)	Pride	Vanity (too much integrity)
Apathy (too little emotion)	Good Temper	Irascibility (too much emotion)
Surliness (being too negative towards others)	Friendliness	Flattery (being too positive towards others)

These rules of conduct ('Be courageous!', etc.) have to be regarded as rules of conduct for normal situations in one's own community. One should not urge enemies to be courageous, and one should not be generous and friendly towards enemies. And even in one's own community there may be situations where it is adequate to lose one's good temper and become angry. Aristotle is quite explicit on this.

No virtuous mean does in itself take any degrees, only the deviations do. The virtuous mean is in this sense like the zero point on the scale for

electric charges. There are degrees of both negative and positive electric charge, but there are no degrees of not being electrically charged.

Hippocrates thought in the same way about doctors. They ought to have the general virtues and then some for their profession specific virtues; one might say that Hippocrates put forward an applied virtue ethics. Especially, he warned against unbridled behavior, vulgarity, extortion, and shamelessness, as well as being insatiable. He even put forward rules for how doctors ought to dress. According to the Hippocratic recommendations, doctors should be moderate when it comes to financial matters, and they should neither be greedy nor extravagant. Moderation is the golden mean between these two extremes. There is quite a structural similarity between his balance thinking about diseases and his balance thinking about morally good behavior. Here comes a Hippocratic list of vicious and virtuous ways of being.

Extreme	**Golden Mean**	**Extreme**
Ignorant	Modest	Pretentious
Tempting	Friendly	Coquettish
Weak	Robust	Domineering
Inferior	Humble	Pompous
Nonchalant	Devoted	Fanatic
Slow-witted	Self-command	Impulsive
Soiled (corrupt)	Decent	Artful (too cunning)
Ignorant	Careful	Finical (too keen on details)
Cynical	Empathic	Hypersensitive

The relationship between moral rules and moral know-how in virtue ethics can be further clarified by reflections on how moral knowledge can develop. If there is a deontological norm such as ‘Do not make false promises!’, then there is, as Kant explicitly noted, three *logically different* ways in which one can conform to it. One can:

- act *in accordance with it*, i.e., act without caring about the rule, but nonetheless happen to act in such a way that one conforms to it
- act *on it*, i.e., know the rule and consciously conform to it, but doing this for reasons that are morally indifferent

- act *for it*, i.e., know the rule and consciously conform to it because one thinks that this is the morally right thing to do.

According to Kant, it is only persons that *act for* the categorical imperative that act morally, and can be regarded as being good (virtuous) persons. In Aristotle one meets Kant's tripartite logical distinctions as a distinction between three necessarily consecutive stages in the moral development of human beings. Small children can only be taught to act *in accordance with* good moral rules by means of rewards and punishments. Today, being more aware of the imitating desires and capacities of children, we might add that even small children can learn to act in accordance with moral standards by imitating good role models. They need not be educated the way dogs are trained.

When children have become capable of understanding rules and of consciously acting *on* such, they enter the second stage. They should then, says Aristotle, be taught to heed morally good rules 'as they heed their father'. Still, they lack moral insight, but they can nonetheless regard it as natural and necessary to follow the rules in question; normally, they even develop some technical skill in doing so.

When entering the third stage, two things happen. First, the technical skill becomes perfected, which means that the rules are left behind in the way stressed by particularists. Second, the skill becomes transformed into phronesis. That is, persons now have an insight that they categorically ought to do what they are doing because this is morally good. The first (rule-trespassing) change makes Aristotle differ from Kant, but the second (insight-creating) change parallels Kant's move from 'acting on' rules to 'acting for' moral rules.

Aristotle's developmental thinking has counterparts in contemporary philosophy and psychology. In Chapter 5.4, we presented the five stages of skill acquisition that the Dreyfus brothers have distinguished: the novice stage, the advanced beginner stage, the competence stage, the proficiency stage, and the expertise stage. They think that these distinctions apply to moral skill and moral maturity too. Since Aristotle's stages allow for having grey zones in-between them as well as for having sub-stages, one can try to relate the Dreyfus' stages to Aristotle's. We would then like to make the novice and the advanced beginner stages sub-stages of Aristotle's

second stage, for since the novices are instructed by means of rules, Aristotle's first stage is already left behind. The competence stage seems to be an intermediary between Aristotle's second and third stage, whereas the proficiency and the expertise stages seem to be sub-stages of the third stage.

At the beginning of the 1970s, a developmental psychologist, Lawrence Kohlberg (1927-1987), claimed to have found empirically a law of moral development. It gave rise to much discussion. Like Aristotle, Kohlberg found three main levels, but unlike Aristotle he divides each of them into two stages, and the last level is in one respect not at all Aristotelian. Kohlberg's main views are summarized in Table 3 below. Level 1 is called 'the pre-conventional level', level 2 'the conventional level', and level 3 'the post-conventional level'. At the pre-conventional level we (human beings) have no conception of either informal moral rules or formal rule-followings. At the conventional level we have, but we simply take the rules for granted. Finally, at the post-conventional level, we have arrived at the insight that there are morals, and at the last sub-level we have even realized that morals can be discussed.

Stages of moral consciousness:	Main idea of the good and just life:	Main kinds of sanctions:
(1a) Punishment-obedience orientation	Maximization of pleasure through obedience	Punishment (as deprivation of physical rewards)
(1b) Instrumental hedonist orientation	Maximization of pleasure through tit-for-tat behavior	Punishment
(2a) Good-boy and nice-girl orientation	Concrete morality of gratifying interaction	Shame (withdrawal of love and social recognition)
(2b) Law-and-order orientation	Concrete morality of a customary system of norms	Shame
(3a) Social contract orientation	Civil liberty and public welfare	Guilt (reaction of conscience)
(3b) Universal ethical principles orientation	Moral freedom	Guilt

Table 3: *Kohlberg's three levels and six stages of moral development.*

In relation to this table, Kohlberg says something like the following. Children from 0-9 years live on the pre-conventional level where, first (1a), they only strive for immediate hedonistic satisfaction, but later (1b) learn that by means of tit-for-tat behavior one can increase one's amount of pleasure. They will act in accordance with morals either by chance or because adults reward them for moral behavior and punish them for immoral behavior; punishment is here taken in such a broad sense that a mere negative attitude counts as punishment.

Children and teenagers in the age of 9-20 are normally living on the conventional level. They do now perceive that there is something called 'being good' and 'being bad', respectively. At first (2a), they only apprehend it in an informal way when they concretely interact with other people, but later (2b) they can connect 'being good' and 'being bad' to the conformance of impersonal rules. Something in human nature makes people ashamed when they regard themselves as having done something bad. The living on this level is deontological in the sense that role conformance and rule following are regarded as being important quite independently of their consequences.

Some people may forever stay on the conventional level, but many adults proceed to the post-conventional level. First (3a) they realize that at bottom of the informal and formal rules that they have earlier conformed to, there is an implicit or explicit social contract between people. Then, perhaps, they also come to the conclusion (3b) that good social contracts ought to be based on universal ethical principles. Here, Kohlberg is much more a Kantian and/or utilitarian than an Aristotelian. On level 3, people react with a feeling of guilt if they think that they have done something that is seriously morally wrong.

(It has been argued that 'ontogeny recapitulates phylogeny', i.e. that the biological development of an organism mirrors the evolutionary development of the species. Similarly, Kohlberg tends towards the view that the moral development of the individual recapitulates the historical moral development of societies. He says: "My finding that our two highest stages are absent in preliterate or semiliterate village culture, and other evidence, also suggests a mild doctrine of social evolutionism (Kohlberg 1981 p. 128).")

Leaving societal moral development aside, what to say about Kohlberg's schema? Well, our aim here is mainly to show that questions about moral development emerge naturally, and that there are reasons to think that discussions of them will continue for quite a time. We will only add three brief remarks and then make a historical point.

First, does the schema contain as many levels and stages that it ought to contain? From an empirical point of view, Kohlberg later came to doubt that one should speak of stage (3b) as a *general* stage of development; not many people reach it. From a moral-philosophical point of view, however, he has speculated about the need for a seventh level, and from such a perspective Habermas has proposed within level 3 a more dialogical stage (3c), which fits discourse ethics better as an end point than (3b) does. One might also ask whether all kinds of emotional sanctions have been taken into account. Is there only shame and guilt? Some philosophers have argued that even remorse is a kind of 'morally punishing' emotion – and a better one.

Second, what about the last stage? It seems to make no difference between deontology and consequentialism; it only focuses on what these have in common, namely a stress on the existence of universal ethical principles. However, this stage differs from both old-fashioned deontology and consequentialism in bringing in fallibilism; on stage (3b) people are prepared to discuss their own moral norms.

Third, what about the place afforded to virtue ethics? It seems to have no place on any stage above (2a). The stress on know-how typical of virtue ethics matches Kohlberg's stress on the informal aspect of the moral interaction that characterizes this stage. But then there is only explicit rule-following and/or discussions of such rules. The Dreyfus brothers have argued that the kind of phronesis that virtue ethics regards as central has to be made part of the description of the last stage.

Kohlberg's first empirical investigations gave rise to a very intense debate about gender and moral development. According to his first results, on average, girls score lower on moral development than boys do. One of Kohlberg's students, Carol Gilligan (b. 1936), then argued that this was due to the fact that Kohlberg, unthinkingly, favored a principle-seeking way of reasoning to the detriment of a relation-seeking way. Both are necessary, both can be more or less developed, and none of them can be

given a context independent moral priority over the other. The first way is the primary one in relation to questions of justice, and the second way in relation to moral problems in caring. According to Gilligan, given traditional sex roles, faced with moral problems boys focus on principles and justice, whereas girls focus on face-to-face relations and caring. If both kinds of moral issues are given their moral-developmental due, then, Gilligan argues, there is no longer any sex difference with respect to moral development. Her writings became one seminal source of what is now known as 'the ethics of care'. In this school of thought one stresses (and investigates) the moral importance of responding to other persons as particular individuals with characteristic features.

In our exposition of virtue ethics, we have so far made a distinction between Aristotle and modern virtue ethicist thinking in only one respect: the latter needs a notion of 'rule phronesis'. But we think there are three more respects in which modern virtue ethics has to differ from the traditional one. First, fallibilism has to be integrated. Aristotle was an infallibilist, but as we have seen even expert know-how can miss its target (Chapter 5.4-5). Even if the expert in front of a contesting non-expert sometimes has to say 'I cannot in detail tell you why, but this is simply the best way to act!', it may turn out that the non-expert was more right than the expert.

Second, according to thinkers such as Socrates, Plato, and (but to a lesser extent) Aristotle, there cannot arise any real conflicts between acting morally right and acting in the light of the happiness of one's true self. Such apparent conflicts arise, they claim, only when a person lacks knowledge about what is truly good for him. Socrates thought it was best for his true self to drink the hemlock. But this view is hard to defend. The dream of what might be called an 'unconflicted psychology' has to be given up even among virtue ethicists. When in their original culture, an elderly Eskimo was asked by his son to build an igloo in which he could 'travel on his own', he was expected to be aware that he had become a burden to his tribe, and follow its duties and sacrifice himself. According to classical virtue ethics, the old Eskimo might be said to act also in his own individual self-interest, but there is no good reason to cling to this view. Virtue ethicists have to accept that conflicts can arise between what their moral know-how tells them to do, and what their true self-interest

wants them to do. The existence of such conflicts is for duty ethicists and utilitarians such a trivial and elementary fact that for many of them its denial immediately disqualifies classical virtue ethics for consideration. Even though a deadly sick elderly utilitarian patient, who wants to live longer, may for reasons of maximizing happiness come to the conclusion that younger patients with the same disease should have his place in the operation queue, he would never dream of saying that his choice furthers his own self-interest.

Third, we have to repeat that experiments in modern social psychology have shown that character traits are not as situation independent as classical virtue ethicists assumed.

In what follows, we will take it for granted that modern virtue ethics, just as duty ethics and utilitarian ethics, accepts the existence of conflicts between acting morally right and acting in the light of one's true self. This means that all three have a motivational problem:

- what can in a situation of conflict make a man act morally instead of only trying to satisfy his self-interest?

We will not present and discuss any proposed solutions to this problem, only put virtue ethics on a par with deontology and consequentialism. The latter might argue that morality has the sort of authority over us that only a rule can provide, but virtue ethicists can then insist that situations and individuals can come in such a resonance with each other that, so to speak, a moral demand arises from the world. The situation simply demands that self-interest has to surrender.

Our little plea for a modern fallibilist virtue ethics has structural similarities with our earlier strong plea for a fallibilist epistemology. Before we bring it out, let us here a second time (cf. p. 00) quote Thomas Nagel about the inevitability of trying to find truths and trying to act right:

> Once we enter the world for our temporary stay in it, there is no alternative but to try to decide what to believe and how to live, and the only way to do that is by trying to decide what is the case and what is right. Even if we distance ourselves from some of our thoughts and impulses, and regard them from the outside, the

> process of trying to place ourselves in the world leads eventually to thoughts that we cannot think of as merely "ours." If we think at all, we must think of ourselves, individually and collectively, as submitting to the order of reasons rather than creating it (Nagel 1997, p. 143).

When we focused on truth-seeking, we linked this quotation to Peirce's view that we should "trust rather to the multitude and variety of [the] arguments than to the conclusiveness of any one[; our] reasoning should not form a chain which is no stronger than its weakest link, but a cable whose fibers may be ever so slender, provided they are sufficiently numerous and intimately connected." Having now focused on right-act-seeking and introduced the notion of phronesis, we can say that Peirce's view is a way of stressing the importance of phronesis in epistemology. The methodological rules taught in a science should be looked upon as default rules.

9.5 Abortion in the light of different ethical systems

We have presented the three main ethical paradigms and some of their sub-paradigms, and we have shown by means of the notion of 'reflective equilibrium' how discussions about what paradigm to accept can make cognitive sense. Next we will present what one and the same problem may look like from within the various paradigms. Our example will be one of the big issues of the twentieth century, abortion. We leave it to the reader to find structural similarities between this and other cases. Some aspects of the problem of abortion will probably, because of the rapid development of medical technology, soon also appear in other areas.

The problem we shall discuss is how to find a definite *rule* that speaks for or against abortion in general or at some date. In relation to a single case of a woman who wants a forbidden abortion, a prima facie duty ethicist may then nonetheless come to the conclusion that the rule is overridden by other prima facie duties, an act utilitarian that it is overridden by some act utilitarian considerations, and a virtue ethicist that his practical wisdom requires that he makes an exception to the rule. But this is beside the rule discussion below.

In all deontological ethical systems put forward so far, the rule for abortion (pro or against) is derived from some more basic norm. Let us start with a religious duty ethicist who believes in the sanctity of human life and regards the norm 'You shall not kill human beings!' as being absolute, and one from which he claims to be able to derive the rule that abortions are prohibited. From a pure knowing-that point of view he has no problem, but from an application point of view he has. The application cannot be allowed to be a matter of convention, because deontological norms should be *found*, not created. That is, the norm presupposes that there is in nature a discontinuity between being human and not being human. By definition, where in nature there are only continuities, every discontinuity must be man-made and in this sense conventional. To take an example: the line between orange and yellow colors is conventional, and such a kind of line cannot be the base of a deontological norm. So, what does the first part of the developmental spectrum for human beings look like? Where in the development '(egg + sperm) → zygote → embryo → fetus → child' is there a discontinuity to be found between life and non-life? Here is a modern summary of prototypical stages and possible discontinuities (Smith and Brogaard 2003):

a. the stage of the fertilized egg, i.e., the single-cell zygote, which contains the DNA from both the parents (day 0)
b. the stage of the multi-cell zygote (days 0-3)
c. the stage of the morula; each of the cells still has the potential to become a human being (day 3)
d. the stage of the early blastula; inner cells (from which the embryo and some extraembryonic tissue will come) are distinguished from outer cells (from which the placenta will come) (day 4)
e. implantation (nidation); the blastula attaches to the wall of the uterus, and the connection between the mother and the embryo begins to form (days 6-13)
f. gastrulation; the embryo becomes distinct from the extraembryonic tissue, which means that from now on twinning is impossible (days 14-16)
g. onset of neurulation; neural tissue is created (from day 16)
h. formation of the brain stem (days 40-43)

i. end of the first trimester (day 98)
j. viability; can survive outside the uterus (around day 130 [should be 147])
k. sentience; capacity for sensation and feeling (around day 140)
l. quickening; the first kicks of the fetus (around day 150)
m. birth (day 266)
n. the development of self-consciousness

First some general comments on the list. Discontinuity i (the end of the first trimester) is obviously conventional; and discontinuity j (viability) is conventional in the sense that it depends on medical technology. Several medieval theologians discussed an event not mentioned in the list above, the date for 'ensoulment' of the fetus, i.e., the day (assumed to be different for boys and girls) at which the soul entered the body. Some ancient philosophers, including Plato, thought that it was not until the child had proven capable of normal surviving that it could be regarded as a human being. This view might have influenced the Christian tradition of not baptizing a child until it is six months old. Children who managed to survive the first six months of their lives were at that time assumed to have good chances of becoming adults. Aristotle took quickening to be the decisive thing; and so did once upon a time the British Empire. British common law allowed abortions to be performed before, but not after, quickening.

The Catholic Church takes it to be quite *possible* that the fertilized egg is a living human being, and that, therefore, abortion might well be murder. The coming into being of the zygote marks a real discontinuity in the process that starts with eggs and sperms, but to regard the zygote as possibly a human seems to erase another presumed radical discontinuity, that between human beings, other animals, and even plants. If single cells are actual human beings, what about similar cells in other animals and in plants? And even if it is accepted that the human zygote is a form of human life, it cannot truly be claimed that an embryo is a human *individual* before the sixteenth day, because up until then there is a possibility that it may become twins (or more).

Faced by facts like these, some religious deontologists take recourse to the philosophical distinction between *actually* being of a certain kind and

having the *potentiality* of becoming a being of this kind. Then they say (in what is often called ‘the potentiality argument’): neither a sperm nor an egg has in itself the potentiality of becoming a human being, but the fertilized egg and each of its later stages has. Whereupon they interpret (or re-interpret?) the norm of not killing as saying: ‘You shall not kill living entities that either actually or potentially are human beings!’ The main counter (reductio in absurdum) argument goes as follow: if from a moral point of view we should treat something that is potentially H (or: naturally develops into H) as already actually being H, then we could treat all living human beings as we treat their corpses, since naturally we grow old and die; potentially, we are always corpses.

Even if, the comments above notwithstanding, a duty ethicist arrives at the conclusion that life begins at conception, abortions may still pose moral problems for him. How should he deal with extra-uterine pregnancy? If the embryo is not removed, then the woman’s life is seriously at risk, and gynecologists are not yet able to remove the embryo without destroying it. The duty to sustain the life of a fetus (a potential human being) is here in conflict with the duty to help an actual human being in distress, or even in life danger. The Catholic Church solves the problem by an old moral principle called ‘the Principle of Double Effect’. It says that if an act has two effects, one of which is good and one of which is evil, then the act may be allowed if the agent intends only the good effect. In the case at hand, it means that if the abortion is intended only as a means to save the life of the mother it can be accepted, even though a potential human being is killed; a precondition is of course that there is no alternative action with only good effects, i.e., an action that saves the life of both the fetus and the mother.

Let us take the opportunity to say some more words about the principle of double effect. It has won acceptance even among many non-religious people. For instance, think about whether to provide painkillers (e.g., morphine) to terminally ill patients. Morphine has two effects. It relieves the patient from pain, but it also suppresses the respiratory function, which may hasten the patient’s death. Here, also, the intention of the action might be regarded as crucial. If the intention is solely to relieve the patient’s pain, the action might be acceptable, but if the intention is to shorten life then, surely, the action is unacceptable.

Let us next look at another well-know principle, the principle of respecting the autonomy of the individual, and regard it as a deontological norm. What was earlier a problem of how to find a discontinuity that demarcates living from non-living human beings becomes now a problem of how to find a discontinuity that demarcates individuals from non-individuals. The principle can give to a woman (or the parents) an absolute right to choose abortion only if neither the embryo nor the fetus is regarded as an individual. Note that if an embryo and/or a fetus is regarded as a *part* of the pregnant woman on a par with her organs, then it cannot possibly be regarded as an individual human. But here a peculiar feature appears. The duty to respect others does not imply a duty to take care of others. Therefore, even if a fetus is regarded as an individual human, it seems as if abortion is acceptable from the autonomy principle; only a deontological principle of caring could directly imply a prohibition on abortion. In a much discussed thought experiment (J. J. Thomson 1971), the readers are asked to think that they one day suddenly wake up and find themselves being used as living dialysis machines for persons who have suffered renal failure. Isn't one then allowed just to rise, unplug oneself, and leave the room? But what if one has consented to be such a dialysis machine?

The arguments pro and con abortion that we have now presented in relation to classical deontology can also be arguments pro and con corresponding prima facie duties.

Some clergies do not consider abortion as a sin in itself, but use consequence reasoning in order to show that to allow abortions will probably lead to more sinful actions, i.e., actions that break some deontological norm. Already in the seventeenth century a Jewish rabbi, Yair Bacharach (1639-1702), who thought that fetuses are not human beings, argued that to allow abortions might open floodgates of amorality and lechery.

In rule utilitarianism we find a series of different arguments both for and against abortion. Here, the structure of the arguments is different, since there is no basic moral notion of human being or personhood. Utilitarians can very well accept moral rules that rely on completely conventionally created boundaries in the zygote-embryo-fetus development, if only the rules in question are thought to maximize utility.

One basic utilitarian argument for abortion is that an illegalization of abortion produces negative consequences such as deterioration of the quality of life of the parents' and their already existing children, or gives the (potential) child itself a poor quality of life. Such considerations become increasingly strong when the fetus has a genetic or chromosomal anomaly, which will result in a severe disease or mental handicap. In utilitarianism, however, there is no direct argument to defend that the mother or the parents should be sovereign in the abortion decision. Since the consequences can influence not only the family, but also the society at large in psychological, sociological, as well as economic respects, it is also a matter of utility calculations to find out who should be the final decision maker.

New medical technology has introduced quite new consequential aspects of abortion. A legalization of selective abortion or selective fetus reduction based on information about what sex, intelligence, risks for various diseases and disabilities the potential child has will probably also affect how already existing people with similar features experience their lives.

As soon as someone reaches the conclusion that abortion cannot in general be prohibited, he has to face another question: should there be a time limit for how late during pregnancy abortions can be allowed? One might even have to consider Singer's view that abortion can, so to speak, be extended to infanticide. Now we have to look at the development list again. Does it contain any discontinuities of special importance for the new question? And for the utilitarians there are. Especially, there is stage k, sentience. When there is a capacity for sensation and feeling, there is a capacity for pleasure and pain, which means that from now on an abortion may cause pain in the fetus. Abortion pills such as mifepristone might be painless, whereas prostaglandin abortions can cause the fetus to die through (on the assumptions given) painful asphyxiation. For Singer the discontinuity where self-awareness can be assumed to arise (around the third month after birth) is of very special importance.

In many countries where abortion is legal, the woman can decide for herself as long as the fetus is less than twelve weeks old; in some countries the limit is eighteen weeks, and with special permission from authorities it can be allowed up to twenty-two weeks. These limits are not only based on what properties the embryo as such has, they are also based on facts such

as the frequency of spontaneous abortion during different weeks, and the vulnerability of the womb before and after twelve weeks of pregnancy; after twelve weeks the womb is rather vulnerable to surgical intervention. If one day all abortions can be made by means of medical intervention a limit like this have to be re-thought.

Stage j (viability) in our list, which seems unimportant to deontologists because it is heavily dependent on the development of medical technology, receives a special significance in utilitarianism. Why? Because it may make quite a difference to the parents' experiences and preference satisfactions if the aborted fetus could have become a child even outside the uterus. But such experiences can also undergo changes when people become used to new technologies. Here again we meet the utilitarians' problems with actual utility calculations. It depends on the result of his utility calculation whether he should put forward the norm: abortions after the moment of which a fetus becomes capable of surviving outside the womb should not be allowed.

We have defined a modern virtue ethicist as a person who:

(i) accepts moral particularism
(ii) accepts a conflicting psychology
(iii) accepts moral fallibilism
(iv) realizes that characters can be situation-bound
(v) accepts that he has to discuss the moral aspects of certain social rules.

Such a virtue ethicist cannot rest content with merely saying that virtue ethics leaves all rules behind. What does the problem of abortion look like to him? In one sense it is similar to that of the rule utilitarian, i.e., he ought to think about the consequences, but he can do it having some rights of persons as default norms. Modern virtue ethics is so to speak both quasi-deontological, since it accepts default norms, and quasi-consequentialist, since one of its default rules is take also consequences into account. But there is still something to be added: the virtue ethicist is careful as long as he has had no personal encounter with people who have considered abortions, having aborted, and who have abstained from abortion. The more experience of this kind he has, the better. Why? Because through

such experience he may acquire tacit knowledge that influences what abortion rule he will opt for.

9.6 Medical ethics and the four principles

Due to the fact that neither classical deontology nor consequentialism have managed to come up with norms that provide reasonable and comprehensive guidance in the medical realm, the American philosophers Tom Beauchamp and Jim Childress have, with great resonance in the medical community, claimed that the latter in both its clinical practice and research ought to rely on four *prima facie principles* (in Ross' sense). Relying on our comments on moral particularism in the last two subchapters, we would like to reinterpret these principles as *default principles*. But this change is more a philosophical than practical. It means, though, that we think that these rules fit better into virtue ethics than into deontological and utilitarian ethics. The four principles are rules of:

1) Beneficence. There is an obligation to try to optimize benefits and to balance benefits against risks; a practitioner should act in the best interest of the patient and the people. In Latin, briefly: *Salus aegroti suprema lex*.

2) Non-maleficence. There is an obligation to avoid, or at least to minimize, causing harm. In Latin, briefly: *Primum non nocere*. Taken together, these two principles have an obvious affinity with the utility principle; taken on its own, the latter has affinity with so-called 'negative consequentialism'.

3) Respect for autonomy. There is an obligation to respect the agency (cf. Chapters 2.1 and 7.5), reason, and decision-making capacity of autonomous persons; the patient has the right to refuse or choose his treatment. In Latin, briefly: *Voluntas aegroti suprema lex*. This principle has affinity with Kant's imperative never to treat people only as means, and it implies sub-rules such as 'Don't make false promises!', 'Don't lie or cheat!', and 'Make yourself understood!' When respect for autonomy is combined with beneficence, one gets a sub-rule also to

enhance the autonomy of patients; it might be called ‘the principle of empowerment’.

4) Justice. There are obligations of being fair in the allocation of scarce health resources, in decisions of who is given what treatment, and in the distribution of benefits and risks. It should be noted that justice means treating equals equally (horizontal equity) and treating unequals unequally (vertical equity).

These principles are as prima facie principles or default principles independent of people’s personal life stance, ethnicity, politics, and religion. One may confess to Buddhism, Hinduism, Christianity, or Islam, or be an atheist, but still subscribe to the four principles. They might be regarded as the lowest common denominator for medical ethics in all cultures. In specific situations, however, the four principles well may come in conflict with religious deontological norms, secular deontological norms, ordinary socio-political laws, and various forms of utilitarian thinking. How to behave in such conflicts lies outside the principles themselves; they do not provide a method from which medical people can deduce how to act in each and every situation.

Normally, the principles are not hierarchically ordered, but it has been argued (Gillon 2004) that if any of the four principles should take precedence, it should be the principle of respect for autonomy. Sometimes they are presented as (with our italics) ‘four principles *plus attention to scope*’ (Gillon 1994). Obviously, we can neither have a duty of beneficence to everyone nor a duty to take everyone into account when it comes to distributive justice. What scope do then principles one and four have? In relation to principle three one can ask: who is autonomous? Those who subscribe to the four principles have to be aware of this ‘scope problem’. In the terminology we have earlier introduced, we regard this ‘scope problem’ as a ‘phronesis problem’.

Often, in the medical realm ethical reasoning is performed by persons working in teams. This adds yet another feature to the complexity we have already presented. We have spoken of one man’s phronesis and moral decision when confronted with a specific situation. But in teams, such decisions ought to be consensual. For instance, a ward round can contain a

chief physician, specialist physicians, nurses, and assistant nurses. When, afterwards, they discuss possible continuations of the treatment, even moral matters can become relevant, and the differences of opinion can be of such a character that literal negotiations with moral implications have to take place. Then, one might say that it befalls on the group to develop its collective phronesis. In clinical ethical committees, which can be found in almost every Western hospital, similar kinds of situations occur. In such committees, which are not to be confused with research ethics committees (see Chapter 10), representatives from various specialties and professions convene in order to solve especially hard and critical situations.

9.6.1 The principles of beneficence and non-maleficence

As is usually done, we will remark on the first two principles simultaneously. One reason for bringing them together is the empirical fact that many medical therapies and cures have real bad side effects; surgery and chemotherapy might cure a patient in one respect, but at the same time cause some handicap in another. Sometimes one dangerous disease is treated with another dangerous disease; for instance, in 1927 the Austrian physician Julius Wagner-Jauregg (1857-1940) received the Nobel Prize for treating syphilis with malaria parasites.

But the two principles belong together even from a philosophical point of view. If there are several possible beneficial, but not equally beneficial, alternatives by means of which a patient can be helped, then to choose the least beneficial is, one might say, a way to treat the patient in a maleficent way. As in utilitarianism degrees of pleasures and pains naturally belong to the same scale, here, degrees of beneficent and maleficent behavior belong together. If a treatment has both desired and undesired effects, it is important to balance these effects and try to make the treatment optimal. The same holds true of diagnostic processes; there can be over-examinations as well as under-examinations, and both are detrimental to the patient.

According to the Hippocratic Oath, each physician is expected to testify:

- I will apply dietetic measures for the benefit of the sick according to my ability and judgment; I will keep them from harm and injustice.

This has been regarded a basic norm in medicine since ancient times. In more general words: physicians should try to do what is supposed to be good for the patient and try to avoid what may harm or wrong him. The ambition to avoid making harm is weaker than the ambition ‘first of all do not harm’ (‘primum non nocere’), a saying which is often, but falsely, thought to come from Hippocrates. The latter rule did first see the light in the mid nineteenth century. It was occasioned, first, by the rather dangerous strategy of bloodletting, and, later, the increased use of surgery that followed the introduction of anesthesia. Some patients died as a result of bloodletting; several physicians recommended that as much as 1200 ml blood should be let out, or that the bloodletting should continue until the patient fainted.

While minimizing-harm might appear to be an obvious ambition, the history of medicine unfortunately provides several examples where one has not been careful enough. In Chapter 3.3, we told the sad story about Semmelweis and the Allgemeine Krankenhaus in Vienna. The current perception of what constitutes good clinical practice does not automatically prevent medical disasters. Therefore, in most modern societies, physicians have to be legally authorized before they are allowed to treat patients on their own.

Proposed cures may be counter-productive, especially when a treatment initially connected to a very specific indication is extended to other more non-specific indications. Cases of lobotomy and sterilization provide examples. Lobotomy was invented by António Egas Moniz (1874-1955) to cure psychiatric anxiety disorders. At the time of the invention, no other treatments were available, it was applied only in extremely severe cases, and used only on vital indication. Moniz’s discovery was regarded as being quite a progress, and he was awarded the Nobel Prize in 1949. Unfortunately, it was later used in order to cure also disorders such as schizophrenia and neurosis; it was even applied in some cases such as homosexuality, where today we see no psychiatric disorder.

Before the birth control pills were introduced in the 1960s, it was difficult to prevent pregnancy in mentally handicapped individuals. Therefore, many of these individuals were sterilized. The reason was of course that a mentally disabled person is assumed not to be able to take the

responsibility of parenthood, and that children have a right to have at least a chance to grow up in a caring milieu. Sterilization meant that the mentally handicapped didn't need to be kept in asylums only in order to stop them from becoming pregnant. Today we use birth control pills and other contraceptive medication for the same purpose. Sterilization, however, was also used in connection with eugenic strategies, which are very controversial because mental handicap is not always inherited; and when it in fact is, it is neither monogenetic nor a dominant trait. This means that (according to the Hardy-Weinberg law of population genetics, presented 1908) a sterilization of all mentally handicapped individuals has almost no effect in a population-based perspective. Nevertheless, in several Western countries in the 1940s and 1950s, thousands of mentally disabled individuals, as well as others with social problems, were sterilized on eugenic grounds.

It is easy to be wise after the event; the challenge is to be wise *in* the event. Therefore, doctors have to consider whether their actions will produce more harm than good. It seems to be extremely rare that individual physicians deliberately harm patients, but, as Sherlock Holmes said about Dr Roylot (in 'The Speckled Band), when doctors serve evil they are truly dangerous. The ideal of beneficence and non-maleficence is the ideal for physicians, nurses, and health care staffs. In many countries this fact is reflected in the existence of a special board, which assesses every instance in which patients are put at risk by unprofessional behavior. Also, omitting taking care of patients in medical need is usually understood as severe misconduct.

In the Hippocratic Oath, a physician is also expected to testify that:

- I will neither give a deadly drug to anybody who asked for it, nor will I make a suggestion to this effect. Similarly I will not give to a woman an abortive remedy. In purity and holiness I will guard my life and my art.

This part of the oath has to be re-thought in the light of the beneficence and non-maleficence principles. We have already commented upon abortion. What about euthanasia? Is it defensible to help terminally ill patients suffering unbearable pain to die? What is to be done when no

curative treatment is available, and palliatives are insufficient? Doctors are sometimes asked to provide lethal doses of barbiturate. Do physicians harm a patient that they help to die? Do they harm someone else? As in the case of abortion, there are many complex kinds of cases. Within palliative medicine, a terminally ill patient might be given sedation that make him fall asleep, during sleep no further treatment or nutrition is given, and because of this he dies within some time. Here, again, 'the principle of double effect' might be brought in. One may say that the intention is only to make the patient sleep, which is a good effect, but then inevitably there happens to be also a bad effect, the patient dies. Usually, this is not understood as euthanasia.

The original Greek meaning of 'euthanasia' is 'good death', but the word acquired an opposite negative ring due to the measures imposed by the Nazis during the years 1942-1945. They called euthanasia the systematic killing of people not regarded as worth living (often chronically ill patients and feeble-minded persons). Today, euthanasia is defined as:

- a doctor's intentional killing of a person who is suffering 'unbearably' and 'hopelessly' – at the latter's voluntary, explicit, and repeated request.

Euthanasia is then distinguished from 'physician-assisted suicide', which can be defined as:

- a doctor's intentional helping/assisting/co-operating in the suicide of a person who is suffering 'unbearably' and 'hopelessly' – at the latter's voluntary, explicit, and repeated request.

The Nazi misuse of the word euthanasia has given emphasis to a slippery slope argument against euthanasia: if 'good euthanasia' is legalized, then also 'bad euthanasia' will sooner or later be accepted. And, it is added, when this happens, the trust in the whole health care system will be undermined. In Holland and Belgium, where euthanasia was legalized in 2002 and 2003, respectively, nothing of the sort has so far happened. In a democratic setting, the risk of entering a slippery slope might be very small, but there are still very important concrete issues to discuss. If

society legalizes euthanasia, under what conditions should it be accepted and who should perform it?

In 1789, the French physician Joseph Guillotine (1738-1814) suggested a less harmful method of execution than the traditional ones. Hanging had come to be regarded as inhumane to both the criminal and the audience. The guillotine was introduced with the best of intentions, but during the French Revolution it became a handy instrument of mass executions. Dr Guillotine dissociated himself from its use and left Paris in protest. Against this background, it is worth noting that the American Medical Association has forbidden its members to participate in capital punishment by means of drug injections. Even though chemical methods (e.g., sleeping medicine in combination with curare or insulin) might appear more human and less harming than execution by hanging, gas, or the electric chair, it might be discussed whether it is at all acceptable for a physician to participate in these kinds of activity. In brief: does the principle of non-maleficence imply that physicians ought to refuse to participate in capital punishments?

9.6.2 The principle of respect for autonomy

Patient autonomy might be defined as a patient's right to take part in medical decision-makings that lead to decisions by which he is more or less directly affected. If he is the only one who is affected, then he should be allowed to make the decision completely on his own, but this is rather uncommon. Autonomy means more than integrity and dignity. To respect a patient's integrity or dignity is to respect his wishes, values, and opinions even though he is in the end not allowed to be part of the final decision. Autonomy takes degrees, it might be strong (as in normal adults), weak (as in small children), or entirely absent (as in unconscious patients). Integrity, on the other hand, is usually regarded as non-gradable. Wishes might be important and should be respected as far as possible even in patients that lack autonomy and in dead patients. Such wishes may be manifested orally or in so-called 'advance directories'. Since, advance directories are rather uncommon, relatives who knew the patient may have to interpret his wishes in relation to whether to withhold or to withdraw a treatment or to donate organs or tissues.

If we speak about the patient's right to take part in decision making, it is necessary to distinguish between, on the one hand, the patient's right to

decline examination, treatment, or information and, on the other hand, his entitlement to exercise his positive rights. Usually, it is easy in principle to respect the autonomous patients' right to say no, i.e., to respect their negative rights, but it might nonetheless sometimes be hard in practice. The competent Jehovah's Witness, who rejects life supporting treatment if it includes blood transfusion, is the classical illustrative case. The positive rights, however, are complicated even in principle since their realization involve costs and have repercussions on queuing patients.

Sometimes a patient requests examination and treatment where there seems to be no clear medical need. In such cases, to respect autonomy means initiating negotiations. If a patient with a two day old headache visits his GP and asks for a computer tomography (CT) of his skull, the GP might well challenge the reason for conducting this examination. If, after having examined the patient, the GP finds that the symptoms more probably derive from the muscles of the neck, he should suggest physiotherapy – and ask if this might be an acceptable alternative. Unless the GP suspects that a brain tumor or other pathological change might be the cause, the autonomy principle by no means implies that he should grant the patient's wish. This would be much more than to respect the autonomy of the patient; it would mean that the GP subordinates himself to the patient. A doctor has to make reasonable estimations of probabilities, and react (Africa apart) according the saying, 'if you hear the clopping of hooves outside the window, the first thing that comes to mind is not a zebra'.

Were all patients presenting a headache referred to a CT of the skull, radiologists and CT-scanners would be swamped. The skilled GP should be able to distinguish between a patient with a tension-based headache and a brain tumor. The reason why the GP does not simply agree is thus partly his skill and knowledge. Patients have to accept a special kind of autonomy, the 'professional autonomy' of the physicians. Doctors should consider the patients' problem with empathy and with respect of their autonomy, while at the same time respecting *his* own professional autonomy. Ideally, the negotiations between doctors and patients should be based on mutual trust and aim at reaching consensus. This view is referred to as 'patient-centered medicine', in contrast to 'physician-centered medicine'. The latter is characterized by the doctor setting the agenda, and

since the doctor knows more about medicine, he is also assumed to know independently of the patients' wishes what is best for the patient.

Since autonomy has degrees, the autonomy principle cannot possibly altogether forbid paternalistic behavior on the part of physicians and health care workers. The word paternalism comes from the Latin 'pater', which means father, and today it refers to the decision making role of the father in traditional families. Parents of small children must decide what is in the best interests of their children, and now and then there are structurally similar situations in medicine. When there is a considerably reduced autonomy, as in some psychotic patients, there might be reasons not even to respect the patient's negative right to refuse treatment. This might be called 'mild paternalism'. All paternalistic actions must be for the benefit of the patient; some paternalistic actions can be defended by an assumption that the patient will approve of the action when he recovers from his disease or disorder; this is so-called 'presumed informed consent'. Of special concern are patients suffering from Alzheimer's disease. When the disease has progressed, the patients are often in a state in which they do not know their own best interest, and here weak paternalism is the rule. Not being paternalistic would amount to cynicism.

During the last decades of the twentieth century, the doctor-patient relationship became both formally (in laws) and really 'democratized', i.e., the doctor's paternalism was decreased and the patient's autonomy increased. The patient-centered method is an expression of this change. In the Hippocratic period, medical knowledge was secret, and not supposed to be made freely available; during the last centuries it was public in theory but hard to find for the laymen. Today, internet has changed the picture completely. Rather fast, and with a very small effort, many patients check a little on the internet what their symptoms can mean before they meet their GP. And after having been diagnosed, some of them seek contact with other patients with the same diagnosis in order to discuss treatments and side effects or they ask another doctor for a second opinion.

9.6.3 The principle of justice

Modern medical research has brought about many new medical technologies within preventive and curative medicine, as well as very advanced high tech examinations, and our knowledge of diseases, their

etiology, and their pathogenesis has grown. In other words, health care systems are today able to do and offer much more than only, say, three or four decades ago. During the 1970s and 1980s, many health care systems were in a phase of rapid expansion from which quite a number of new specialties emanated. Many modern hospitals and health care centers were built, and an increasing number of physicians and nurses were educated to meet the needs.

At the same time, quite naturally, the patients' expectations rose. During the 1980s, however, the costs of the expansion became a public issue much discussed by politicians. To put it briefly and a bit simplified, it became clear that the costs of the health care system had reached a limit. This insight put to the fore priority settings. Problems of what a fair health care distribution and a fair cost distribution should look like came to be regarded as very serious. Should the healthcare system be based on public finances, private finances, or a mixture of the two? Should patients be provided with health care through a random principle? Should health care be provided depending on the patients' verbal capacity, financial situation, and social influence? Should it be based on some egalitarian principle? Let us take a brief look at some of these options.

By means of thought experiments, it might be easy to conceive of situations where priority settings based on a random principle makes sense. For example, assume that Smith and Thomson (who are of the same sex and age, and have the same position in society, etc.) have suffered from a heart attack, and have received the same diagnosis and the same prognosis both with and without treatment. Furthermore, they arrive at the same emergency clinic at the same time, where, unhappily, it is only possible to treat one of them with the high tech means available to the cardiology intensive care unit. Here it seems adequate to flip a coin. If Smith wins, the doctor might have to say to Thomson: 'Sorry, you were out of luck today. Please, go home and try to take it easy the next couple of weeks. You may die or you may survive, and if you survive you are welcome back should you suffer a new heart attack'. Reasonable as it may seem, one might nonetheless ask if this would be fair to Thomson. In one sense Smith and Thomson are treated equally (when the coin is flipped), and in another they are treated unequally (only one is given the treatment). If both Thomson and Smith had voted for the political party that had introduced the random

system, it might seem fair to treat them in the way described, but what if the procedure has been decided quite independently of Smith's and Thomson's views? Would it be fairer to treat neither Thomson nor Smith at the clinic, and offer both of them low tech oriented care and painkillers?

In thought experiments one can imagine people to be exactly similar, but in real life this is hardly ever the case. Different patients suffer from different diseases, and some of them have had the disease a long time and others only a short time. Furthermore, some diseases are life threatening and others not. Some patients are young and others old, some are male and some are female, and they come from different ethnic cultures. Also, they have different social status, different incomes, and pay different amounts of taxes. Some are employed and some unemployed, some are refugees and some have lived in the society in question for generations. How to apply a random principle in the midst of such variations? In several countries where priority rules have been developed, it is said that those most in need of a treatment should be prioritized and assessed in relation to the severity of the disease. Sometimes, however, so-called VIPs (famous politicians, movie stars, sports heroes, etc) become (especially on operation lists) prioritized before other patients who have been waiting for a long time.

Brainstormers have argued that the state should at birth supply each member of the society with a certain amount of money, which should be the only money he is allowed to use for his health care. Furthermore, during the first seventy years he is not allowed to use this sum on anything else but health care, but then he is free to use what remains in any way he wants. This means that when the money runs out, it is impossible to receive more health care services. If you are unlucky and suffer from diabetes or a chronic disease early in your life, then the health care money may run out when you are quite young, whereas if you are lucky you can have much money to spend on possible later diseases. Even in this proposal, individuals are in one sense treated equally (they receive the same sum at birth) and in another unequally (people who suffer the same disease may nonetheless later in life not be able to buy the same treatment). The proposal is in conflict with the kind of solidarity based justice that says that it is fair that the healthy citizens subsidize the sick ones. Others claim that such solidarity justice is theft, and that the only fair way of

distributing health care resources is to let only people who can pay for it receive it. In the case of Thomson and Smith described, the reasonable thing would then be to sell the high tech treatment by means of an auction where Thomson and Smith can bid.

A special problem of justice in priority settings appears in relation to wars and war-like situations. Is it fair to extend to enemies a principle such as ‘patients who are most in need of treatment should be given treatment first’? The extreme case is a suicide bomber that survives together with many injured people. Assume that he is the most injured, and that to save his life would imply a long and complicated operation that would take resources away from the innocently injured, and cause them further suffering; some might even die. What to do? Justice seems to be justice in a pre-given group, and then one can ask what this group looks like for physicians. Does he belong to mankind as a whole, to the nation to which he belongs, or to some other community? Is perhaps the question of justice wrongly formulated? Perhaps an analogy with counsels for the defense in law courts is adequate? The judge and the jury are obliged to come to a fair decision (and a possible punishment), but the counsel is not. Society has prescribed a division of labor according to which the task of the counsel is to work solely in the interest of his clients. Perhaps the world community should place physicians in a similar situation in relation to prospective patients?

Reference list

Apel K-O. *Towards a Transformation of Philosophy*. Routledge. London 1980.

Aristotle. *Nicomachean Ethics*. (Many editions)

Beauchamp TL, Childress JF. *Principles of Biomedical Ethics*. Oxford University Press. Oxford 2001.

Bentham J. *An Introduction to the Principles of Morals and Legislation*. (Many editions)

Bernstein RJ. *Beyond Objectivism and Relativism*. Basil Blackwell. Oxford 1983.

Bok S. *Lying. Moral Choice in Public and Private Life*. Vintage Books. New York 1979.

Brandt, RB. *A Theory of the Good and the Right*. Oxford University Press. Oxford 1979.

Broad CD. *Five Types of Ethical Theory*. London. Routledge 2000.

Capron AM, Zucker HD, Zucker MB. *Medical Futility: And the Evaluation of Life-Sustaining Interventions*. Cambridge University Press. Cambridge 1997.

Crisp R, Slote M (eds.). *Virtue Ethics*. Oxford University Press, Oxford 1997.

Curran WJ, Casscells W. The Ethics of Medical Participation of Capital Punishment by Intravenous Drug Injection. *New England Journal of Medicine* 1980; 302: 226-30.

Dancy J. *Ethics Without Principles*. Oxford University Press. Oxford 2004.

Dancy J. Moral Particularism. *Stanford (online) Encyclopedia of Philosophy* (April 11, 2005)

Doris JM. *Lack of Character: Personality and Moral Behavior*. Cambridge University Press. New York 2002.

Dreyfus H, Dreyfus S. What is Moral Maturity? Towards a Phenomenology of Ethical Expertise. In Ogilvy J. (ed.). *Revisioning Philosophy*. Suny Press. New York 1986.

Edelstein L. *The Hippocratic Oath: Text, Translations and Interpretation*. John Hopkins Press. Baltimore 1943.

Fulford KWM, Gillet G, Soskice JM. *Medicine and Moral Reasoning*. Cambridge University Press. Cambridge 1994.

Fletcher J. *Situation Ethics: The New Morality*. John Knox Press. Louisville, Ky. Westminster 1997.

Fesmire S. *John Dewey: Moral Imagination. Pragmatism in Ethics*. Indiana University Press. Bloomington 2003.

Gewirth A. *Reason and Morality*. The University of Chicago Press. Chicago 1978.

Gilligan C. *In a Different Voice: Psychological Theory and Women's Development*. Harvard University Press. Cambridge Mass. 1982.

Gillon R. Lloyd A. *Health Care Ethics*. Wiley. Chichester 1993.

Gillon R. Medical Ethics: Four Principles Plus Attention to Scope. *British Medical Journal* 1994; 309: 184.

Gillon R. Ethics needs principles – four can encompass the rest – and respect for autonomy should be 'first among equals'. *Journal of Medical Ethics* 2003; 29: 307-12.
Habermas J. *Moral Consciousness and Communicative Action*. Polity Press. Cambridge 1990.
Hare RM. *Moral Thinking*. Oxford University Press. Oxford 1981.
Helgesson G, Lynöe N. Should Physicians Fake Diagnoses to Help their Patients? *Journal of Medical Ethics* (forthcoming).
Holm S. The Second Phase of Priority Setting. *British Medical Journal* 1998; 317: 1000-7.
Kjellström R. Senilicid and Invalidicid among Eskimos. *Folk* 1974/75; 16-17: 117-24.
Kant I. *Groundwork of the Metaphysics of Morals*. (Many editions.)
Kohlberg L. *The Philosophy of Moral Development*. Harper & Row. San Francisco 1981.
Kuhse H, Singer P. *Should the Baby Live?* Oxford University Press. Oxford 1985.
Lynöe N. Race Enhancement Through Sterilization. Swedish Experiences. *International Journal of Mental Health* 2007; 36: 18-27.
McGee G. *Pragmatic Bioethics*. Bradford Books. Cambridge Mass. 2003.
MacIntyre A. *After Virtue*. Duckworth. London 1999.
Mill JS. *Utilitarianism*. (Many editions.)
Norman R. *The Moral Philosophers. An Introduction to Ethics*. Oxford University Press. Oxford 1998.
Nord E. *Cost-Value Analysis in Health Care. Making Sense of QALYs*. Cambridge University Press. New York 1999.
Nussbaum, MC. *Love's Knowledge: Essays on Philosophy and Literature*. Oxford University Press. New York 1990.
Nussbaum M. *Upheavals of Thought: The Intelligence of Emotions*. Cambridge University Press. Cambridge 2001.
Rawls J. *A Theory of Justice*. Harvard University Press. Cambridge Mass. 1971.
Ross WD. *The Right and the Good*. Oxford University Press. Oxford 2002 .
Sharpe VA, Faden AI. *Medical Harm: Historical, Conceptual, and Ethical Dimension of Iatrogenic Illness*. Cambridge University Press. Cambridge 1998,
Singer P. *Practical Ethics*. Cambridge University Press. Cambridge 1993.
Smith B, Brogaard B. Sixteen Days. *The Journal of Medicine and Philosophy* 2003; 28: 45-78.
Smith CM. Origin and Uses of Primum Non Nocere – Above All, Do No Harm! *Journal of Clinical Pharmacology* 2005; 45: 371-7.
Statman D. (Ed.) *Virtue Ethics. A Critical Reader*. Edinburgh University Press. Edinburgh 1997.
Swanton C. *Virtue Ethics*. A Pluralistic View. Oxford University Press. Oxford 2003.
Thomson JJ. A Defense of Abortion. *Philosophy & Public Affairs* 1971; 1: 47-66.

Umefjord G, Hamberg K, Malker H, Petersson G. The use of Internet-based Ask the Doctor Service involving family physicians: evaluation by a web survey. *Family Practice* 2006; 23: 159-66.

World Medical Association, Declaration of Geneva 1948. See e.g. < www.cirp.org/library/ethics/geneva >

10. Medical Research Ethics

Medical research ethics is of course part of both medical ethics and general research ethics. From medical ethics it takes over ‘the Four Principles’ just presented (Chapter 9.6), and from general research ethics it inherits the norm ‘Be Honest in Research!’, which we touched upon when discussing what a scientific fact is (Chapter 3.1). Apart from these *moral norms*, research also contains intra-scientific *methodological norms*, i.e., each scientific disciplines default rules for how to do good and efficient research. Since moral norms overrule all other kinds of norms, it may seem as if the relationship between the moral and the methodological norms is one-sided; moral norms can put constraints on methodological norms – period. But this is spurious. One then forgets that ethics (classical deontology apart) often needs to take consequences into account, too. And to constrain the methodological rules for moral reasons may have bad consequences. In this way, moral and methodological norms may interact. This interaction will be highlighted below, and in the last subchapter (Chapter 10.5) we will present a mixture of moral and methodological norms that are called ‘the CUDOS norms’.

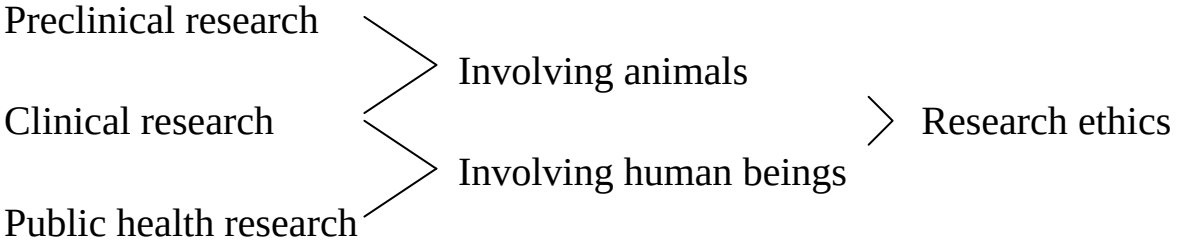
Since medical research is typically done on human beings and animals, medical research has a moral aspect that is absent from research in physics and chemistry. But it should not be forgotten that much medical research (e.g., on tissues, genes, and cell-lines) is also done in physico-chemical laboratories. A simplified but common division of medical research is: 1) preclinical research, 2) clinical research, and 3) public health research.

1) Preclinical research overlaps with natural scientific research. One might define it as natural scientific research that aims at finding knowledge that is helpful in inventing new medical therapeutic strategies. It is often concerned with chemical, physiological, microbiological, anatomical, histological, molecular-biological, and genetic analyses. Both experimental and observational studies are used; the experiments can be made on animals.

2) Clinical research involves healthy and/or more or less seriously ill human beings, and it tries directly to test newly invented drugs and medical diagnostic and therapeutic devices, involving all clinical specialties. Sometimes, this kind of research is divided up into four phases, I-IV, preceded by phase 0, which is preclinical research. In phase III, one performs the kind of randomized control trials that we presented in Chapter 6.3; phase 4 is mainly post-launch safety and generalization surveillance. In phase I there can be experiments on animals; it is a kind of safety studies made in order not to put the human participants of phases II and III into too risky situations.

3) Public health research is dominated by epidemiology, but it also contains specialties such as environmental medicine, social medicine and health care sciences. It often aims at finding preventive measures; and to this effect the researchers study very large groups or whole populations. Often, in this undertaking phenotype and genotype information from different registers and bio-banks are used. When, here, researchers move from their statistical findings to making claims about causal relations, they may use Bradford Hill's criteria, which we presented and commented upon in Chapter 6.2.

To repeat: this tripartition is a simplification. It does not, for instance, consider mutual interdisciplinary collaborations typical of some fields. Nor is it complete; research in disciplines such as medical sociology, medical anthropology, and medical ethics, have got no place in the list. As a main memory device in relation to the list, the following chart can be used:



Today, there exist a number of research ethical guidelines that regulate how to deal with research in different medical fields. We will describe below their historical background, as well as comment more on the overarching research ethical aspects that medicine has in common with other scientific areas. First of all, however, we would like to make it clear why it can be regarded as unethical to abstain from carrying out medical research.

10.1 Is it unethical not to carry out medical research?

The aim of health care is to prevent illness; when this is not possible, to cure; and when this is not possible, to alleviate suffering; and when even this is not possible, to comfort the patient. In all these undertakings, one should avoid harming people. These rankings are in accordance with the principles of beneficence and non-maleficence. Now, in order to prevent, cure, alleviate, comfort, and even to avoid harming, we need knowledge; therefore, we must to produce knowledge. We need of course knowledge in relation to illnesses, diseases, and disabilities that we cannot at the moment do anything at all about, but often we also need new knowledge in order to be able to improve on already existing therapies. Therefore, it is unethical to abstain from medical research. Let us concretize.

When cytostatics were introduced in cancer treatment, they were given to all cancer patients, even to terminally ill patients. The latter were treated with cytostatics for palliative reasons, and sometimes simply in order not to take away the patient’s hope of being cured. In the latter case, the doctor did not know whether the treatment would have any effect other than the side effects, which for most cytostatics include nausea, vomiting, diarrhea, hair loss and more life-threatening conditions as leucopenia with risk of lethal infections. Sometimes the main effect of the treatment on the terminally ill patients was that their last days on earth were made excruciating. In this way the treatment harmed the patients, but if it kept their hope of being cured alive, then at the same time it comforted the patients a little. How to weigh these factors against each others? Answer: research. How to improve the situation radically? Answer: research. As there are numerous different cancer diseases, as well as plenty of different mechanisms by means of which treatment functions, we need to find out whether certain combinations of cytostatics and other drugs prolong

survival in relation to quality of life, and which treatments might be useful for palliative purposes for different kinds of cancer.

The argument in favor of research in the two paragraphs above seems straightforward, but it is much too simplified. It forgets the fact that research has its costs. If in order to make some sick people healthy, we have – in research – to make some healthy people suffer, it is no longer obvious that the research in question should be performed. We have to start to weigh the pros and cons. And if many research projects are involved, we have to start thinking about making rules that prohibit that research costs too much from a human suffering point of view. The main principle that has emerged during the development of medical research is the principle of informed consent:

- Participants in medical research investigations of all kinds must have given their explicit consent – after having been informed about what the investigations amount to.

The relationship between this principle and the principle of respect for autonomy is obvious. However, if patients in a trial are informed in detail and in advance about what kind of treatment they are receiving, e.g., placebo or a promising new treatment, this information may render the result of the trial hard to assess (see Chapter 6.3). But problems are there to be solved. The introduction of the randomized double-blind test solves this problem. Instead of being given detailed information about what treatment they will be provided, the participants are informed generally about the design of the study, and are then asked to participate. That is, when they consent they know that neither they nor the doctors will until afterwards know whether they belong to the experimental group or the control group (which is given placebo or the old treatment). This respects the autonomy of the participants, even though among some of them it may create uneasiness that they do not know whether they receive a treatment or not.

Sometimes, one wants to investigate a new treatment in relation to a life-threatening disease where previously no treatment has been available. If the new treatment proves to provoke numerous bad side effects and is ineffective, the placebo group might well have received the less harmful treatment. But if the new treatment proves to be effective and life saving,

most of the patients in the placebo group have missed an effective treatment; several of them might even have died before the result of the trial had become clear and they could have received the new treatment. It is only in retrospect, and then in a metaphorical sense, that one can say that the placebo patients were ‘sacrificed’ for the happiness of future patients.

Autonomy, as we have said, takes degrees. Therefore, if possible, already acceptable designs should be improved on. One proposal that works when the control group receives an old treatment, is the so-called ‘Zelen’s pre-randomization procedure’; after Marvin Zelen (b. 1927).

In this design, the participants are randomized to the control group and the experimental group, respectively, before the informed consent is asked for; and the consent is then asked for only in the experimental group. In this way, no participant needs to live with the uneasiness of not knowing whether he receives a treatment or not. Furthermore, it seems to respect the autonomy of the members of the experimental group even more than in the ask-first-randomize-then procedure, since now the patients in the experimental group is given the information that they belong to this group. But what about the required respect for the autonomy of the patients in the control group? They are unknowingly becoming part of a medical research trial. Is this to respect their autonomy? It is an empirical fact that some patients afterwards, when they have become aware of the whole thing, have become morally quite upset. However, moral views are fallible and so are moral reactions; moral views can be discussed. One may well challenge these reactions by the following line of argument:

premise 1:	apart from the experimental group and the control group, there is a third group, ‘the rest group’, i.e., the much larger group of all those who receive the old treatment without being part of the investigation
premise 2:	there seems to be no norm saying that the rest group has to be informed about the fact that a new treatment is being investigated; and if there were such a norm, it would nonetheless be practically impossible to implement
premise 3:	there is no important difference between the control group and the rest group
hence:	----------
conclusion:	Zelen’s procedure does not violate the autonomy principle; it does not even decrease the autonomy of the control group

Ethically good research is not marked by ‘consent after as much information as possible’; it is marked by ‘consent after adequate information’.

Trivially, an investigation that has a high ethical quality may nonetheless have big intra-scientific methodological flaws. For instance, the data can be badly collected and treated badly from a statistical point of view. The four-fold matrix in Figure 1 is good to keep in mind in what follows, even though it turns two variables that take many degrees into simple plus-or-minus variables:

		The ethical quality of the research is:	
		High	Low
The methodological quality is:	High	1	2
	Low	3	4

Figure 1: *Ethical and methodological quality as two dimensions of research.*

Historical examples show that medical research of high methodological standard and low ethical standard (square 2) are not only in principle possible but has been conducted; and even conducted without meeting the most elementary ethical requirements.

We can now return to the title of this subchapter: ‘Is it unethical not to carry out medical research?’ Our answer is that – when the economic constraints have been given their due – yes, it is unethical not to do research that fits square 1 in Figure 1.

10.2 The development of modern research ethics

There is no absolute gap between clinical practice and clinical research. Consequently, there is no absolute gap between ethical rules for physicians and such rules for medical researchers. Nonetheless, it is convenient to draw a line and keep them distinct. In the great civilizations in history there have always been oaths and codes for doctors. We have several times mentioned the Hippocratic Oath, but there have been other similar ones. Already at the ancient Vedic times, long before Hippocrates, India had its ‘Charaka Samita’, which is a main text in book *Ayurveda* (‘Science of a long life’); and ancient China had the Taoist writer Sun Szu-miao (682-581 BC). In modern times, the English physician Thomas Percival’s (1740-1804) book *Medical Ethics; or, a Code of Institutes and Precepts Adapted to the Professional Conduct of Physicians and Surgeons* (1803) is a landmark. The first code adopted by a national organization of physicians is the *Code of Ethics* that the American Medical Association put forward in 1846. It very much took its departure from Percival’s writings. However, pure research guidelines had to await a more pronounced division of labor between clinical practice and clinical research. Furthermore, in order to become publicly significant, such guidelines had to await a medical-moral catastrophe: the experiments conducted in the Nazi concentration camps.

These experiments were for a long time perceived as being also methodologically poor, i.e., as also being bad science. They were simply regarded as bad (square 4 in Figure 1); period. However, during the last decades, historians of research have shown that, apart from some very strongly ideologically influenced projects, many concentration camp experiments were conducted in a methodologically proper manner; at least as measured by the contemporary methodological standards. The post-war discussion of the Nazi-experiments lead directly to the Nuremberg Code (1947), which is a forerunner of the first truly modern medical research ethical guidelines, the Helsinki Declaration of 1964, which, in turn, has later undergone several revisions and clarifications. The fact that the Nazis

had experimented on concentration camp prisoners as if the latter were laboratory animals was publicly revealed during the post-war Nuremberg trials, where a group of prominent Nazi leaders such as Hermann Göring, Rudolf Hess, Joachim von Ribbentrop, and Albert Speer were put on trial. This first series of trials started in 1945, and Göring's defense happened to refer to the concentration camps experiments in such a way that the Americans started to scrutinize this issue. A huge amount of documentation was found, which lead to another series of trials, referred to as the Medical Case, which ran during 1947. The prosecution focused on two issues:

1) experiments on human beings conducted in the concentration camps (1942-1943);
2) the so-called 'Euthanasia program' (1942-1945), which was a euphemism for the systematic killing of human beings whose lives were considered not worth living (e.g., mentally or somatically chronically ill persons); in the Nazi terminology they were referred to as 'useless eaters'.

The charges did not include the Nazi sterilization program. Probably, the reason for this was that sterilizations had also been conducted in several other European nations, as well as in the US. However, research about sterilization procedures, which was conducted in the concentration camps of Auschwitz and Ravensbrück, became part of the trial.

The chief counsel for the prosecution was Brigadier General Telford Taylor who, in his opening statements on the research projects, characterized the experiments as being: (a) bestial, (b) unscientific, and (3) useless. For some (perhaps utilitarian or rhetoric) reason he claimed that: 'the accused doctors become guiltier if it is shown that the experiments were of no use. It was then a useless sacrifice of human beings.'

The strategy of the American prosecution was to demonstrate that the Nazi research was perverse both ethically and scientifically, indicating that bad ethics and bad science go together; and that bad science leads to useless results, which makes what was from the start ethically bad even worse. The German defense of the twenty-three German doctors on trial accepted, that if it could be proven that the results were of no use, then the

crimes of the doctors they defended would be worse. But, they also said, this means, conversely, that if it can be shown that the results were of great use, then the severity of the crime would have to be regarded as being small. The assumed relationship is represented by the diagonal line in Figure 2 below.

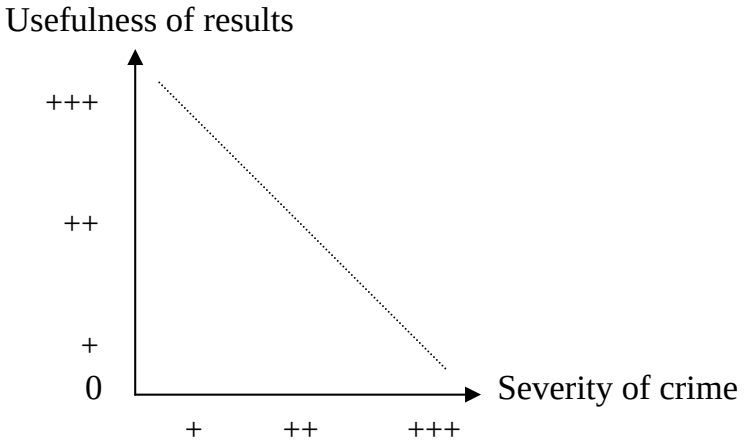


Figure 2: *Proposed relationship between the results and the ethical standards of research.*

The strategy of the counsels for the defense was now clear. Their ambition was to prove that the concentration camp research had been conducted according to good scientific practice, and that it had led to useful results. While admitting that the research required the sacrifice of a large number of human beings, it was, they claimed, for the greater benefit of mankind, or at least for the German military and its soldiers.

The idea to use prisoners from the concentration camps as experimental objects was initiated by Dr. Sigmund Rascher (1909-1945). He performed research on how the human body may react in high altitudes. This research had as its background the English introduction of radar technology. In order to be able better to discuss how to avoid the radar, the German military wanted to know on what altitude and for how long a pilot was able to fly with very little oxygen. Rascher and his co-workers had conducted experiments with monkeys in pressure chambers, but it proved impossible to use animals, since they refused to keep quiet. In the spring of 1941, Rasher wrote to Heinrich Himmler, the leader of the SS and the man responsible for the concentration camps:

> considerable regret was expressed that no experiments on human beings have so far been possible for us because such experiments are very dangerous and nobody is volunteering. I therefore put the serious question: is there any possibility that two or three professional criminals can be made available for these experiments? … The experiment, in which the experimental subjects of course may die, would take place with my collaboration… Feeble minded individuals also could be used as experimental material. (*Trial of War Criminals Before the Nuernberg Military Tribunals*. Volume I 'The Medical Case'. Nuernberg October 1946 – April 1949)

Himmler's adjutant promptly replied that they should gladly make prisoners available for the high altitude research. This simple request is in retrospect merely the beginning of a large-scale research program spanning a variety of disciplines and involving thousands of prisoners. It is notable that Rascher uses the term 'professional criminals', although the only 'crime' committed by the majority of the people in the concentration camps was that they had been born into a certain ethnic minority, like Jews and Gypsies, or that they were political dissidents.

Apart from the high altitude experiments, there were in the camps Auschwitz, Buchenwald, Dachau, Natzweiler, Ravenbrück and Sachsenhausen, experiments concerned with freezing, malaria, sulfonamide, epidemic jaundice, and spotted fever, as well as bone, muscle and nerve regeneration and transplantation. Experiments of purely military nature dealt with poison, mustard gas, and explosives.

However cruel and inhumane these experiment were, the true point of the defense was that the experiments were systematically and carefully conducted from a scientific methodological point of view; the discussions around the data obtained seemed proper, and the conclusions seemed to be derived in a scientific manner. Below, we present an overview of the structure of the presentation of one of the high altitude studies. Obviously, the structure of this research report is quite similar to contemporary empirical clinical research. This is the way it looks:

I. Introduction and statement of the problem.
II. Procedure of the experiment.
III. Results of the experiment: (a) descending experiment without O_2 breathing; (b) descending experiment with O_2 breathing; (c) falling experiment without O_2 breathing; (c) falling experiment with O_2 breathing.
IV. Discussion of the results.
V. Conclusions from the results.
VI. Summery.
VII. References.

The report in question contains 28 pages with 3 figures and 6 tables; authors were Dr. Rascher and Dr. Romberg of the German Aviation Research Institute.

This military influenced research had no ideological bias towards the Nazi world view; nor is there any reason to suspect that the results obtained were deliberately distorted. One of the most prominent researchers within the aviation research was Dr. Hubertus Strughold (1898-1987), who at least intellectually supported the research conducted in Dachau. After the war he was recruited by the Americans (in the so-called ‘Project Paperclip’). This was a medical counterpart to the recruitment of Wernher von Braun (1912-1977) for American rocket development; Strughold became regarded as the father of American space medicine. Accordingly, his support and approval of the high altitude research in Dachau indicates that it was scientifically properly planned and conducted. In May 2006, let us add, Strughold’s name was removed from the International Space Hall of Fame by unanimous vote of the board of New Mexico Museum of Space History. But this post-war part of the Strughold story belongs rather to war ethics than to medical ethics; or, more precisely, to the ethics of the cold war.

Although Jews and Gypsies were used in the experiments mentioned, there were of course no gains to be made by obtaining results which were not applicable to German pilots. In this military influenced research, the researchers had no motivation to distort the research process in order to ‘prove’ something ideologically convenient, e.g., that Aryan pilots are superior to Jewish pilots. Accordingly, it seems reasonable to assume that

the research was conducted in a methodologically optimal manner, although the treatment of the research participants was horrible. In fact, it was even discussed whether results derived from experiments with Jews and Gypsies could be generalized to Aryan German men.

In many cases the research in the concentration camps was conducted in collaboration with universities. The prominent professor Gerhard Rose, from the Robert Koch Institute in Berlin, was deeply involved in vaccination experiments; and the above-mentioned Rascher collaborated with Professor Eric Holzlöner, from the University of Kiel. Rose was sentenced to life imprisonment in the Nuremberg Trials; Rasher was killed by the Nazi SS just before the end of the war; and Holzlöner committed suicide after the war. Some experiments were conducted in close cooperation with the Kaiser Wilhelm Institute in Berlin. Obviously, many prominent medical researchers regarded concentration camp based research as an exceptional opportunity to do research on human beings without any ethical constraints.

When recruiting participants from the camps, the researchers were choosy; the prisoners had to be in good nutritional conditions, for each experiment only one sex was recruited (mostly men), and the participants in the experiment should be of similar age and have the same ethnic background in order to make the biological variability minimal. Furthermore, when experimenting with for example typhus vaccines, the quantity of typhus material injected in the individuals were exactly the same and subsequently even the set-out of the disease as well as the duration of the disease. Today, when conducting randomized controlled trials these biological variables are not under control. Dr. Erwin-Oskar Ding-Schuler (1912-1945) conducted a vaccine experiment with the results presented in Figure 3 below.

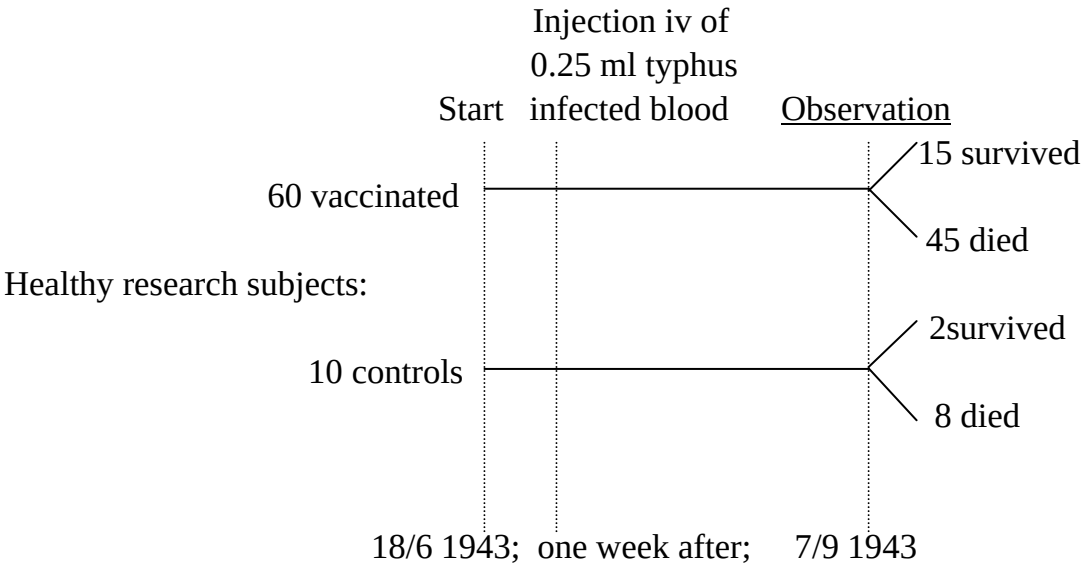


Figure 3: *Result of Nazi experiment with typhus injections.*

In this study, Ding-Schuler had almost all relevant biological variables under control; it is as if it had been an animal experiment with identical guinea pigs. A control group was also included, receiving the standardized injection of typhus but not vaccinated in advance.

Several vaccines that were supposed to be effective were in this way systematically studied, and most of them turned out not to be effective; as in the case summarized in Figure 2. This does not mean that the experiments were useless in Telford Taylor's sense. To exclude ineffective medical technologies is useful. Accordingly, these negative results are no arguments against the research in question. But, apart from the obvious crimes committed on the participants, there are two other aspects that also shed light on what research ethics is concerned with.

First, scientific research is supposed to be made publicly available for two reasons: (a) no one should be permitted to own basic knowledge, and (b) the scientific community should be able to criticize and scrutinize presumed knowledge. But, of course, this military research was not published in any journal. It was top secret, and in this respect it failed to follow international scientific standards, but so did some of the Allies war research.

Second, and more interestingly, the possibility of unethical research makes problematic the common requirement that experiments should be

reproducible. Should the Nazi experiments we are considering be regarded as reproducible or not? They are *not* reproducible if normal ethical guidelines shall be followed, but if the latter are abstracted away, then of course the experiments are reproducible. As a matter of post-war fact, when researchers (who have made purely observational studies on patients with accidentally caused severe cold injuries) have made references to the Nazi-experiments, debate has ensued. At least one medical journal, the *Norwegian Medical Journal* (‘Tidsskrift for Den norske laegeforening’*)*, has completely forbidden references to Nazi research. The journal finds it morally unacceptable to use results gained from experiments conducted under monstrous conditions. As it happens, most of the results from the concentration camp experiments are today obsolete and no longer of any medical interest, but there are some exceptions such as results from freezing experiments. Having sailors and pilots in mind, the Nazis tried to find out for how long people can survive in cold and very cold water. In connection with such research, the first life jackets equipped with a neck-brace in order to keep the neck (medulla oblongata) above the water were constructed.

Summing up, chief counsel Telford Taylor’s original framework, which worked only with a dichotomy between good and bad science (in Figure 1: squares 1 and 4, respectively), could not do the job it was supposed to, since much of the concentration camp research was methodologically acceptable and gave useful results. Of course, the intuition remained that something was seriously morally wrong with these experiments, i.e., square 2 in Figure 1 started to become clearly visible; also to the defense. They therefore began to argue that there was in principle no difference between the Nazi concentration camp research and some research made in US penitentiaries during the war. Prisoners were not allowed to become US soldiers, but they asked whether they might show their loyalty to the American nation and serve their country in other ways; medical research (e.g., malaria studies) was suggested, and many accepted to become participants in such military related experiments. In fact, prisoners continued to be used in this way even after the war; this kind of research was only regulated in 1973.

When the German defense compared the experimentation with prisoners in Nazi concentration camps with those of prisoners in US penitentiaries,

they probably thought that in both cases one could equally well say that the rule ‘let the end sanctify the means’ had been applied. That is, even though it might superficially look as if the Nazi experiments belong in square 2 (good science, bad ethics), a more careful look, which brings in the fact that some ends can sanctify some extraordinary means, places the experiments in square 1 (good science, good ethics). But now the American lawyers saw clearly the essence of the bad ethics in the Nazi experiments: in contrast to the German prisoners used, all the American ones had voluntarily participated; and before they volunteered they had been informed about the risks. In modern words, the US prisoners had given informed consent, the concentrations camp prisoners had not. The US experiments were ethically acceptable (square 1), but the Nazi experiments were not (square 2). At this moment, the judges adjourned the trial, medical advisers were called upon, and the so-called ‘Nuremberg Code’ was composed.

10.3 The Nuremberg Code and informed consent

As the simple framework first suggested by Telford Taylor had shown to be insufficient, the American lawyers now created a list of thoroughly considered principles, which ought to function as a kind of informal laws in the case at hand. The first point in this list introduces the essence of informed consent; and the list is concise enough to be quoted in its entirety:

> 1. The voluntary consent of the human subject is absolutely essential. This means that the person involved should have legal capacity to give consent; should be situated as to be able to exercise free power of choice, without the intervention of any element of force, fraud, deceit, duress, over-reaching, or other ulterior form of constraint or coercion, and should have sufficient knowledge and comprehension of the elements of the subject matter involved as to enable him to make an understanding and enlightened decision. This latter element requires that before the acceptance of an affirmative decision by the experimental subject there should be made known to him the nature, duration, and purpose of the experiment; the method and means by which it is

to be conducted; all inconveniences and hazards reasonably to be expected; and the effects upon his health or person which may possibly come from his participation in the experiment.

The duty and responsibility for ascertaining the quality of the consent rests upon each individual who initiates, directs or engages in the experiment. It is a personal duty and responsibility which may not be delegated to another with impunity.

2. The experiment should be such as to yield fruitful results for the good of society, unprocurable by other methods or means of study, and not random and unnecessary in nature.

3. The experiment should be so designed and based on the results of animal experimentation and a knowledge of the natural history of the disease or other problem under study that the anticipated results will justify the performance of the experiment.

4. The experiment should be so conducted as to avoid all unnecessary physical and mental suffering and injury.

5. No experiment should be conducted where there is an a priori reason to believe that death or disabling injury will occur; except, perhaps, in those experiments where the experimental physicians also serve as subjects.

6. The degree of risk to be taken should never exceed that determined by the humanitarian importance of the problem to be solved by the experiment.

7. Proper preparations should be made and adequate facilities provided to protect the experimental subject against even remote possibilities of injury disability or death.

8. The experiment should be conducted only by scientifically qualified persons. The highest degree of skill and care should be required through all stages of the experiment of those who conduct or engage in the experiment.

> 9. During the course of the experiment the human subject should be at liberty to bring the experiment to an end if he has reached the physical or mental state where continuation of the experiment seems to him to be impossible.
>
> 10. During the course of the experiment the scientist in charge must be prepared to terminate the experiment at any stage, if he has probable cause to believe, in the exercise of the good faith, superior skill and careful judgement required by him that a continuation of the experiment is likely to result in injury, disability, or death to the experimental subject.

Let us make some remarks. Note that the first point says ‘without *any* element of force’, and requires that the participants should comprehend ‘the nature, duration, and purpose of the experiment; the method and means by which it is to be conducted; *all* inconveniences and hazards reasonably to be expected’. Point 2 makes it clear that since participants in medical research often have to suffer, it is unethical to do such research just for the fun of research; some good consequences have to be expected. The requirement of point 3, that animal experimentation should take place before human beings become involved, should of course be seen in the light of the fact that human beings taken from the concentration camps were actually used as if they were guinea pigs. And as previously stated, several scientists considered the situation to be a historically unique research situation, which made it possible to conduct studies without restrictions on what the research subjects were exposed to and what the consequences for them (being severely injured or even killed) should be.

Note that point 8 of the code stresses the scientific qualifications of the researcher. Whereas the other points can be said to denounce squares 2 and 4 of Figure 1, i.e., all unethical research, this point denounces squares 3 and 4, i.e., all methodologically bad research. Left then is only square 1.

The Second World War had quite an impact on the view that science is a field free from ordinary moral considerations. It was not only medicine that was affected. After the development, production, and use of the atomic bomb on Hiroshima and Nagasaki in fall 1945, physicists started to discuss what in fact they should be allowed to do. At the beginning, in 1942, all

physicists involved in the so-called ‘Manhattan project’ were in favor of creating the bomb, since they all strongly feared that the Germans were close to construct such a weapon. But after the German capitulation in May 1945, some of them were strongly against using it. In the aftermath, when it was all over, several physicists ended in a moral quandary.

But back to medicine: what happened after the Medical Case in Nuremberg? In 1948 the newly established World Medical Association updated the Hippocratic Oath in the ‘Declaration of Geneva’ (it starts: ‘I solemnly pledge myself to consecrate my life to the service of humanity’), and in 1949 it put forward an ‘International Code of Medical Ethics’. It kept the ethical issues alive, and inaugurated the Helsinki Declaration of 1964 as well as its many revisions. In point 5 of its present version, this declaration states:

- (5) In medical research on human subjects, considerations related to the well-being of the human subject should take precedence over the interests of science and society.

We will return to the Helsinki declaration, but let us first present the sad fact that the Nuremberg Code did not make all medical researchers aware of the importance of informed consent. Several studies, which ought to have been stopped, were not stopped; and new bad ones were initiated even after the Nuremberg Code had become well known.

One reason why medical scientists did not react properly has been suggested: ‘this was [regarded] a good code for barbarians but unnecessary to ordinary physicians’. Most ordinary physicians and clinical scientists seem to have taken it simply for granted that what they were doing could in no way whatsoever be comparable to the research made in the Nazi concentration camps.

A number of events and scandals in the late 1950s and early 1960s paved the way for a more general research ethical awakening among clinical researchers. In 1966, Henry K. Beecher (1904-1976), a prominent Harvard Medical School anesthesiologist, published in *New England Journal of Medicine* a paper called ‘Ethics and Clinical Research’. Here, he presented a number of medical articles that were based on research involving human beings who had never received information and provided

consent. One such case was the Willowbrook Hepatitis experiments. Here, the researchers deliberately incurred severely mentally retarded and institutionalized children with hepatitis. They wanted to study the natural development of the disease, although gamma globulin was already available and employed as an effective preventive treatment.

At the beginning of the 1960s, the thalidomide (Neurosedyn) disaster became a scientific fact. During 1956 to 1962, around 10,000 children were born with severely malformed limbs and organs, especially phocomelia, because their mothers had eaten thalidomide pills during pregnancy. Thalidomide was used as a tranquilizer, and also against nausea, and many pregnant women used it. The effect of thalidomide (in certain doses) is that it constricts the vessels. In several countries the disaster led to the enactment of new regulations, which prescribed stronger safety tests before pharmaceutical products could receive approval for sale. (By the way: although thalidomide has not been used for several years, it has recently become interesting in the treatment of cancer.)

In 1962 occurred the ‘Jewish Chronic Disease Hospital Case’; so called because it took place at this hospital in Brooklyn, NY. A cancer researcher, Dr. Chester Southam of the Sloan Kettering Institute for Cancer Research, wanted to study whether chronically ill patients with immunodeficiency could reject implanted cancer transplants. He had previously studied both healthy prisoners and cancer patients, and found that they all eventually reject such transplants. In fact, many young doctors who Southam asked for help refused to assist him, but he managed to persuade a Dr. Mandel at the Jewish Chronic Disease Hospital to let a certain number of patients have a suspension of living cancer cells injected into their muscles. Twenty-two patients received such injections; none of them had received information, none had consented, and several of them were not even capable of making their own decisions.

However, when the responsible consultant physician realized that an experiment involving his patients was proceeding without his consent, he became furious. The project was exposed, and both Mandel and Southam were condemned for unprofessional conduct. Since Southam was a well-known cancer researcher, this case is considered extremely important for awakening American doctors and clinical researchers to the problem of informed consent.

Another case of great significance, due to the general publicity it generated, is the so-called US Public Health Service Syphilis Study conducted in Tuskegee, Alabama, 1932-1972. It became a symbol for the exploitation of Afro-Americans in the US. The study was initiated at the end of the 1920s as a part of a special project concerned with the conditions of poor Afro-Americans in the southern states. Originally, it was an intervention project in order to help Afro-Americans suffering from health problems, and it was supported by black institutions and by local black doctors. At that time in this region, the incidence and prevalence of syphilis was very high, and at the outset the ambition was to study the participants without treatment over a six-month period, and then provide the current treatment (Salversan, mercurial ointments, and bismuth). The Rosenwald Fund provided financial support for the planned treatment, and 400 syphilitic and 200 healthy men were recruited. However, the Wall Street Crash of 1929 caused the Rosenwald Fund to withdraw its funding. Inspired by another study of untreated syphilis, which was presented in a scientific journal in 1929 (the Boeck and Bruusgaard Oslo studies), the Tuskegee study was then changed into a purely observational study of the spontaneous course of the disease.

The participants were offered a ‘special treatment’, which was a euphemism for bloodletting and spinal taps, as they were told that they were suffering from ‘bad blood’. In order to convince their patients to participate, the researchers sent out a very misleading letter, see Box 1 below. It illustrates the way in which the authorities informed the participants. (The real reason for letting the participants remain in the hospital was probably that the ‘special treatment’ offered was lumbar puncture.)

Macon County Health Department

Alabama State Board Of Health and US Public Health Service Cooperating With Tuskegee Institute

Dear Sir:

Some time ago you were given a thorough examination and since that time we hope that you have gotten a great deal of treatment for bad blood. You will now be given your last chance to get a second examination. This examination is a very special one and after it is finished you will be given a special treatment if it is believed you are in a condition to stand it.

If you want this special examination and treatment you must meet the nurse at _____ on _____ at _____M. She will bring you to the Tuskegee Institute Hospital for this free treatment. We will be very busy when these examinations and treatments are being given, and will have lots of people to wait on. You will remember that you had to wait for some time when you had your last good examination, and we wish to let you know that because we expect to be so busy it may be necessary for you to remain in the hospital over one night. If this is necessary you will be furnised your meals and a bed, as well the examination and treatment without cost.

REMEMBER THIS IS YOUR LAST CHANCE FOR SPECIAL FREE TREATMENT. BE SURE TO MEET THE NURSE.

Macon County Health Department

Box 1: *Letter to the participants in the Tuskegee study.*

A morally crucial point emerged in 1947, when penicillin became the standard therapy for syphilis. Several programs, sponsored by the US government, were then initiated in order to eradicate syphilis, but the participants in the Tuskegee study were prevented from using penicillin. When later, in 1972, the study was forced to be concluded, only 74 of the original 400 syphilis patients were alive; 28 had died from syphilis and 100 were dead due to related complications; 40 of the participants' wives had been infected, and 19 of their children had been born with congenital syphilis. Nonetheless the goal had been to continue the study until all the participants had died and been autopsied.

In 1968, Dr. Peter Buxtun a venereal disease investigator with the US Public Health Service (Syphilis Study at Tuskegee) became concerned about the ethical aspects of the study, and tried – but in vain – to receive support from his colleagues. He then went to the press, and on July 25, 1972 the *Washington Star* published an article based on Buxtun's information. The following day, it became front-page news in the *New York Times*. Immediately, an ad hoc advisory committee was appointed, the study was terminated, and the surviving participants and their relatives were treated. These events resulted in a change of the composition of the Institutional Review Boards (IRBs) – laypersons became included, and it gave in 1978 rise to the report 'Ethical Principles and Guidelines for the Protection of Human Subjects of Research', often called 'the Belmont Report'. In 1997, on behalf of the United States government, President Clinton officially apologized to the eight still living survivors of the study. Sociological studies have shown that many Afro-Americans distrust US public health authorities, and the Tuskegee study is supposed to be a significant factor behind the low participation of Afro-Americans in clinical trials, in organ donations, and in preventive programs; however, other historical and socio cultural factors, have also been suggested to influence Afro-Americans non-willingness to participate in clinical research.

A structurally similar case (but less known and less devastating) comes from Sweden. The study is concerned with the relationship between the exposure to carbohydrates in the form of candies and toffees and the incidence of tooth decay (caries), and it is referred to as the 'Vipeholm Experiments'. During the period 1946-1951, dental researchers conducted

a series of experiments at an asylum for mentally handicapped people (both young and old) at Vipeholm, situated close to Lund in southern Sweden. At this time around 800 persons were placed at this institution, most of whom would spend the rest of their lives there. All who were able to cooperate were included in these odontological studies. The research was initiated by dental researchers in cooperation with Swedish health care authorities, and it was sponsored by the chocolate and candy industries; the latter did not at the time believe that sweets were associated with tooth decay. Vipeholm was regarded as an ideal place for the research, since the food consumption, also between the meals, was under complete control.

The research subjects were divided in different groups and subgroups and exposed to different doses of toffee that were especially produced for the study. The experimental groups (which received 24 toffees per day over two years) developed much more cavities than the control group, which received no toffees at all. The researchers concluded that the correlation found between exposure to carbohydrates and the frequency of caries was sign of a causal relation, and as a consequence of the study it was publicly recommended that children should be allowed to eat candy only on Saturdays; in Sweden referred to as ‘Saturday sweeties.’ The result was quite good, caries decreased significantly in the next generations. Many similar studies were subsequently made, but this Swedish study is still considered an important starting point for modern preventive dental care. The participants, however, were not informed. This may be excused (or perhaps be made more culpable!) by the fact that they were not considered competent to make their own decisions, but not even their guardians were informed. The patients contracted more caries than normal, and the high level of carbohydrate exposure of those belonging to the 24 toffees/day group may have caused them even other medical complications. However, after the study ended, their teeth were properly treated. For good or for bad, they were afterwards never given candy at all, not even Saturday sweeties.

Another controversial US case was the ‘radiation experiments’, going on between 1942 and 1974, which came to public notice in 1994. It was at first meant to study only to what degree radiations from different radioactive substances could be dangerous to the staff working at the Manhattan project in Los Alamos, New Mexico, during 1943-1945. Later

on, the project was continued by the Atomic Commission; the experiments and observations included:

1) X-ray exposure for chronic diseases such as Rheumatoid Arthritis and cancer, but even healthy individuals were exposed (1942-1944)
2) Plutonium experiments with e.g. terminally ill patients (1945-46)
3) The Tennessee-Vanderbilt Project, where radioactive iron (Fe-59) was provided to pregnant women (1948); result in 1963: four children had developed cancer
4) Therapeutic research concerning X-rays; treating (1950) children with chronic infection of the inner ear (otitis media and otosalpingitis)
5) Experiments with different groups of children, both healthy and mentally handicapped (1950s), where radioactive calcium (Ca-45) was provided in breakfast cereal.
6) Different plutonium, uranium and polonium experiments during 1953-1957.
7) Observation studies on Navajo Indians working in uranium mines in the border of Utah, Colorado and New Mexico (1949); 410 workers out of 4100 had contracted cancer; the researchers had expected only 75.
9) Observation studies (1950s) on the inhabitants of the Marshall Islands (Bikini atolls) of the effect of testing atom bombs 1000 times more powerful than the Hiroshima bomb
10) Experiments on soldiers in the Nevada desert in connection with the testing of atom bombs (1951-1953)
11) Radioactive Iodine (I-131) experiments in Alaska on Eskimos and Indians (1956-1957)
12) Radioactive experiments on prisoners about the genetic effect of radioactivity on genitals (1963-1973).

Many of the participants recruited were terminally ill patients, mentally handicapped individuals, children, pregnant women, ethnic minorities, or prisoners. In most cases, the participants or their parents were not informed at all, or at least not adequately, and no consent had been obtained. When the radiation experiments were disclosed in 1994, a presidential advisory committee was appointed. Two years later it presented a major report,

which from a modern medical-ethical point of view assessed medical research conducted in the past. The committee stressed that the medical community has to acknowledge more or less eternal ethical principles. In line with the Belmont Report, they also underlined the importance of fairness and of not using groups who lack autonomy. If it is a burden to participate in some medical research, this burden should be fairly distributed; and, similarly, if participation in some medical research is an advantage, then this advantage should be distributed fairly.

These are the most well known cases of infractions of the Nuremberg Code. Let us now circumscribe this code by taking a look at some events that occurred before the Second World War.

In Chapter 3.1, we mentioned Edward Jenner's cowpox experiment on the eight year old James Phipps in 1796, which led to the discovery of the smallpox vaccine. Although it is not fair to evaluate events of the past with current ethical principles and values, it is interesting to contrast Jenner's famous and significant experiment with the Nuremberg Code. Jenner transgressed a number of rules. He was inexperienced and had no affiliation with academic research, his research hypothesis was purely based on hearsay, no animal experiments were performed prior to the trials on Phipps, the risk of exposure from cowpox in children was not known, Jenner did not know whether or not the extract was purified or had been polluted by other agents. Finally, Jenner used a minor as experimental subject, a child whose father was a gardener at Jenner's household and who might have been dependent on Jenner in such a way that he was unable to withhold the boy's participation. Nevertheless, Jenner eventually (after his academic colleagues had succeeded in purifying the cowpox substance) convinced his initially reluctant colleagues that his technology functioned. In fact, Jenner's research has to be placed in square 4 of Figure 1. But it is interesting in that it shows that even initially methodologically bad research may hit upon truths and in the end result in good consequences.

The Norwegian Gerhard Armauer Hansen (1841-1912) is well known for his discovery of the leprosy bacillus in 1873. In retrospect one may say that he managed to demonstrate that leprosy was an infectious disease and not God's punishment for a sinful life, which was the common opinion at the time. For a long time, however, he had problems in proving his

hypothesis. In order to do this, he tried to transmit bacteria from diseased individuals to individuals who did not suffer from leprosy. First he tried on animals, but when this was not successful, he tried to transmit leprosy bacilli into the eyes a woman. For this he was in 1880 charged and convicted by the court in Bergen. One reason why Hansen was so eager to attain immediate proof of his hypothesis was the fact that he was competing with a German colleague, Albert Neisser (1855-1916), who in 1879 had visited Hansen in Bergen and received leprosy material. Back in Breslau, Neisser succeeded in identifying the bacteria using a special color technique; he published his results and claimed, a bit astonishingly, complete priority for the discovery of the cause of leprosy.

Neisser also had his own ethical problems. In 1898, he tried to develop a new cure for syphilis. The syphilis spirochete was not yet identified, but serum therapy was applied by several of his contemporary colleagues. This is the case of the Danish physician and Nobel Prize winner (1926) Johannes Fibiger (1867-1928) a randomized study regarding serum treatment of diphtheritic patients, which is sometimes regarded as the first randomized trial in medical history. Serum treatment was even used as a preventive measure, and Neisser thought that by providing human beings with serum from patients with syphilis, he would be able to prevent healthy individuals from contracting by syphilis. Therefore, he provided a cohort of non-syphilitic female prostitutes with serum from patients suffering from syphilis, but he found that the treatment was not an effective preventive measure; instead, several of the participants became infected. Neisser, though, claimed that the prostitutes had not contracted syphilis through his experiment, but in their activities as prostitutes. Although Neisser received support from almost all his colleagues – apart from a psychiatrist Albert Moll (1862-1939), one of the founders of modern sexology – his experiment caused public debate. Moll wrote a book in which he presented a long list of unethical experiments, and eventually the Prussian minister of education presented some guidelines for how to use human beings in medical research.

Around the turn of the nineteenth century, beriberi meant weak muscles, including the heart muscle; the consequence of untreated beriberi was often that the heart stopped functioning. In 1905-06, William Fletcher conducted beriberi experiments at a lunatic asylum in Kuala Lumpur. The theory of

vitamin (‘vital amines’) deficiency was not yet established, and some of Fletcher’s colleagues were convinced – quite in accordance with the microbiological paradigm (Chapter 2.5) – that beriberi is an infectious disease. Fletcher divided the patients into two groups, which were given different diets; one group received cured (brown) rice and the other uncured (white) rice. Amongst the 120 patients who received uncured rice, there were 43 cases of beriberi and 18 deaths; of the 123 patients who were given cured rice, there were only two cases of beriberi (which already existed up on admission) and no deaths. Fletcher also changed the study design so that ten of the participants (or ‘lunatics’ as he refers to them) suffering from beriberi were then placed on a diet of cured rice – and all of them recovered; of the 26 who were not put on the cured rice diet, 18 died. In his report, published in *The Lancet* in 1907, Fletcher stressed that he did not believe that a certain kind of rice might be the cause of the disease, but that a ‘proteid matter’ in the rice was a possible cause. Another hypothesis launched by Fletcher (in accordance with the microbiological paradigm) was that a lack of something in the uncured rice made it easy for external agents like bacteria or protozoa to infect the individual and thus cause beriberi. Even though Fletcher’s hypothesis was erroneous, his clinical study was the first empirical one that indicated the existence of deficiency diseases; and the vitamin concept was finally established in 1912. It is striking that Fletcher described and seem to have treated the participants as if they were nothing but laboratory animals.

In 1983 the German ethicist Hans-Martin Sass paid attention to the so-called ‘Reichsrundschreiben’ or ‘Reichsrichtlinien’ from 1931, some pre-Nuremberg German regulations concerned with new therapies and human experimentation. The reason for the creation of these Richtlinien (guidelines) was some clinical research about the effect of tuberculosis vaccine. In these trials, which included even children, many persons died. A trial conducted in Lübeck is especially well known, where 75 children died in a pediatric tuberculosis vaccine experiment. Even though the Reichsrichtlinien were put forward in the Weimar Republic, they were not considered when the Nazis got into power (1933). It is an ironical historical fact that these Reichsrichtlinien from 1931 are more comprehensive and restrictive than the Nuremberg Code. If the Reichsrichtlinien had been known and considered during the Nuremberg

medical trial when Telford Taylor presented his prosecution speech, the Nuremberg Code might have been rendered superfluous.

10.4 The Helsinki Declarations and research ethics committees

As we have already made clear, the Helsinki Declaration of 1964 is the result of a long development and reaction on the Medical Case in the Nuremberg trial, which after the Second World War was (and still is) promoted by the World Medical Association. The original Helsinki Declaration from 1964 contains 14 paragraphs; in 1975 they were expanded into 22, and after the latest revision, in 2000, the declaration contains 32 points divided into three sections: (A) 'Introduction', (B) 'Basic principles for all medical research', and (C) 'Additional principles for medical research combined with medical care'. Here we find all the things we have already spoken of: the requirement of informed consent, a stress on the competence of the scientists, and that there should be harm and risk assessments of the study. We will now quote three paragraphs; point 8 because it underlines what is meant by vulnerable groups, point 12 because it makes animals into moral objects, and point 13 because here the aims of the research ethics committees are stated:

- (8) Medical research is subject to ethical standards that promote respect for all human beings and protect their health and rights. Some research populations are vulnerable and need special protection. The particular needs of the economically and medically disadvantaged must be recognized. Special attention is also required for those who cannot give or refuse consent for themselves, for those who may be subject to giving consent under duress, for those who will not benefit personally from the research and for those for whom the research is combined with care.

- (12) Appropriate caution must be exercised in the conduct of research which may affect the environment, and the welfare of animals used for research must be respected. [Our comment: note

that it is ‘the welfare’ of the animals, not ‘the autonomy’, which should be respected.]

- (13) The design and performance of each experimental procedure involving human subjects should be clearly formulated in an experimental protocol. This protocol should be submitted for consideration, comment, guidance, and where appropriate, approval to a specially appointed ethical review committee, which must be independent of the investigator, the sponsor or any other kind of undue influence. This independent committee should be in conformity with the laws and regulations of the country in which the research experiment is performed. The committee has the right to monitor ongoing trials. The researcher has the obligation to provide monitoring information to the committee, especially any serious adverse events. The researcher should also submit to the committee, for review, information regarding funding, sponsors, institutional affiliations, other potential conflicts of interest and incentives for subjects.

The explicit idea of research ethics committees (REC) were first conceived in the Tokyo revisions of 1975 (sometimes called the second Helsinki Declaration). However, even before the Helsinki Declarations there existed in several countries a kind of research ethics committees, called Institutional Review Boards, but they were almost solely composed of researchers representing different medical specialties. Modern research ethics committees include representatives of the general public. In several countries today, half of the committees consist of people representing the research community and the other half consists of people who represent the general public or society at large; mostly, the total number is 10-15 persons. One reason for including representatives from the general public is of course the fact that earlier in history the research community has taken a bit lightly on ethical issues. But let us add some more words about the RECs.

Research Ethics Committees (RECs) are committees convened to provide independent advice to participants, researchers, sponsors, employers, and professionals on the extent to which proposals for research

studies comply with recognized ethical standards. The purpose of a REC in reviewing the proposed study is to protect the dignity, rights, safety, and well-being of all actual or potential research participants. RECs are responsible for acting primarily in the interest of potential research participants and concerned communities, but they should also take into account the interests, needs, and safety of researchers who are trying to undertake research of good quality.

RECs also need to take into consideration the principle of justice. This requires that the benefits and burdens of research be distributed fairly among all groups and classes in society; in particular, taking into account gender, economic status, culture, and ethnic considerations. In this context the contribution of previous research participants should also be recalled.

One of the reasons for the latest revision of the Declaration was to secure that the population of poor countries do not become exploited by medical researchers. The following statement is addressing this issue:

- (19) Medical research is only justified if there is a reasonable likelihood that the populations in which the research is carried out stand to benefit from the results of the research.

This statement, as well as the one below (30), should be seen in the light of some unethical trials conducted in Africa and Asia regarding the treatment and prevention of HIV.

- (30) At the conclusion of the study, every patient entered into the study should be assured of access to the best proven prophylactic, diagnostic and therapeutic methods identified by the study.

Accordingly, a study where it is not possible to make the treatment available to the participants after the study is completed, should not be conducted in the first place. The demands for information provided to participants have been made more specific:

> (22) In any research on human beings, each potential subject must be adequately informed of the aims, methods, sources of funding, any possible conflicts of interest, institutional affiliations of the researcher,

> the anticipated benefits and potential risks of the study and the discomfort it may entail. The subject should be informed of the right to abstain from participation in the study or to withdraw consent to participate at any time without reprisal. After ensuring that the subject has understood the information, the physician should then obtain the subject's freely-given informed consent, preferably in writing. If the consent cannot be obtained in writing, the non-written consent must be formally documented and witnessed.

It is also said that when persons who are not competent to give consent are included in medical research, it is especially important that they, or the group to which they belong, will benefit from the research. Furthermore, the special reasons for including such participants shall be transmitted to the research ethics committee.

In what follows we will make further comments on the contemporary content of informed consent. It is comprised of two components: the information component and the consent component.

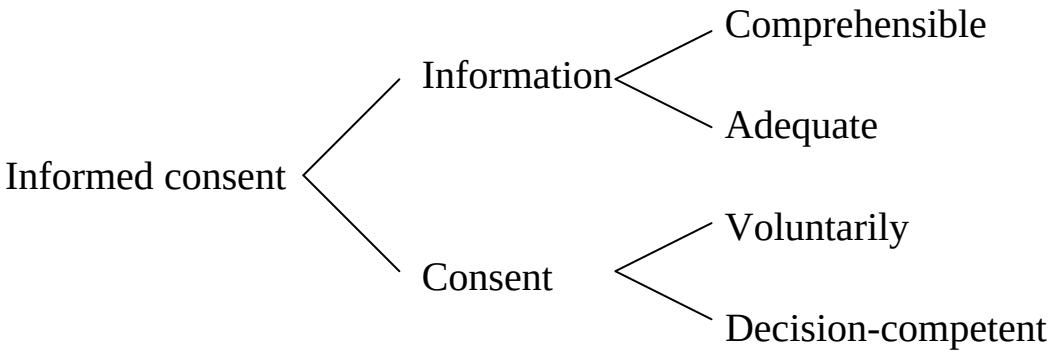


Figure 4: *The structure and content of informed consent.*

The information should be comprehensible, which means that plain and straightforward language should be used and difficult words and technical terms avoided. Studies about participants' comprehension of terms such as 'randomization' and 'randomized trial' indicate that these issues are delicate tasks; and that participants sometimes do not understand all aspects of a study they are being told about. Whether this is due to poor memory or to poor communication skills is hard to say. But it has also been shown that if the participants receive both oral and written

information, they understand the whole thing better. However, sometimes researchers meet participants who do not want to receive information, either orally or in written form. The reason might be that they simply do not want to face the pros and cons of an upcoming treatment; especially if the disease is serious, e.g., cancer. Researchers have to respect even such wishes, since the overarching principle is the principle of respect for autonomy. Even though the patients do not want to be informed, they might nonetheless very well wish to consent. In such cases there is literally not informed consent, but there is voluntary consent after having been offered information.

Information should also be adequate, which means that the information should be in accordance with the study protocol as well as with the Declaration of Helsinki, which specifies information about the 1) aims, 2) methods, 3) sources of funding, 4) any possible conflicts of interest, 5) institutional affiliations of the researcher, 6) anticipated benefits and potential risks of the study and the discomfort it may entail. The subject should be informed of 7) the right to abstain from participation in the study or 8) to withdraw consent to participate at any time without reprisal.

The second part of informed consent concerns the consent component, and it says (first) that consent or refusal should be provided voluntarily, which means that no dependence or pressure may occur from the researcher. When recruiting participants, it might be argued that the payment offered is too high, since it might then be psychologically impossible for a very poor person to refuse participating in a very risky investigation. Some kind of payment seems reasonable, but it is then important that it be neither too low nor too high. The same kind of reasoning is applicable in relation to various kinds of organ donation.

Also (second), the participants are supposed to be truly decision-competent. This does not imply that they need to have a Ph.D. in philosophy or medical science. It means that a person of normal intelligence is, based on the adequate information provided, supposed to be able to make a rational decision as to whether or not to participate. However, if a potential participant is not capable of understanding the information or of making a real decision, this does not necessarily exclude him from participation. Fairness has also to be considered. It cannot be fair that some groups such as small children and psychotic patients cannot have

their diseases and conditions investigated. In such situations, as stated in the Helsinki Declaration, surrogate decisions and proxy consent can be accepted.

To sum up, good clinical trial practice implies that both design and other methodological aspects are optimal, and that the study is aligned with current research ethical guidelines. As many safety regulations in factories and in society at large have their origin in some actual accidents, many of the stages towards today's research ethical codes and declarations have their origin, as we have seen, in some specific cases of morally questionable, sometimes morally outrageous, research. The list below gives an overview of the important stages in the development of medical research ethics in the Western world that we have described. What during history in this respect has happened in other corners of the world, we do at the moment not know much about.

Event	Research ethical reaction
Edward Jenner's cowpox experiment (1796)	No reaction at the time
Controversies between pharmacists, surgeons, and medical doctors in Manchester (1799)	Percival's guidelines 1803
Gerhard Armauer Hansen's leprosy experiments (1878)	Trial at the civil court in Bergen 1880
Albert Neisser's syphilis experiments (1898)	Public reaction and directive from the Prussian Minister 1900; the writings of Moll 1902.
William Fletcher's beriberi experiments (1905-06)	No reaction at the time
Tuberculosis vaccination experiments (Lübeck) including children of whom many died (1927)	Guidelines by Reichgesundheitsrat 1931 (Reichsrichtlinien)

The medical experiments in the Nazi-concentration camps (1942-43)	Nuremberg Code 1947; Declaration of Geneva 1948, World Medical Association Int.; Code of Medical Ethics 1949; The Helsinki Declaration 1964
Carbohydrate experiments in the Vipeholm studies of caries (1948-52)	Some reactions at the time – mainly discussed during the 1990s
The Willowbrook hepatitis experiments in the 1950s	Henry Beecher's article 1965 – also discussed in the 1990s
The Jewish Chronic Disease case (1962)	Public debate of researchers' responsibility; Helsinki Declaration 1964
Experimentation in US prisons (1944-1974)	Special guidelines 1973
Tuskegee syphilis study (1932-72)	Belmont Report 1978 and Helsinki Declaration 1975; President Clinton's apology 1998
Radiation experiment in the US (1943-74)	No reaction at the time – in 1996 an advisory committee presented a report
HIV- vaccination and treatment experiments in Africa and in Asia 1997	Revision of Helsinki declaration 2000.

10.5 Between cowboy ethics and scout morals: the CUDOS norms

In Chapter 10.3 we mentioned in passing Albert Neisser's attempt to achieve priority for the discovery of the leprosy bacilli without adequately referring to the works of Gerhard A. Hansen. Now we will say a little more about problems of this kind. History has much to tell about how far researchers have been prepared to go in dishonest and deceitful practice in order to promote their own ideas, to persuade skeptical colleagues, to

achieve status, and to be the first to present an invention or a discovery. How far the norms of the medical community should allow them to go, is a question of importance for the future.

In some of the stories told above (and in some to be told below), the researchers seem to have transformed the cowboy saying ‘shoot first, ask questions later’ into: deceive first, answer possible questions later. This contradicts the Helsinki Declaration. On the other hand, there seems to be cases in medical research that underline the view we defended in the preceding chapter: all norms seem to require exceptions. The scout law that says ‘obey orders first, ask for reasons only afterwards’ (Baden-Powell’s original seventh law) should probably not be transformed into a corresponding deontological law for research that says: obey your scientific community first, ask for reasons only afterwards. Modern virtue ethics exists in-between this kind of cowboy ethics and this kind of scout morals. (Let us within parenthesis present the original eighth scout law with ‘researcher’ substituted for ‘scout’. It gives us: ‘A researcher smiles and whistles under all circumstances. Researchers never grouse at hardships, nor whine at each other, nor swear when put out.’ Something for research groups to think about?)

We have earlier talked about misconduct in order to make the concept of ‘scientific fact’ clear; here, we will talk about it in order to shed light on research ethics. In Chapter 3.1 (‘Deceit and ideological pressure’) we mentioned some cases of pure fraud: Burt and his monozygotic twin studies, and Schön and the single-molecule transistor. We also mentioned the less clear case of Banting, Macleod, and the discovery of insulin. We will now tell the last story in more detail.

The relatively inexperienced rural physician Frederic Banting and the prominent Toronto professor of physiology John Macleod received in 1923 the Nobel Prize for the discovery, isolation, and purification of insulin. Banting, who initially had received no external recognition, became angry that his assistant, medical student Charles Best, was not also awarded, and decided to share his part of the Prize with Best. The head of department, Macleod, who had since long enjoyed good reputation, decided to share his part with James Collip, a chemist and experienced academic researcher who had been responsible for purifying and testing the insulin. In

hindsight, it seems reasonable to assume that all four were necessary for the breakthrough.

At the outset, it was Banting who contacted Macleod. Banting had read an article suggesting that obstruction of the external function of the pancreas might cause the death of all pancreas cells apart from those responsible for internal secretion. Banting suggested to Macleod that it would be possible by surgery to bring about the same condition and thus to study the internal secretion more closely. Macleod was first reluctant towards Banting's idea, but he eventually accepted it, and in the summer 1921 he provided Banting with a room, dogs to experiment with, and an assistant, Charles Best, who was a medical student skilled when it came to measuring the concentration of sugar in urine and blood.

Banting and Best conducted a series of studies during the summer 1921; Banting operated on the dogs initially in order to stop the external function of the pancreas, and to isolate the internal secretion. Later on this isolated substance was provided to dogs that Banting had made diabetic. This was very hard, but they managed to simplify the strategy, and the process of extracting insulin was facilitated. During the summer, Macleod went on holiday, not returning until October 1921, but then he discovered that something interesting was going on. When Banting thought that they had problems and perhaps had come to a dead end, Macleod helped them to change direction of the investigation and brought the chemist James Collip into the project. Collip found out how to administer insulin, what doses were needed, and the effect on the liver of insulin treatment.

In Banting and Best's first article (published in '*The Journal of Laboratory and Clinical Medicine*' January 1922), they reported that the effect of adding an extract of the internal secretion of the pancreas to diabetic animals (dogs) had been positive in all 75 experiments conducted. Afterwards, historians have scrutinized the protocols and found that out of the 75 experiments, only 42 were actually positive; 22 were classified as negative, and 11 were referred to as 'inconclusive'. Although neither Banting nor Best knew that they were actually in competition with several other researchers; one of whom later claimed priority in the discovery of insulin, the Romanian physiologist Nicolae Paulesco (1869-1931). Paulesco had already in August 1921 published a paper in *Archives Internationales de Physiologie*, a Belgian scientific journal, and today after

the events, many scientists find that Paulesco should have had his share of the Nobel Prize.

The whole idea of the existence of hormones as the result of the internal hormone secretion of the thyroidal and testicular glands was, however, already acknowledged. Therefore, it was possible to understand the internal secretion of the pancreas in analogy with these glands; accordingly, Banting and Bests' discovery were not exposed to paradigmatic resistance.

The question is whether we should characterize the presentation of the results by Banting and Best in their first publication in terms of normal scientific activity, poor science, or pure fraud, i.e., as intentional distortion of results. Why did they not simply report the results that were stated in their protocols? Probably, they were themselves convinced that they were right, but were afraid that the results were not sufficiently convincing to others. Worth noting is the fact that Macleod was not a co-author. Does this indicate that he did not want to put his name on the paper? At this time, the leader of a department usually placed his name as the last one in order to inform editors that the paper had his approval. But perhaps Macleod felt that he had not contributed enough to be reckoned as a co-author; or that he had no control over the experiments, since he had not participated in these. What is at stake is the behavior of two young researchers, Dr. Banting, rather inexperienced as a researcher, and the medical student Best, who is not supposed to know how accurately the results needed to be presented. Was the presentation deliberately deceitful in order to convince potential reluctant colleagues, or was it merely the result of a lack of experience in writing scientific papers and presenting data in a proper manner? Focusing only on the consequences, and looking beyond their debatable manner of conducting research, they actually discovered a treatment which saved the lives of millions of suffering patients.

Since several decades, most medical researchers are well aware of the existence of scientific misconduct, but such awareness has been fluctuating. During most of the twentieth century, it was more or less taken for granted that fraud in science is a rare exception. When, in 1983, W. Broad and N. Wade published the path-breaking book *Betrayers of the Truth. Fraud and Deceit in the Halls of Science*, many readers were quite

astonished about how many examples of misconduct the authors could present. However, as the authors themselves made clear, their message had been voiced before. Already in 1830 appeared a book called *Reflections on the Decline of Science in England* which made the same general points. The author was the mathematician Charles Babbage (1791-1871); today mostly remembered as inventor of some computing engines and as a very important fore-runner of computer science.

Babbage found reasons to divide scientific misconduct into three different categories; in increasing degree of misconduct they are:

- trimming (e.g., taking outliers away)
- cooking (e.g., only mentioning positive observations)
- forging (e.g., inventing data).

Babbage's classification still makes good sense. One thing to be added, though, is:

- plagiarism (e.g., stealing of data, hypotheses, or methods; some the sub-forms have been dubbed 'citation amnesia', 'the disregard syndrome', and 'bibliographic negligence').

In 1989, it should be noted, the US National Institutes of Health created a special office called 'The Office of Scientific Integrity (ISO)'. Many countries have followed. Nowadays, there are many official 'watchdogs of science'. Apart from the four factors above, one talks also of 'Other Serious Deviations from accepted practices' and calls it

- the OSD clause.

Sometimes the demarcation line between normal scientific practice and misconduct is quite well-defined, but sometimes it is hazy. Here come some questions. Is it fraud or normal scientific practice: 1) to use short citations without acknowledging the source? 2) to exclude seemingly accidental single data that are not in accordance with the hypothesis? 3) not to mention literature which report observations that contradicts one's own empirical findings? 4) to apply advanced statistical methods to trivial results in order to make them appear more scientific? 5) to choose an

inappropriate statistical method in order to achieve (seemingly) significant results? 6) to retouch an electron microscopic picture for what is truly believed to be an artifact?, 7) to let the chief of the department, with high external recognition, present results which were actually produced by an unknown young colleague?, and, as the chief of a department, 8) to use your authority to suppress critics of a certain study? 9) to be listed as co-author of a paper when you have only contributed by recruiting patients or by registering data? 10) to promise - as incitement for promptly recruiting patients in a clinical trial - the doctor who recruits the most patients to be ranked as the first author? 11) to let a colleague be a co-author despite the fact that he has not contributed to the paper, if he, in return, lists you as a co-author on his paper, to which you have not contributed either? We leave the questions unanswered.

A common denominator in the definitions of ‘misconduct’ is that it contains an *intentional* distortion of the research process. Some definitions, however, do not stress the intention due to the fact that it might be difficult to reveal or make evident an intention to deceive. For instance, it has been argued that misconducting scientists are suffering from psychiatric personality disorders, and if such individuals are excluded from the scientific community the problem would be solved. The common opinion, however, seems to be that fraud is a common feature in our society seen in all sectors. There are insider and corruption issues in the commercial and financial world, pederasty among catholic priests, war crimes in the military, corruption within sports, nepotism and corruption among politicians and so forth and so on; and there seems to be no general reason why the scientific community should be free from things such as these.

Here comes an illustration of what conflict of interests might mean. In a Lancet paper and a subsequent press-conference in 1998, the gastro-enterologist Andrew Wakefield and his collaborators presented results that supported the hypothesis that the triple vaccine (against measles, mumps, and rubella) caused autism. It had an enormous impact on parents’ decisions about whether or not to provide their children with this vaccination. The worldwide vaccination rate decreased with approximately ten percent (but most in UK); and there was a threatening risk of developing an epidemic of the diseases. The paper was retracted in 2004, when it had become evident both that linking the triple vaccine with autism

was a result of severely selected data and that Wakefield and his research group was financially supported (more than £400.000) by a lawyer who – on behalf of parents with children suffering from autism – was trying to prove that the vaccine was unsafe. Wakefield had actually been recruited in order to support a lawsuit against the vaccine manufacturers. Later, the Editor-in-Chief of the Lancet, Richard Horton, stated that if this conflict of interest had been disclosed when the paper was submitted, it would never have been published. It was a journalist at Sunday Times (Brian Deer) who revealed the story.

We have given much space to the lucky case of Banting-Best-Macleod-Collip and their trimming and cooking, but this does not mean that they are very special. Historians of science have found that some of the truly great heroes of the scientific revolution – e.g., Galilei and Newton in physics, Dalton in chemistry, and Mendel in genetics – trimmed the data they reported (Broad and Wade). The leading hypothesis in this kind of historical investigation has been the view that when empirical results look perfect, one might suspect misconduct. We now know very well from the empirical sciences, that some statistical variation in observational data is the rule. It is often statisticians who disclose misconduct of the being-too-perfect kind. Today, in fact, some researchers who arrive at results which are extremely good hesitate to publishing them.

Why do some scientists trim, cook, forge, and plagiarize? Looking at the known cases in the light of general facts about human nature, it is rather easy to come up with a list of possible – and very human – reasons. We have produced such a list below. It should be noted that people may cheat both for high purposes and for very egoistic reasons. Even a real passion for truth and knowledge may cause misconduct, i.e., the scientist in question believes that he has found the truth, but also that he cannot without misconduct make other persons see the presumed truth. And then he lets the means (cheating) sanctify the end (make truth visible). All the six kinds of human ‘hunger’ that we list can very well exist at one and the same time; a scientist may tend towards misconduct for more than one reason at a time. Here is our list:

Believing that one knows the right answer:

- passion for truth as such; truth-hunger
- thinking that for altruistic reasons truth is important; norm-hunger

Promoting one's egoistic interests:

- to be first with results in order to become famous; fame-hunger
- to obtain a better paid position; money-hunger
- to obtain a more powerful position; power-hunger
- to become part of a desired group (institution, faculty, etc.); community-hunger

One further factor, but not an original motivation, is of course the belief on part of the cheating person that he can get away with his fraud. Now one may ask: since fraud in science is bad for science, what can we do, apart from making it hard to get away with fraud? Since all the interests listed might be regarded as being part of human nature, but being differently distributed in different persons, should the scientific community make psychological tests of prospective researchers? Our negative answer and its reasons will come in due course. Let us make a quotation before we present our reflections. It comes from the prominent Hungarian researcher and Nobel Prize winner (1937; for works around vitamin C) Albert Szent-Györgyi (1893-1986), who in 1961 in a speech at an international medical congress, said:

> The desire to alleviate suffering is of small value in research – such a person should be advised to work for a charity. Research wants egotists, damned egotists, who seek their own pleasure and satisfaction, but find it in solving the puzzles of nature.

What is the message of these quotes? Behind Szent-Györgyi's view, one might guess there is a line of thought analogous to that made famous by the Scottish moral philosopher and political economist Adam Smith (1723-1790). The latter claimed that in the long run everyone would benefit if everyone would act egoistically in the market (but, by no means, in the family or in friendship circles!). The market is 'an invisible hand', which, so to speak, transforms egoism into altruism. It does not, though, house any

war of all against all, or egoism pure. In order to install the kind of competition wanted by Smith, everyone has to subdue his egoism to laws for property rights (to material as well as to intellectual property); to laws for contract keeping; to patents; and to the laws regulating money and bank transactions. The egoism advertised is an egoism that functions within a normative network.

Like Szent-Györgyi, we think that something similar can be said about science, but only if there is something in science that functions somewhat as the norms for the economic market do. As it happens, we think that the CUDOS-norms, which we will soon present, can function this way; especially the norm of organized skepticism. One might then even take a step further than Szent-Györgyi (who mentions only the pleasures of puzzle-solving), and claim that as long as the fame-hunger, money-hunger, power-hunger, and community-hunger of scientists can be satisfied only by doing good and ethical research, then even science can accept the private egoism of the researchers. Science can, if it has the right normative network, function as something that transforms the egoisms of the individual researchers into an almost altruistic aim: the growth of common knowledge. The aim might be called ‘altruistic’, since on the basis of new knowledge new useful technological devices and new useful medical therapies can be invented.

After having worked for several years with sociological questions related to science, in 1942 the American sociologist Robert Merton (1910-2003) presented a paper in which he claimed that scientific communities can be sociologically defined. They are held together by four norms: ‘Communism’, ‘Universalism’, ‘Disinterestedness’, and ‘Organized Skepticism’. As a name for all of them together, he created the acronym ‘CUDOS norms’. He thought these norms were good norms; he did not intend his sociological discovery to lead to any transformation of classical science. To the contrary, his thoughts were also influenced by what had taken place in the Nazi influenced scientific community; especially the denunciation of Einstein’s theories as ‘Jewish physics’.

Merton wrote his article a couple of decades before the late twentieth century criticism of science departed, but another sociologist of knowledge, John Ziman (1925-2005), has advertised the same acronym after the post-modern turn in mainstream sociology of knowledge. It is,

however, a use with a difference. Keeping the acronym, Ziman expands the four principles into five, and gives some of the old ones a partly new and weaker content. He calls his principles: ‘Communalism’, ‘Universalism’, ‘Disinterestedness’, ‘Originality’, and ‘Skepticism’. When he presents them (Ziman 2000), he also defends epistemological views that we endorse. For instance, he claims that “science is a genuine amalgam of ‘construction’ and ‘discovery’ (p. 236)”, and he also states:

> What I am really saying is that the post-modern critique effectively demolishes the Legend [that science is a ‘method’ of guaranteed, unassailable competence], along with the more general philosophical structures that support it. That does not mean that we should go to the opposite extreme of a purely anarchic or existential philosophy where everything is ‘socially relative’, and ‘anything goes’. ‘Negativism’ is not the only alternative to undue ‘positivism’! (p. 327)

We will now simultaneously present and comment on Merton’s and Ziman’s principles. We find the tension between them illuminating, and we would under the name of ‘CUDOS norms’ like to retain some more Merton-like views within Ziman’s framework. It seems to us as if also researchers and editors like Marcia Angel and Sheldon Krimsky want to actualize the CUDOS norms again.

- Communism and Communalism:

Knowledge is an intellectual achievement. Therefore, this question arises: *should scientists be allowed to acquire intellectual property rights for their discoveries in about the same way as authors can acquire copyrights and inventors can acquire patents?*

With the rapid development within the life sciences and computer technology, problems around intellectual property have become quite a significant public issue. Since 1967 the United Nations shelters an agency called ‘World Intellectual Property Organization’ (WIPO). To Merton, ‘communism’ simply means common ownership; nothing more. According to him, all knowledge should always be common property. This implies, among other things, that everyone should be allowed to use new

knowledge at once. The discoverers of new knowledge have either to rest content with due recognition and esteem for their work or, on its basis, to try to make an invention and patent it before others do it. Ziman's communalism is weaker. The essence of it, says Ziman, is "its prohibition *of secrecy* (p. 35)". But such a prohibition is quite consistent with the existence of private intellectual property rights for scientific discoveries. Let us make just a few observations and remarks, and then leave the future discussion to the readers.

It seems odd that a scientist, or a group of scientists, should be allowed to own overarching theories and laws, for instance, that Copernicus for some time could have owned the theory that is called 'Copernicus' heliocentric theory, that Newton could have owned his laws of motion, Einstein his relativity theories, Mendeleev the periodic table of the elements, Bohr his model of the atom, Harvey his views about the heart and the blood circulation, and Crick and Watson their model of the DNA molecule. But it seems less odd to claim that discoverers of small knowledge pieces relevant for technology, e.g., about how a certain molecule is constituted, might be given intellectual property rights *for some time*. Currently, scientists are in some areas allowed to take out patent on their discoveries; such patents are usually running for 20 years. Probably, the norm ought to be: *communism for some kinds of scientific discoveries and communalism for some*.

It might be argued that knowledge should always be for the common good, and that, therefore, it should always be common property. But this is a hasty conclusion. Even copyrights and patent laws have been construed with the common good in mind. However, even if there are and ought to be intellectual property rights in some parts of science, nothing can of course stop scientists from making (to use an expression from Chapter 9.1) 'supererogatory good actions'. Banting, whom we have earlier commented on critically, should in our opinion really be credited for having sold his patent on insulin for one dollar. Why did he do this? Answer: in order to help diabetic patients.

The C-norm (the norm of some-communism-and-some-communalism), is in one sense a moral norm, since it is based on considerations on what is right, fair, and good for society at large. This notwithstanding, it is also an

overarching methodological norm, since it promotes knowledge, which is the utmost goal of all methodological norms.

- Universalism (in relation to persons and ideas):

Ziman writes: "The norm of *universalism* […] requires that contributions to science should not be excluded because of race, nationality, religion, social status, or other irrelevant criteria. Observe that this norm applies to *persons* [in their role as scientists], not to *ideas*. […] Thus *gender* bias is firmly forbidden. […] However elitist and self-serving it may be to outsiders, the scientific community is enjoined to be democratic and fair to its own members (pp. 36-37)."

Merton's universalism, on the other hand, did not stress persons but how scientific ideas should be tested. According to his version, universalism means that presumed research results shall be evaluated by means of universal and impersonal criteria. There is no such thing as 'German physics'; apart, of course, from physics done – well or bad – by Germans. From Merton's criteria-centered universalism, Ziman's person-centered universalism can be derived (since if there are impersonal criteria, then criteria such as race, gender, and ethnicity are superfluous), but from Ziman's universalism Merton's cannot be derived; therefore, they are not identical. The reason behind Ziman's weakening of the principle of universalism is all the problems connected with perceptual structuring, the necessity of auxiliary hypotheses and the very notion of 'paradigms', which we have presented in Chapters 2.4-2.5 and 3.2-3.5.

To put it briefly, all basic criteria seem to be paradigm-bound or paradigm-impregnated, and therefore not truly universal and impersonal; and in this we agree. Nonetheless we think that Ziman has weakened Merton's principle too much. From the point of view that we have defended – fallibilism-and-truthlikeness plus default methodological rules – it makes good sense to say that hypotheses should be impersonally tested even though there are no infallible overarching criteria.

Since all real and presumed knowledge is *about* something (be this something either nature as it is independently of man, social reality as it has been created by man, or both at once), it is the relationship between the hypothesis and what it is about that is at issue, not the relationship between the hypothesis and the social classification of its creator. It is only in

sociology of knowledge that the latter relationship is of interest, but then the relationship between the sociologist himself and his hypothesis is left out of account; or it has to be studied elsewhere (which leaves the new researcher's social position out of account; and so on). Of course, every test in every science is made by socially specific people and takes place under certain given specific social circumstances, but the tests should nonetheless aim at finding truthlikeness. And truthlikeness (Chapter 3.5) is a *relation* between a hypothesis and what the hypothesis is presumed to be about; not the *expression* of a culture or sub-culture the way angry words might be an expression of a personal state of anger. One has to keep one's tongue in check here, because even though a *theory as such* can be (and often is) the expression of a culture, the *truthlikeness of the theory* cannot be such an expression, since a true theory (in contradistinction to a novel) cannot create the objects it is about. This is quite consistent with the fact that an *interest in truthlikeness* may be promoted in some cultures and disavowed in others.

The norm of universalism can in its relations to persons be reckoned a moral norm, since it says that it is wrong to let social background play any role in science, but in its relations to ideas it is an overarching methodological norm.

- Disinterestedness (in evaluations of research):

According to the traditional mid-twentieth century picture of the scientist, he should in the whole of his research (but not in his private life!) detach himself from his emotions in order to keep his mind open and let his reason work freely. The only emotional attitudes allowed in research should be a passion for knowledge combined with humbleness in front of truth. Research should be free of external and ideological interests whether political, financial, religious etc., and on the personal level it should be free from emotions. Even though perhaps most scientists in the 1930s produced a less emotional kind of public image than what is usual among scientists today, and that therefore there was something to Merton's observations, he did exaggerate these things. After all what history and sociology of science has now taught us, one simply has to write as Ziman does:

> The notion that academic scientists have to be *humble* and *disinterested* seems to contradict all our impressions of the research world. Scientists are usually passionate advocates of their own ideas, and often extremely vain. Once again, this [disinterestedness] is a norm that really only applies to the way that they present themselves and their work in formal scientific settings. In particular, it governs the style and approach of formal scientific communications, written or spoken (p. 38).

However, again we would like to draw Ziman a bit more in the direction of Merton. But let us first make our agreement with Ziman clear. In everyday life it is good not to quarrel all the time even if there are conflicts; and some degree of politeness usually makes the overall business of life easier. In research, we think that the disinterested style of communication has the same advantages and should be adhered to. It makes it easier to apprehend the opponents' views.

Now some words about a harder kind of disinterestedness. Even though no researcher is supposed to give up his research hypothesis too easily, he must have the capacity to evaluate tests and counter-arguments that are critical even of his favorite research hypothesis. If you are given a birthday present that you find ugly, you may afterwards (or perhaps even immediately!) in an angry mood throw it in a dustbin, but if you are a researcher meeting falsifying data or a powerful counter-argument, you are not allowed to throw them away. In this sense, scientific evaluation will always require some disinterestedness, even though the working out of the hypothesis can be as disinterested as a love affair; and researchers can literally fall in love with ideas. That is, the disinterestedness required in evaluation can co-exist with a strong non-disinterestedness in the truth of the hypothesis under scrutiny. A researcher should be allowed to have a personal interest (e.g., 'If my hypothesis is true, the pharmaceutical company will raise my salary considerably!'), a political interest ('If my statistics represent the truth, my party comes out much better in relation to its old promises!), or an ideological interest ('It must be the case that Xs in respect of Ws are superior to Ys!') in the truth of a certain hypothesis, but he has to be able to be a bit disinterested when evaluating data and listening to counter-arguments.

The norm of disinterestedness in evaluations of research is an overarching methodological norm, but it gets a moral touch because society at large may say that since you have chosen to become a researcher, it is your moral duty to try to be a good one.

- Originality:

It makes good sense to buy a chair that is a copy of a chair one already has. But to seek knowledge one already has is senseless. As the saying goes, there is no need to invent the wheel twice. From the perspective of the whole scientific community, these simple observations give rise to a norm: even though you should teach already gathered knowledge, and it might be good to teach the same pupils the same thing twice, *in research you are obliged to seek new knowledge*; be it radically new knowledge or only a small but nonetheless new piece of information. This norm is, by the way, one reason why plagiarism cannot be allowed.

The norm of originality may seem easy to follow, but there is a complication. Man is a social animal, and most persons have a strong tendency to be a bit conformist, but new knowledge might well be a threat to those who have a vested interest in the old and trespassed knowledge. Therefore, researchers may risk their social well-being if they stick to what they have found. This means that the norm of originality implies a sub-norm: scientific communities should train their members to become independent and self-reliant; it might be called ‘courage of creativity’ (Bauhn 2003). It also implies a sub-norm to the effect that researchers are obliged to try to keep updated with the developments in their specialty on a world-wide scale; not an easy task.

The norm of originality is in its general form somewhat moral; it states that it is a waste of scarce resources to discover what is already discovered. However, the research virtues that it promotes (being self-reliant and keeping one informed about others’ research) can be regarded as methodological virtues. As there are both moral norms and moral virtues, there are both methodological norms and methodological virtues.

- Skepticism (or: organized criticism):

In Chapter 3.5 we wrote: “Fallibilism is the view that no knowledge, not even scientific knowledge, is absolutely certain or infallible, but in

contradistinction to epistemological skepticism it is affirmative and claims that it is incredible to think that we have no knowledge at all." Should we stick to this terminology, according to which skepticism is wholly negative, then the norm of skepticism would need to be renamed into 'the norm of organized criticism', but here we will follow suit and continue to use the term 'skepticism'. Researchers have an obligation to be critical and skeptical both towards their own work and towards the work of others. A researcher should openly point to sources of possible errors, doubts, and weak spots in his research; also, he should wait until evidence-based knowledge is obtained before he presents his findings in public. In so far as the norm of skepticism requires that researchers should be critical or skeptical towards their own work, it overlaps with the norm of disinterestedness. Therefore, what should here be highlighted is the norm of mutual criticism or organized skepticism. It is a norm that the scientific community as a whole should force on its members the way a state forces its laws on the citizens. There is no need to require that each and every researcher should take this norm to heart, only that they act in accordance with it.

The way the norm works is well known. Students have to have their papers criticized in seminars, doctoral dissertations are scrutinized by chosen experts, and even scientific papers by famous researchers undergo so-called 'peer-reviewing' before they are allowed to be published in scientific journals; and as soon as research job is applied for, there are further evaluations. Many kinds of review-loops arise. The tightest one is this: at one occasion researcher R_1 reviews a paper by researcher R_2, and at another occasion R_2 reviews a paper by R_1. A somewhat less tight loop is that first R_1 reviews R_2, and then R_3, a colleague of R_2, reviews R_1; and there are many kinds of larger and far less tight loops. Such loops are inevitable because many researchers are not only researchers and authors of research papers; they also function as editors, referees, applicants for research money, and consultants for research funds.

The norm of organized skepticism is a methodological norm in a special sense. Normal discipline-specific (default) methodological rules are norms that tell each individual scientist how to behave. And the same is true of the norms of universalism, disinterestedness, and originality, but the norm of skepticism is primarily a socio-methodological norm that rules

interaction among researchers. To the extent that research is good for society at large, it can also, just like the norm of communism-and-communalism, be regarded as a moral norm.

We have already said that there is an analogy of some kind between the egoism of agents on the economic market and the egoism of researchers in the scientific community. Now, after having spelled out all the norms of science, in particular the CUDOS norms, we can make the analogy clearer. In both cases there is egoism-in-competition, not egoism-in-warfare. Firms compete on the economic market (with its laws and regulations) for the money of the consumers, and researchers compete in the scientific community (with its norms) for fame, money, power, and being part of desired research communities. In order to see also the differences, we have to take a closer look at some different forms of competition. When presenting these forms we talk neither of economic nor of scientific competition. They will be brought in afterwards.

Our first distinction (i) is between *counter-competition* and *parallel competition*. It is easy to exemplify with sports. In soccer, wrestling, and chess there is counter-competition; and in weight lifting, figure skating, and running there is parallel competition. In soccer, wrestling, and chess there are necessarily at one and the same time two teams or two individuals who compete against (counter) each other, but in weight lifting and figure skating the competitors can perform one by one (in parallel, so to speak) against the weights and the expert judges, respectively. Running is also parallel competition, because even if it is often performed by persons running at the same time on the same road, the competition could in principle have been performed by letting them run one by one against the clock. No such transformation is possible in counter-competition.

The second distinction (ii) is between *public-oriented competition* and *actor-oriented competition*. This distinction is easy to exemplify by means of competitions in cultural activities such as song, music, theatre, and poetry. Since long there are song contests and music charts, and since some decades there are also theatresports and poetry slams. In these, the public (or randomly chosen members of the audience) decides by simple voting who is to be reckoned the best performer. This is public-oriented competition. When the audience has no norm at all in common, then this kind of competition is only a competition about who can make the most

popular performance, not about who can perform best, but there are many degrees between being the best according to a referee or jury and having made only the most popular performance. In public-oriented competition, the competitors compete about the judgments or the favors of one, some or all persons in the audience; and the competitors and the audience are kept distinct. In actor-oriented competition, on the other hand, the same individuals are both actors and judges of the performances. Mostly, this kind of competition is only informal, and it exists, e.g., in many cultural avant-garde groups. The members of such groups often compete about an informal internal ranking in their esoteric group. However, nothing in principle prevents the staging of formal actor-oriented poetry slams where everyone in the audience is also a participant.

Some cases of what we call ‘public-oriented competition’ are better described by names such as ‘referee-oriented competition’ or ‘jury-oriented competition’, but we wanted a single term. Traditional sports contains these specific kinds of public-oriented competition; there is a referee, a group of referees, a jury, or a competition committee who finally decides who has won.

The two distinctions introduced relate to different aspects of a competition: (i) how are the competitors related to each other in the competition?, and, (ii) who decides who has won?, respectively. This means that the distinctions can be crossed, whereby we obtain the four kinds of competition represented in Figure 5.

Competition:

	Public-oriented	Actor-oriented
Parallel competition	1	2
Counter-competition	3	4

Figure 5: *Four different kinds of competition.*

What now to say about competition in science and in ordinary economic competition? The latter must in the main be regarded as belonging to square 1. Take, for instance, the competition between car producers. It is public-oriented competition; each producer tries to sell his cars to people in general, not to his competitors. The consumers decide who wins, not the producers and competitors themselves. It is also parallel competition, since the consumers can in the absence of the other cars find out about the good and bad sides of each car.

Basic research, on the other hand, is within each discipline (or sub-discipline) very much a matter of actor-oriented competition. It is the same group of people who both produce and judge the research results. In a 'research slam', the performers and the audience cannot be kept apart. We made the same point earlier in terms of 'review-loops'.

When basic research in a discipline is done on the basis of well entrenched and well specified methodological norms, then the corresponding competition is parallel competition, since each research project can be judged without immediately comparing them to other; such basic research belongs to square 2. However, because of the fallibilism of science and of the paradigm-boundedness of methodological norms, basic research may now and then end in situation where its results cannot be measured against a pre-given methodology, but only against competing theories. Since in such cases there is no meta-norm, the competition in question has to be a counter-competition. The pros and cons of the theory can only be seen in relation to another theory in about the way a boxer can only show his skills – really – when actually competing against another boxer. This kind of competition is the natural kind of competition in many philosophical areas. Why? Because parts of philosophy are by definition concerned with the utmost questions, and in relation to these there can by definition be no pre-given norms or criteria. And, as we have argued (Chapter 1), science and philosophy overlap. Using Thomas Kuhn's terms, we can say that normal science contains competition of the square-2 kind, whereas revolutionary science contains competition of the square-4 kind.

(Just for the sake of completeness we will mention also an example of competition of the square-3 kind. Normally, philosophy is actor-oriented counter-competition, but if one day 'philosophy slams' enter the cultural scene, then we would have philosophy also in the form of *public-oriented*

counter-competition. In each match in such slams, two philosophers should argue against each other on some topic, and the audience should decide who has produced the best arguments.)

Medical research requires yet another comment. From what has already been said, it follows that basic medical research is normally done within a framework of actor-oriented parallel competition (square 2), but that now and then it moves into a framework of actor-oriented counter-competition (square 4). But there is also a close link to public-oriented parallel competition (square 1). The reason is that in square 1 we find the kind of competition that medical drugs and medical therapies are involved in. Such things are not evaluated only by researchers and clinicians; what patients choose and reject can in many cases play quite a role for determining what drugs and what therapies will in the long run survive or ‘win’.

We would like to conclude this chapter by mentioning the problem of civil courage in research: what are we supposed to do if in some way or other we are the first to become aware of scientific misconduct? The positive way to describe a person who reveals fraud is to call him a whistle blower; as if everyone wanted this to be done. The negative way is to describe him as a betrayer, defector, traitor, or conspirator; he is putting science in a dark and unfavorable position. We have mentioned cowboy ethics in the derogatory sense, but old Western movies also contain the moral sheriff or outsider hero who, at the very risk of his life, manages to track down some villains and restore law, order, and ordinary living in some little Western community. Here are some modern moral hero stories.

Dr. Buxtun, who exposed the Tuskegee Syphilis Study worked for six years in order to succeed, and in the end he had to contact newspapers. Looking very decent in retrospect, he was probably in 1966-72 seen as a rather inconvenient person.

The young doctors who refused to assist Dr. Southam at the Jewish Chronic Disease Hospital, and even complained to their boss, Dr. Mandel, felt for leaving their positions at the hospital.

The two rocket scientists who told the investigators about the possible reason why the Challenger space shuttle exploded (1986) were discharged from their positions at the responsible company; in the aftermath, the US Senate provided them with a pension.

When a bank in Switzerland (1997) was going to destroy documents that proved their involvement in laundering Nazi-gold derived from the concentration camps, a guard discovered the nature of the documents and contacted the authorities. This guard was promptly fired, a judicial investigation was opened, and he got political asylum in the United States.

On Swedish water, a Russian oil tanker ran 1977 aground, and its captain simply claimed that something has to be wrong on the Swedish made sea-map. The relevant Swedish authorities bluntly denied this, and put the whole blame on the captain, but a specialist at the responsible department stated openly two years later that the captain was correct. This man was not fired, but it became difficult for him to continue to work at his old department, and eventually he agreed to leave with a financial compensation.

Once (1993) a French soccer player told media that a certain top team tried to bribe him. This other team was immediately by the soccer authorities replaced to a lower division, and its chairman ended in an ordinary jail. What happened to the player? He became completely isolated as a soccer player and found no new team to play with.

It is not easy to stand up and become a whistle blower. However, it is not especially hard simply to follow the Four Principles, the Helsinki Declaration, and the CUDOS norms. In the best possible scenario, no whistle blowers are needed.

Reference list

Angel M. *The Truth about the Drug Companies: How They Deceive Us and What to Do about it*. Random House. New York 2004.

Bauhn P. *The Value of Courage*. Nordic Academic Press. Lund 2003.

Bird K, Sherwin MJ. *American Prometheus. The Triumph and Tragedy of J. Robert Oppenheimer*. Vintage Books. New York 2006.

Bliss M. *The Discovery of Insulin*. The University of Chicago Press. Chicago 1982.

Broad W, Wade N. *Betrayers of the Truth. Fraud and Deceit in the Halls of Science*. Touchstone Books. 1983.

Bunge M. A. Critical Examination of the New Sociology of Science: Part 1. *Philosophy of the Social Sciences* 1991; 21: 524-60.

Deer B. < http://www.briandeer.com/ >

Doyal L, Tobias J (eds.). *Informed Consent in Medical Research*. Blackwell. Oxford 2000.

Horton R. The lessons of MMR. *The Lancet* 2004; 363: 747-9.

International Ethical Guidelines for Biomedical Research Involving Human Subjects. Council for International Organization of Medical Sciences (CIOMS). World Health Organization. Geneva 2002.

Fletcher W. Rice and beri-beri: preliminary report on an experiment conducted at the Kuala Lumpur lunatic asylum. *The Lancet* 1907; 1: 1776-9.

Fulford KWM, Gillett D, Soskice (eds.). *Medicine and Moral Reasoning*. Cambridge University Press. Cambridge 2000.

Gjestland T. The Oslo study of untreated syphilis: an epidemiologic investigation of the natural course of syphilitic infection based on a restudy of the Boeck-Bruusgaard material. *Acta Dermato-Venerologica* 1955; 35 (Suppl 34): 1-368.

Goodman KW. *Ethics and Evidence-Based Medicine. Fallibility and Responsibility in Clinical Science*. Cambridge University Press. Cambridge 2002.

Hunt L. *Secret Agenda. The United States Government, Nazi Scientists and Project Paperclip, 1945-1990*. St. Martin's Press. New York 1991.

Johansson I. Pluralism and Rationality in the Social Sciences. *Philosophy of the Social Sciences* 1991; 21: 427-43.

Jones JH. *Bad Blood: The Tuskagee Syphilis Experiment*. The Free Press. New York 1993.

Judson HF. *The Great Betrayal. Fraud in Science*. Harcourt, Inc. Orlando 2004.

Katz J. *Experimentation with Human Beings*. Russell Sage Foundation. New York 1973

Krasse B. The Vipeholm Dental Caries Study: Recollection and Reflections 50 Years Later. *Journal of Dental Research* 2001; 80: 1785-8.

Krimsky S. *Science in the Private Interest. Has the Lure of Profit Corrupted the Virtue of Biomedical Research*. Rowman & Littlefield. Lanham 2003.

Loue S. *Textbook of Research Ethics. Theory and Practice*. Kluwer Academic. New York 2000.

Lynöe N, Jacobsson L, Lundgren E. Fraud misconduct or normal science in medical research - an empirical study of demarcation. *Journal of Medical Ethics* 1999; 25: 501-6.

Lynöe N, Sandlund M, Dahlquist G, Jacobsson L. Informed Consent: study of quality of information given to participants in a clinical trial. *British Medical Journal* 1991; 303: 610-3.

Lynöe N. *Between Cowboy Ethics and Scout Morals* (in Swedish). Liber. Stockholm 1999.

McCallum JM, Arekere DM, Green BL, Katz RV, Rivers BM. Awareness and Knowledge of the U.S Public Health Service Study at Tuskegee: Implications for

Biomedical Research. *Journal of Health Care for the Poor and Underserved* 2006; 17: 716-33.

Rockwell DH. The Tuskagee study of untreated syphilis. *Archives of Internal Medicine* 1964; 114: 792-7.

Snowden C, Elbourne D, Garcia J. Zelen randomization: attitudes of parents participating in a neonatal clinical trial. *Controlled Clinical Trials* 1998; 20: 149-71.

The Human Radiation Experiments. Final Report of the President's Advisory Committe. Oxford University Press. Oxford 1996.

Ziman J. *Real Science. What it is, and what it means*. Cambridge University Press. Cambridge 2000.

11. Taxonomy, Partonomy, and Ontology

Classifying entities and discerning part-whole relations belong to the normal activity of everyday life. As soon as we describe a particular thing, we are classifying it, and as soon as we enter a house or a town, we are dividing it into parts. Such ordinary classifications and partitions have a practical purpose, and the classification and partition schemas used need not be systematic. In science, however, unsystematic classification schemas are often developed into well structured and principled general taxonomies such as the classic biological taxonomies of plants and animals. Similarly, parts and their positions in larger wholes are systematized by science into partonomies such as the anatomy of the human body and the double helix structure of the gene.

11.1 What is it that we classify and partition?

In Chapters 2 and 3, we talked a little about the scientific revolution of the seventeenth century. In taxonomy (the science of classification), the revolution occurred in the eighteenth century. The Swedish botanist and zoologist Linnaeus (Carl von Linné, 1707-1778) is often regarded as the creator of modern taxonomy, with Aristotle as its ancient founding father. During medieval times, alchemists made extensive classifications of chemical substances, and herbalists made the same with respect to plants, but in neither case was a real taxonomy created. This is probably because the alchemists and the herbalists were too practically minded. With the advent of modern chemistry and botany things changed. Better organized taxonomies were created, and helped to speed up subsequent scientific development. But with the recent information explosion and the advent of the computer revolution, scientific work with the construction of taxonomies and partonomies has taken on quite a new dimension. A new stage in the need for taxonomic classifications seems to have been reached. In new disciplines such as medical informatics, bioinformatics, genomics, proteomics, and a range of similar disciplines, classificatory issues play a prominent role.

11.1.1 Classification of particulars and classification of classes

There are two clearly distinct kinds of classifications: *classifications of particulars* (e.g., ‘this is a lion’, ‘this is a case of measles’, and ‘this is an example of turquoise blue’) and *classifications of classes* (e.g., ‘the lion is a mammal’, ‘measles is a virus disease’, and ‘turquoise blue is a color’). The term ‘particular’ is in this chapter meant to cover all cases where we normally speak of spatiotemporally specific *persons*, spatiotemporally specific *things*, or spatiotemporally specific *instances* of a property. Traditional taxonomies of plants, animals, and diseases are classifications of classes, whereas patient diaries and modern electronic health records rest on classifications of particulars (persons).

Assertions by means of which we *classify particulars* – such as ‘this is a lion’, ‘this is a case of measles’, and ‘this is an example of turquoise blue’ – are worth making because the animal could have been of another kind, the patient could have had a different disease or no disease at all, and the colored object could have been of a different color. To describe a particular in front of us is to convey information about this particular by means of a classificatory term. Subject-predicate sentences such as ‘this lion is somewhat brown’ contain two classifications: one for what falls under the subject term (‘being a lion’), and one for what falls under the predicate term (‘being somewhat brown’). Implicitly, there is a classification also when we describe something negatively, as in ‘this is not a lion’, ‘he does not have the measles’, and ‘this is not turquoise blue’. In the first case, we are implicitly conveying the view that there is an animal, in the second that the person in question has either no disease at all or a disease of some other kind, and in the third that the object in question is of some other color.

It seems impossible to talk about particulars without using classificatory terms. Even when we merely name something, as in the utterance ‘his name is Fido’, the context does normally indicate a classification. In this case, what is named is probably to be classified as a dog; the speaker could equally well have said ‘the name of this *dog* is Fido’. Such implicit classifications are necessary for communication. An utterance containing a pure ‘this’, e.g., the utterance ‘this has the name Fido’, cannot in itself possibly pick out anything in the world. Why? Since one would then not have any clue as to what the ‘this’ might refer to. The world contains at any moment an innumerable number of possible referents. Cut loose from

all classifications, the word ‘this’ has no communicative function at all, and the same goes for pure names. (In some corners of philosophy, especially those linked to Saul Kripke (b. 1940), there is much talk about a kind of pure names called ‘rigid designators’; but such a designator is always introduced by a ‘*description* used to fix its reference’.)

Assertions by means of which we *classify classes* – such as ‘the lion is a mammal’, ‘measles is a virus disease’, and ‘turquoise blue is a color’ – are worth making because it is often possible and expedient to convey in one stroke information about whole collections of particulars. All the three examples above have the linguistic form ‘A is B’ (in informatics, it is often written ‘A is_a B’), and should in what follows be read ‘the class A belongs as a subclass to the class B’ or ‘the class B subsumes class A’. One connection between classifications of classes and classifications of particulars is the necessity that if (i) class A belongs to class B, and (ii) a certain particular is classified as being an A, then (iii) it has to be classified as being a B, too. Necessarily, if Leo is a lion, and lions are mammals, then Leo is a mammal; if patient Joan has the measles, and measles is a viral disease, then patient Joan has a viral disease; if this spot is turquoise blue, and turquoise blue is a color, then this spot is colored.

So far, so simple; but what then do we more precisely mean by ‘class’? First, the term must not be conflated with the more specific biological term ‘class’ that occurs in the hierarchy of biological taxa: ‘species’, ‘genus’, ‘family’, ‘order’, ‘*class*’, ‘phylum’, ‘kingdom’. Second, neither should it be conflated with the broader term ‘class’ that is used in some programming languages. The general concept of ‘class’ is today normally defined (e.g., in Wikipedia, spring 2007) as follows:

- class $=_{\text{def.}}$ a collection of entities between which there are similarity relations.

This general definition being accepted, there is much to say about how different kinds of similarity relations constitute different kinds of classes. We will only talk about classes of spatiotemporal particulars (not about classes of abstract objects such as ‘the class of prime numbers’), but the term ‘class’ will here never be used to refer to *spatially or temporally bounded* collections. A class in our sense has, in contradistinction to

bounded collections such as ‘the lions *in the Copenhagen zoo*’, ‘the *hitherto living* dinosaurs’, or ‘measles patients *in the year 2000*’, no predetermined limits in space and/or time. The class of lions, the class of patients, and the class of blue property instances are open in the sense that if a new lion is born, a person suddenly becomes a patient, and something is painted blue, then these individuals and instances are automatically members of the corresponding classes. Conversely, if, for example, the lion species becomes extinct, it still makes good sense to talk of the class of lions.

(To those familiar with set theory, we have to add that classes cannot be identified with sets in the way that sets are defined in standard set theory. One reason for this is that there is only one ‘empty set’, i.e., a set without a member, but there are many ‘empty classes’. For example, ‘the class of mermaids’ and ‘the class of offspring of two mules’ are two different empty classes, but ‘the set of mermaids’ and ‘the set of offspring of two mules’ is one and the same set: the empty set.)

When we are classifying we are *using* classificatory terms and the concepts that come with the terms, but we are nonetheless *classifying* the classes that the concepts refer to. A concept is for us, to make this clear, merely the meaning that a term shares with synonymous terms. People often and easily conflate the *use* of terms and concepts in taxonomies with *talk about* these terms and concepts. Certainly, terms and concepts are talked about and classified in linguistics. But in all other traditional empirical sciences the terms and concepts are used in order to classify what they refer to, their *referents*. Because of this unhappy conflation of ‘using terms’ and ‘talking about terms’, we must spend some time on the relationship between language (terms and concepts) and reality (classes of particulars), as this relationship is not always as straightforward as it may appear to the un-philosophical eye.

What, for instance, do we classify when we classify classes of cells as being ‘epithelial’ and ‘neural’, respectively? There are not only two logically possible answers to this question: that we either *classify only the concepts* ‘epithelial cell’ and ‘neural cell’, or we must have an ability to see directly and to *carve, infallibly, nature at the joint* where the class of epithelial cells is distinguished from the class of neural cells. Since the realist fallibilism we have defended (Chapter 3.5) applies to taxonomy,

informatics, and the information sciences, both these answers are ruled out. Our *realism* says that science cannot be reduced to a ‘concept game’ (there is not only the concept of ‘cell’; there are cells, too), and our *fallibilism* blocks belief in the existence of infallible identifications of discontinuities in nature (e.g., epithelial–neural). However, since our view is that we should in general (i.e., apart from a possible research area of our own) regard the most truthlike theory available as in practice being true, we do from a practical point of view accept the old-fashioned talk about ‘carving nature at its joints’. But even given this, it is still false to say that a distinction such as that between epithelial cells and neural cells is simply *found* in nature. Even on the assumption that everything we today know about cells is absolutely true, there is a conventional element in the classification mentioned – as well as in many others. Explaining what is meant by this will involve a long detour, after which we will return to the example of classification of cells in Chapter 11.1.5 below.

11.1.2 Nature-given and (partly) language-constituted property classes

Think of how our everyday world visually appears to us. We see many different kinds of persons, animals, and things; and we perceive immediately most of them as having properties such as shape, length, volume, and color. A moment’s reflection on our color terms for perceived colors (not to be conflated with the quantitative terms for the wavelengths of light) tells us that, for instance, our common term ‘red’ does not refer to one and only one kind of perceived color hue; it covers a certain interval in the perceived color spectrum. Furthermore, the same is true across a smaller interval for terms such as ‘light red’. If the interval named is made smaller and smaller, we will in the end arrive at a term that picks out one and only one perceived color hue. In what follows, we will call such terms ‘*nature terms*’, and call what they pick out ‘*nature-given properties*’. What is referred to by ‘red’ and ‘light red’ will be called ‘*(partly) language-constituted properties*’. The so-called ‘Munsell Hue Designations’, which have not become parts of everyday language, come close to being nature terms and to supplying a term for each and every perceived color hue. We will treat them as if they do so; the Munsell system contains terms for one hundred different perceived color hues. For example, ‘red’ is divided into ten reds (1R, 2R, …, 10R), yellow-red into ten yellow-reds (1YR, 2YR,

…, 10YR), and red-purple into ten red-purples (1RP, 2RP, …, 10RP). Each such term can be used to refer to practically one and only one perceived color hue and/or the corresponding class of color instances. In Figure 1 some of the mentioned relationships are illustrated.

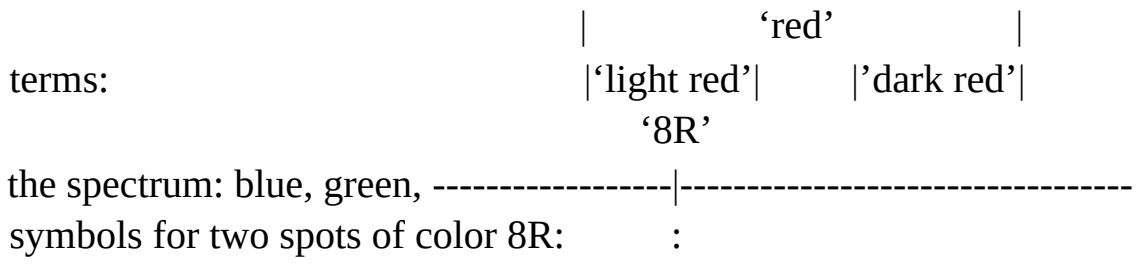


Figure 1: *Illustration of some relations between color terms, color hues, and two instances of a nature-given color hue. When there are vertical lines to the left and to the right of a term, the term covers everything that is in the spectrum interval below them; in case of ‘8R’ there is only a single vertical line below the term, since this term is assumed to refer to one single color hue only*.

In case of a quantified property such as length, every determinate numerical expression such as ‘1.97 m’ is a nature term that picks out one and only one nature-given length. But, of course, we can also choose to use expressions that pick out a range of nature-given lengths. We can, for instance, choose between assertions such as ‘this person is exactly 1.97 m tall’, ‘this person is around 1.97 m tall’, ‘this person is between 1.96 and 1.98 m tall’, and simply ‘this person is very tall’; see Figure 2.

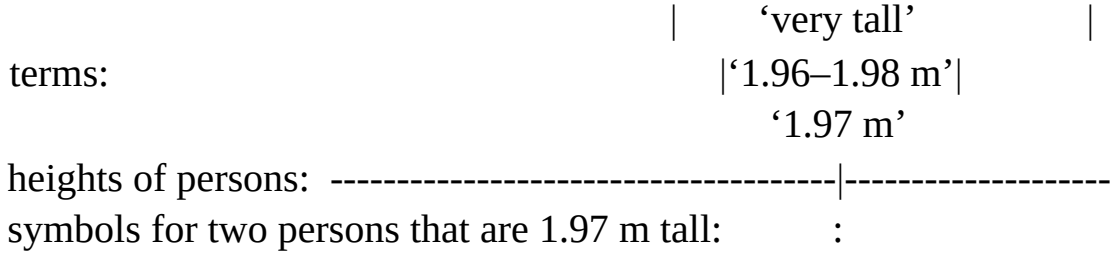


Figure 2: *Illustration of some relations between length terms, heights of persons, and two individual persons.*

These remarks on the property dimensions of color hue and length can easily be generalized to other monadic properties such as mass and electric charge. The upshot is the following: on ordinary assumptions about veridical perceptions and physical reality, each language-independent property dimension has nature-given determinate properties; this holds true for property dimensions in both nature and perception. In order to talk about them we have of course to use terms and concepts. But the determinate properties we talk about have nonetheless a language-independent kind of existence. In Figures 1 and 2, ‘8R’ and ‘1.97 m’, respectively, are nature terms. All the other property terms (‘light red’, ‘red’, ‘1.96–1.98 m’, ‘very tall’) refer to many nature-given properties, and the classes they pick out contain many nature-given classes. These classes, therefore, had better be called ‘partly language-constituted classes’; the boundaries of the classes depend on terms in language, and are in this sense fiat. There is much more to say about the interplay between property terms and nature-given properties, but first we have to articulate a philosophical feature that all nature-given properties have in common:

- nature-given properties can be instantiated in many different places simultaneously.

In Figures 1 and 2, this fact is indicated by the two dots in the lowest line, which represent two color spots and two persons, respectively.

11.1.3 Repeatables (universals)

In the sense we are using the term ‘particular’, a particular (be it an individual person, an individual thing, or a certain property instance as such) can by definition be only *at one place at one time*. But, of course, several different persons can simultaneously have exactly the same hair color, have exactly the same shape of the femurs, and be exactly 1.97 m tall. That is, a nature-given property can via its instances be *at many places simultaneously*; it can have a scattered spatiotemporal existence without losing its identity and unity. Philosophers usually call entities that have this feature ‘universals’, but we will call them ‘repeatables’; we present the reason for our terminological change at the end of this section. To each

repeatable there is a corresponding open-ended class whose members are all the past, present, and future instances of the repeatable in question.

It is an undeniable fact that in everyday life we describe the world as if it contains language-independent repeatables, but many philosophers question their existence and claim that to believe in repeatables (universals) is to fall prey to a linguistic illusion. This existence problem has been dubbed ‘the problem of one-in-many’, i.e., ‘the problem of *one*-repeatable-existing-*in-many*-spatiotemporal-locations’. We are firmly convinced, and will now try to show, that both our perceived common sense world and the world that science investigates have to be regarded as containing language-independent repeatables.

(To those familiar with set theory, we would like to add that in the second half of the twentieth century it was a very widespread view, originating in W.V.O. Quine (1908-2000) that there is one and only one kind of repeatable/universal, namely set.)

A first thing to be acknowledged is this: if there is communication, then there are at least repeatables *in language*. If one person Jack says ‘my car is blue’, and another person Jill hears and understands his utterance, then the semantic content (the proposition) of the assertion ‘my car is blue’ must exist in the minds of two different persons; and so be a repeatable. There are then *two different* instances of *one and the same* semantic content. If there are merely two different instances that have nothing in common, then Jill has not understood what Jack said, and nothing would have been – contrary to our assumption – communicated between them. If Joe then asks Jill ‘what did Jack say?’, and Jill answers ‘he said my car is blue’, the content of the assertion becomes instantiated in Joe’s mind too. It is a brute fact of ordinary language that, in some way or other, it contains repeatables. If the critic of repeatables then asks:

- but *how* can something be in different places simultaneously? Isn’t it a logical contradiction to say that one thing is in many places simultaneously?

the answer is:

- it is only a logical contradiction to say that a *particular* is in many places simultaneously. The question *how* a certain something can be in different places simultaneously has the blunt answer: because this something is a repeatable.

The fact that language contains repeatables does not in itself prove that there are repeatables in language-independent reality, but noticing this fact takes away the widespread general philosophical presumption that there are no repeatables at all. And then it is hard to find any good arguments that would privilege language as being the one and only realm that can contain such entities. Since we can perceive several things in perceptual space as having exactly the same shape or the same color, why shouldn't there also be perceptual shape-repeatables and color-repeatables? And when molecular biologists say that (in real space) all DNA molecules have the shape of a double helix, why shouldn't we be allowed to interpret them literally? There seems to be no reason at all. Look now at the following circle-shaped black spots:

In relation to these five (numerically different) spots we can truly say two different but related things:

- the five spots are *identical* with respect to shape
- the five spots are *exactly similar* with respect to shape.

The identity spoken of in the first sentence is the *identity* of a *repeatable*. Where there is qualitative and/or quantitative identity, there is a repeatable; and where there is a repeatable there is identity. The exact similarity *relations* spoken of in the second sentence are relations between the five *instances* of the shape repeatable. Where there are instances of the same repeatable, there are relations of exact similarity between the instances; and where there are relations of exact similarity between instances, there are instances of the same repeatable.

Here we stumble upon a philosophical problem: are the similarity relations there because the spots have the same shape (instantiate the same repeatable), or do the spots acquire the same shape because (independently of their properties) there are similarity relations between them? We find it more reasonable to think that it is the shape repeatable *circle* that explains the exact shape similarities between the spots, rather than vice versa. From where should any non-arbitrarily projected exact shape similarities come? But be it as it may with this controversy (between 'realism' and 'similarity/resemblance nominalism'); for our future exposition it is important to keep in mind only that a repeatable has an identity of its own, and that between any two spatiotemporal locations where a certain repeatable is instantiated, there is a relation of exact similarity. The concept of 'nature-given class' can be defined as follows:

- nature-given class $=_{def.}$ a collection of particulars between any pair of which there holds a relation of *exact* similarity in the same respect.

The structure of our reasoning in relation to the circle shape is quite general; it could just as well have been made in relation to the black color; and it can in some form or other be made in relation to all nature-given properties. Let us state our conclusion: *all nature-given properties are language-independent repeatables,* and *to each such repeatable there is a corresponding unbounded class of instances*. Figure 1 can be re-drawn as in Figure 3.

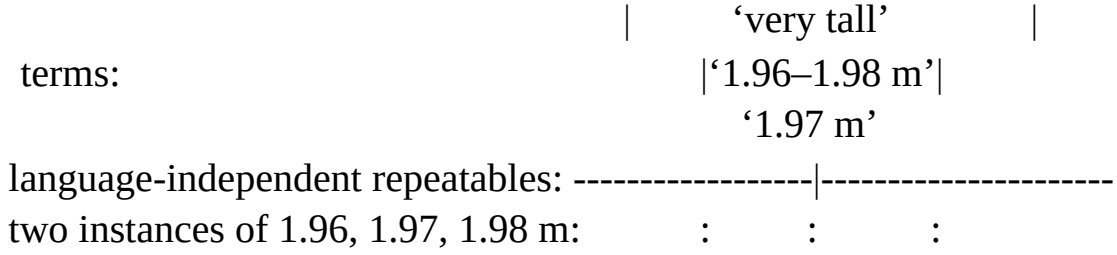


Figure 3: *Illustration of some relations between length terms, repeatables, and instances of repeatables; these instances are at the same time referents of the terms.*

With the help of Figure 3, our distinction between nature-given (language-*in*dependent) repeatables/classes and partly language-constituted repeatables/classes can easily be seen. The class of instances that are exactly 1.96 m long is a nature-given class, and so are the classes of instances that are exactly 1.97 and 1.98 m. The class of instances that are between 1.96 and 1.98 m long, however, depends for its existence as a single class both on the term ‘1.96–1.98 m’ (or on a corresponding term such as ‘77.165–77.953 inches’) and on the fact that the members of the class have similarity relations of such a character that the members can be ordered on a single line; and the same goes for ‘very tall’. These latter classes are *partly* language-constituted classes. Their boundary is a human creation, even though *what* is bounded (i.e., instances of a number of nature-given repeatables with similarity relations) is not. This remark applies not only to length, but to colors, shapes, and other such monadic properties too.

The distinction between ‘(purely) nature-given classes’ and ‘(partly) language-constituted classes’ will re-appear below in relation to natural kinds. Most property terms of everyday language refer to language-constituted classes. The members of such classes need not be (in contradistinction to nature-given classes) *exactly* similar to each other, but there have to be some kinds of similarity relations between them.

Next, we will deliver the explanation why we have chosen to use the term ‘repeatable’ instead of the traditional philosophical term ‘universal’. The latter term had its philosophical meaning established by Plato and Aristotle, who had a non-evolutionary view of the universe. They thought that all the basic elements and all the different kinds of living entities exist as long as the world exists. Therefore, these basic elements and natural kinds were not only regarded as repeatables, they were regarded as also having a temporally universal existence. Modern evolutionary cosmology and biology changed all of this. According to evolutionary biology, there was, for instance, a very first particular dinosaur and later on a very last one. There can be dinosaurs in many places simultaneously, i.e., to be a dinosaur is to instantiate a repeatable, but there are not always dinosaurs. What repeatables and universals have in common is that they are unchangeable. This being said, we can start to talk about natural kinds.

11.1.4 Nature-given and (partly) language-constituted natural kinds

Most sciences contain, like common sense, a distinction between various kinds of entities that are bearers of properties (material things, plants, animals, tools, machines, and other similar devices) and various kinds of properties (shape, color, weight, diffusion capacity, reproductive capacity, having fur, having a spinal chord, being two-legged, having four wheels, etc.). In the life sciences, this distinction between a property and its bearer is obvious, but it has been doubted that it exists or is necessary in modern physics. If one thinks, as many do, that physics is ahead of other scientific disciplines and provides a model for the latter, then what is true of physics in this connection may be of importance to the life sciences too. That is, *if* the property-bearer versus property distinction has become obsolete in physics, then the life sciences should perhaps try to get rid of it as well. Some appearances notwithstanding, however, even modern physics – from Newton's mechanics and Maxwell's electromagnetic field theory to relativity theory and quantum mechanics – contains the distinction in question, and there is therefore no reason to try to eliminate it from the life sciences. The illusion that it has disappeared is probably caused by the following two facts: (i) most kinds of property-bearers postulated in today's subatomic physics are not particles in the traditional sense, and (ii) most mathematical equations in physics do not explicitly mention any property-bearers, but relate only physical variables to each other.

Newton's second law, 'force = mass times acceleration', in itself relates only quantitative values of forces, masses, and accelerations to each other. However, it nonetheless takes for granted that masses are properties *of* material particles (property-bearers), that force is a relation between such particles, and that acceleration is a relation between such a particle and absolute space. Maxwell's equations state relations between variables for electric field strength, electric flux density, magnetic flux density, electric charge density, and current density, but these are always properties *of* electromagnetic fields. Such fields are not material things in the ordinary sense, but they are nonetheless (if the theory is true) property-bearers existing mind-independently; they can exist apart both from the atoms that produce them and any substratum such as the once postulated aether. In philosophical terminology, they can be called 'substances' just like material things.

Neither does the theory of special relativity invalidate the thesis that there exist property-bearers. It places Newton's laws and Maxwell's equations in a new kind of framework, but this framework does not take away the distinction between property-bearers and properties. This is so even if those philosophers who argue (e.g., Bertrand Russell) that special relativity implies that the notions of 'particles' and 'fields' should be replaced by the notion of 'event' would be right. For even if this view were true, all such 'events' would then themselves serve as bearers of properties such as rest mass and electric charge. In the change to the theory of general relativity more radical things have happened, but none radical enough to wipe out the notion of 'property-bearer'. The equation system that is called the theory of general relativity has many solutions. Each solution describes the whole universe in space and time as being one single huge property-bearer that has properties (of mass-energy) in each of its four-dimensionally indexed point-events.

Whatever is to be said about quantum mechanics and its various philosophico-ontological and epistemological interpretations, it holds true that – from its birth to the present day – quantum physicists have distinguished between different kinds of subatomic entities, to which they have ascribed properties. Today, the main properties reckoned with are mass, electric charge, and spin, which are ascribed to property-bearers such as electrons, muons, tau leptons, quarks, antiquarks, photons, and gluons. Sometimes these kinds of property-bearers (which are called 'particles' despite not being particles in the sense of Newton's mechanics) are, in analogy with classical botany, ordered into families and groups. The distinction between property-bearers and properties is very much alive even in present-day quantum physics. But we will in what follows mainly stick to the life sciences and their property-bearers.

In Chapter 11.1.2 we distinguished between nature-given and language-constituted properties; the former being referred to by means of nature terms, the latter by means of language-constituted property terms. Now we will start to discuss whether there is a corresponding distinction between nature-given and language-constituted property-bearers, i.e., instances of natural kinds. Let us take a look at the ranks in a taxonomic hierarchy such as that from the class of animals down to the classes of lions (the species) and Asiatic lions:

Animalia – Chordata – Mammalia – Carnivora –
Felidae (cats) – Panthera (the-four-big-cats) –
Panthera-leo (lion) – Panthera-leo-persica (Asiatic lion).

In relation to natural kind terms such as these, our question can now be put as follows: can some of them be nature terms? Or, in other words: are there any kind repeatables that are to natural kinds in general what the most determinate color hues are in relation to color? Might perhaps our terms for species be nature terms, and the different species (here 'lion') be nature-given repeatables?

The class Asiatic lion is called a subspecies. Asiatic lions have a scantier mane than other lions and a characteristic skin fold at their belly. But is the term 'Asiatic lion' a true nature term? It seems metaphysically odd to assume that there is an infinite progression of classes from subspecies to sub-subspecies, sub-(sub-subspecies), and so on, each one different from (more narrower than) its predecessor. But we can still take some further steps from 'Asiatic lion'. Being a property-bearer, an instance of a natural kind can of course be specified by means of each and every determinate property it is bearer of. Walking down the determinate-properties-line would give us classes where all the members (of each sex and of the same age) have more or less the same bundle of determinate properties. It would give us classes such as that of cloned-Asiatic-lion-of-type-1, cloned-Asiatic-lion-of-type-2, cloned-Asiatic-lion-of-type-3, etc. No doubt, in the life sciences clones have in one sense to be regarded as being nature-given natural kinds. But what are we to say about species in this respect? There is a long-standing intuition to the effect that there is, so to speak, something extra-natural about the natural kinds that species constitute. Is this a mere illusion? Cannot species, just like clones, be nature-given kinds, too? Notwithstanding the fact that their members need not be like clones in having all their properties in common – both yellow lions and white lions are still: lions.

Before we continue, let it be noted that in physics all the most specific natural kinds – such as the isotopes of atoms and the subatomic particles – are such that their instances are identical with respect to *all* their properties. These natural kinds might well be called 'inert-matter clones'.

In order to answer the question whether species can be regarded as nature-given and not as language-constituted natural kinds, we have to return to properties and make some further remarks on them. Not only do the most determinate properties exist independently of the terms we use when we talk about them, the division of determinate properties into property dimensions such as length, time, mass, shape, color, etc., seems also to be outside the sort of conventionality of language we noted with respect to the terms ‘1.96–1.98 m’ and ‘red’. This fact comes out most easily in relation to the basic quantified dimensions of physics, e.g., length, time, mass, electric current, and temperature. Quantities are fusions of numbers and property dimensions (in metrology the latter are called ‘quantity dimensions’, but since we talk also about non-quantified properties we will use the term ‘property dimensions’), and they are subject to a principle of exclusion that is so obviously true that it is hardly ever mentioned in physics:

- no entity can possibly at one and the same time take two specific values of the same property dimension (quantity variable).

Thus, no material object can simultaneously have two masses, two volumes, two electric charges, etc. Such principles of exclusion are most easily understandable in relation to quantities, but they have equally valid counterparts also for property dimensions such as shape and perceived so-called ‘surface colors’. No object can have two determinate shapes at the same time, and no distinct surface can be perceived as having two colors. Both property dimensions and their most determinate properties are repeatables; and they come as unities in the sense that (i) where there is an instance of a determinate property there is also an instance of a property dimension, and (ii) where a property dimension is instantiated there must also be an instance of a determinate property.

Quantified property dimensions can be multiplied and divided as in ‘velocity = length / time’ and ‘body mass index (BMI) = (body weight of person) / (height of person)^2’. It is, however, impossible to create a term such as ‘masslength’ that can cover both the property dimensions ‘mass’ and ‘length’ the way ‘1.96–1.98 m’ can cover both ‘1.96–1.97 m’ and ‘1.971–1.98 m’, and ‘red’ can cover both ‘light red’ and ‘dark red’. The

latter constructions conform to the principle of exclusion, whereas ‘masslength’ does not; nothing can be both ‘1.96–1.97 m’ and ‘1.971–1.98 m’ long, and nothing can be both ‘light red’ and ‘dark red’, but objects that have a certain mass have also a length (along any chosen axis). That is, there seems to be a nature-given discontinuity between property dimensions.

The observation just made is further strengthened in case the quantities are mathematically additive, since only quantities of the same property dimension can be added together to yield a meaningful sum. The expressions ‘5 kg + 3 kg’ and ‘5 m + 3 m’ are perfectly intelligible, but ‘5 kg + 3 m’ makes no sense.

Some words about the term ‘discontinuity’. In a continuum, it is always possible to find, between two points A and B, a third point C. For instance, between any two color hues that you are able to perceive as being distinctly different in the color hue spectrum, there is always a third hue. All the property dimensions mentioned contain in this sense a continuum of determinate properties. If, however, we try to find in a similar way between two *property dimensions* A^d and B^d a third property dimension C^d, we find nothing; between length and mass there is nothing. This being noted, we can return to our question whether species in biology can be regarded as nature-given natural kinds.

A species in the pre-Darwinian sense – a *typological* species – is not just a class whose members are similar with respect to outward appearances (a *morphological* species). At least since Linnaeus’ time, a necessary and sufficient condition for a class to be a species is that its members have the capacity to reproduce, i.e., the capacity to yield fertile offspring. Consequently, the other biological ranks (subspecies, genus, family, etc.) cannot be characterized by this feature. In a discipline such as paleontology, the morphological features of the fossils are automatically regarded as clues to the respective species. In biology, however, the existence of some ‘sibling species’ is regarded as a fact; if S_1 and S_2 are sibling species, then they make up two different typological species but one and the same morphological species.

For members of asexually reproducing species, the capacity to reproduce is a *property* of each member, but in cases of sexually reproducing species,

the capacity to reproduce is a *relation* between members of the sexes in question.

Before the Darwinian revolution, species were regarded as being discontinuous with each other in a way analogous to the way in which property dimensions are discontinuous with each other. It was thought that the members of a species give rise to a new generation of the same species or they do not produce members of any species at all; hybrids such as mules, hinnies, ligers, and tigons were regarded as infertile products, not as members of any species. On these assumptions, each species could truly be regarded as being a unique nature-given referent of its classificatory term, and not due to any fiat discontinuities introduced by us by convention in a process of speciation that in itself is somewhat continuous. But these assumptions are false, and evolutionary biology needs another species concept. However, the need to replace the typological concept does not arise simply because there is evolution; it arises because there is *gradual* evolution. Let us explain.

The following holds true for the typical members of a typological species (what 'typicality' more exactly means is explained in Chapter 11.1.5 below). If the members of generation A produce generation B, and generation B produces generation C, then the members of the different generations, A, B, and C, have such characteristics that, were it not for the time gap, they could together have produced fertile offspring, too. Think next of the following scenario. When generation B is produced, a kind of mutation takes place (because, say, of the influence of cosmic rays). The mutation divides the members of generation B into two groups; on the one hand those who can in principle produce fertile offspring with generation A, and on the other hand those who cannot. Assume further that the mutated individuals are not as such infertile. Within their respective group each can produce a new generation, which in turn can produce still another generation, and so on. If evolution had contained only this or similar kinds of *saltational* origins of new species, there would have been no need to replace the typological species concept; but evolution also contains structures of the following kind:

- Generation A produces generation B, which in principle can produce offspring with generation A; generation B produces generation C, which in principle can produce offspring with generation B, but *not with A.*

According to the twentieth century (many-staged) synthesis of original Darwinism and population genetics, most members of a species differ somewhat with respect to genetic material, and members of one generation can because of mutations and genetic drift produce offspring that have genetic material that is not to be found in the parent generation. Often the offspring is not fertile, but sometimes it is. When this is the case, the offspring can sometimes in principle produce offspring with the preceding generation, but sometimes only with members of its own generation. This means that there is no general nature-given property ‘capacity for reproducing’ that delimits all species.

How then to characterize what is common to all species, or at least to sexual species? Since each individual member has many ordinary (monadic) properties and many relational properties such as ‘can reproduce with individual i’, it is from a God’s-eye point of view possible to construct species classes in the way science has constructed many language-constituted property classes. The problem is that scientists are humans with limited brain capacity; useful language-constituted class concepts of the mentioned sort seem to be epistemologically impossible to construct, since far too many properties and relations are involved.

Evolutionary biology has made another kind of move; it has dropped the pure class concept of species (Ghiselin 1997). A common modern definition of sexual species, propounded by the German ornithologist and philosopher of biology Ernst Mayr (1904-2005), says:

- *Biological* species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups.

The philosophical change involved in the move from the *typological* species concept to the *biological* species concept is indicated by the term ‘population’. According to the typological concept, a species is a class, and

a class has no boundaries in space and/or time. But a population has such boundaries, and more than that. A population is not just a class plus a conventional spatial and/or temporal boundary; the individuals in the population should also be linked by chains of interaction, which makes the population into a kind of spatiotemporal particular. In a sense, the individuals of a population are to the population what the cells of an organism are to the organism. A population is, just like an organism, not a repeatable and class but a particular and a property-bearer. The individual plants or animals of a *biological* species are not members but *parts* of their species. In-between the individuals and the whole population, biologists sometimes discern other kinds of parts, e.g., local populations or ‘demes’.

A thought experiment may highlight the essence of the biological species concept. If, on another planet (P) somewhere in the Universe, there are lion-like animals that in principle can produce fertile offspring together with Earth-lions, these P-lions are nonetheless parts of another population. Therefore, *according to the definition* of biological species, Earth-lions and P-lions have to be regarded as two different biological species. Mayr is quite explicit on this (1988, p. 343). And a similar remark holds for the temporal dimension. If, on Earth, lions become extinct someday, but are then produced a second time by natural selection, then these second-time lions would be a different population and, therefore, also a distinct biological species. The expression ‘reproductively isolated’ is given such a broad sense in Mayr’s definition that not only fertilization barriers (no zygote can be formed), hybrid barriers (the zygote is not viable), and hybrid sterility count as examples of reproductive isolation, but so also do behavioral differences (the courtship rituals do not fit those of their mating partner) and significant physical barriers (oceans, galaxies, etc.).

Typological species (species-as-classes-and-repeatables) are by definition unchangeable, whereas biological species (species-as-populations-and-particulars) allow change, since they are property-bearers, too. Biological species can move on Earth, they can exchange one niche for another, and they can even radically change their genetic material. Does this new species concept therefore make the old typological concept obsolete in the way modern chemistry has made obsolete the concept of phlogiston? In other words: does the old typological species concept refer to something merely fictional? The answer is straightforwardly ‘no, it does

not'. For the notion of 'potential reproduction', which is at the heart of the typological species concept, is still as applicable to *individuals* of different sexes as it ever was, and it is even used in Mayr's definition. In all probability, the typological species concept will always be useful in some restricted contexts. It might even be very useful within central biology. If it is true that speciation only occurs during some relatively brief periods of time (the theory of punctuated equilibria), the typological species concept is as applicable in the normal equilibrium periods as it was in pre-Darwinian theorizing.

On the other hand, it is not only evolutionary biology with its theory of *vertical* gene transfer that has made it hard to apply the typological species concept everywhere. Some organisms with prokaryote cells (e.g., bacteria) and some unicellular eukaryotes show evidence of *horizontal* gene transfer, i.e., some organisms seem to receive genetic material not only from their ancestors. When an organism produces an offspring and there is a non-negligible horizontal gene transfer, then the organism cannot be said to *re*produce itself. Genetic engineering is a form of artificial horizontal gene transfer.

(The move from species-as-repeatables to species-as-particulars is retained in phylogenetic systematics, the discipline that devotes itself exclusively to finding genealogical trees that mirror the process of speciation on Earth. Its founding father is E. W. H. Hennig (1913-1976), and it is nowadays mostly called 'cladistics'. It should be noted, however, that the notion of 'ancestor' used in cladistics is not completely identical with the notion of 'biological species ancestor'. According to Mayrian evolutionary biology, a biological species can give rise to a new species but continue to exist; in mainstream cladistics, on the other hand, it is simply postulated that in such cases, after the speciation event, the old species should be given a new species name. Even though phylogenetic systematics is sometimes called 'cladistic taxonomy', it does not construct taxonomies in the ordinary sense of this word. Its geneaological trees are a kind of temporal *partonomies*, see Chapter 11.1.7.)

We will next describe still another complication that prevails in the relationship between natural kind terms and natural kinds. But it will be the last one. Not even the typological species concept is as simple as we have presented it. Normally, several members of a typological species lack the

capacity to reproduce; be it for anatomical, physiological, or courtship–ritual reasons. In this sense, the class of (typological) lions is wider than the class of lions that *can* reproduce. Even pre-Darwinian species classes are (partly) language-constituted classes. But, as we shall now explain, they are language-constituted in another way than the language-constitution described in Chapter 11.1.2.

11.1.5 Nominal, real-prototypical, and ideal-prototypical terms

Species terms in classical non-evolutionary biology (e.g., 'lion') have a relation to the repeatables and instances that fall under them that is distinct from our description of property terms in the Chapters 11.1.2-3. In the philosophy of science, there is a distinction between two different kinds of terms by means of which *particulars* are classified:

- nominal terms (reflecting nominal classification of particulars)
- prototypical terms (reflecting prototypical classification of particulars)

Figures 1-3 above represent nominal terms and nominal classifications – with the simplification that all vagueness is taken away (i.e., in Figures 1-3 there are definite boundaries for what the terms refer to). This means that each and every term, the most specific as well as the most general, relates to its repeatable(s) and its instances in the same way. In relation to a nominal term, a repeatable either falls under the term or it doesn't, and the same holds for all the corresponding instances; either a red color hue is a red color hue or it isn't; either an instance of red belongs to the class of red instances or it doesn't. There are in the figures no *degrees* of 'being red', 'being light red', 'being 8R', 'being very tall', 'being between 1.96 and 1.98 m', and 'being 1.97 m'. In *nominal* classifications, all referents (repeatables as well as instances) fit their respective classificatory terms either completely or not at all.

In *prototypical* classification, on the other hand, the referents can fit the term more or less. There are, as everyday language has it, *typical* lions, *typical* roses, and so on. A prototypical term such as 'lion' refers directly to a small range of prototypical repeatables and their classes of instances, and indirectly to other classes of individuals. However, for simplicity's

sake, we will in what follows write as if every prototypical term refers to one and only one repeatable. This means that we can say that all prototypical individuals or instances resemble each other exactly in all relevant aspects, whereas the individuals or instances of non-prototypical classes resemble the prototypical ones only to a certain degree. The less an individual resembles the prototypical individuals, the less representative it is of the prototypical term. Such a term functions as a *measuring* norm, but, let it be noted at once, it needs not represent a *moral* norm; more about this at the end of this section. Even many property terms can be given a prototypical reading. For instance, the term ‘red’ is in many everyday contexts equivalent to the explicitly prototypical term ‘typical red’. The relation between this term and its referents – repeatables as well as the corresponding classes of instances – can then be illustrated as in Figure 4.

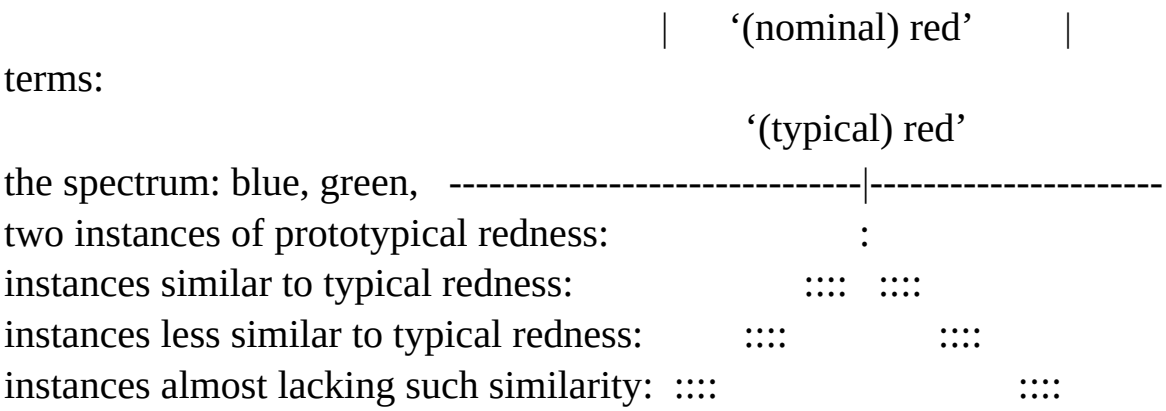


Figure 4: *Illustration of how the referents of a prototypical term can ‘fade off’ from instances of the prototypical repeatable.*

Prototypical terms designate at one and the same time both one specifically chosen repeatable and the corresponding class, the class of prototype instances, and many other but somewhat similar repeatables and their classes. The term ‘(typical) red’ and its referents allow themselves to be illustrated in the simple way of Figure 4, but the term ‘(typical) lion’ does not. This is due to the fact that color hues can be ordered on one single line, whereas lions have many different property dimensions along which they can differ (e.g., size, weight, general color, mane shape, mane color, tail length, and others). Nonetheless, nature has endowed us with

such perceptual and linguistic capacities that we can identify and communicate about non-typical lions by means of the term ‘lion’.

Cognitive scientists and linguists have made extensive studies of what people take to be typical referents of various terms; both terms for artifacts such as ‘furniture’ and terms for natural kinds such as ‘bird’. It seems to be quite clear that many perceptions have an in-built feature of typicality, and that when children learn a language, they first learn prototypical terms. When prototypical terms are taught in floras and animal atlases, normally a mix of words and pictures are used; but it is often not enough only to read and look in such books in order to become capable of applying the prototypical terms in question. The reading and looking has to be complemented by activities that give rise to ‘tacit knowledge’ in the sense spoken of in Chapter 5. To be able to identify in perception something as being the referent of a prototypical term does often require know-how. This is evident in disease identification.

In Figures 1-4, the relations between terms (nominal as well as prototypical) and referents (the corresponding repeatables and their classes of instances) have been represented as if there is a clear and distinct border for each term. But, well defined quantitative terms apart, this is seldom the case. Most non-quantitative nominal terms are *theoretically vague* in the sense that there are repeatables in relation to which it is unclear whether the term is applicable or not. And the same lack of definite application boundaries is characteristic of prototypical terms. It is almost never made clear where the limits for prototypical terms are situated, in fact this is so seldom the case that such terms are not even called vague; theoretical vagueness seems to be regarded as a normal feature of prototypical terms. Of course, even quantitative terms can because of measuring problems be *epistemologically vague* in the sense that it might be hard to know exactly when something falls under them.

In cases where it is possible to obtain an overview of all the dimensions along which a nominal or prototypical term can have referents, language communities can in principle make a decision to draw borders by fiat at certain pragmatically useful places. In the case of color hues, where a linear ordering exists, it would only be a practical problem to define by decree exactly where, for instance, red turns into light red and dark red, respectively. In relation to some terms, such taking-vagueness-away

definitions have actually been proposed and endorsed; if only within smaller expert communities. One such case is the continuum of wind intensities; here the Beaufort scale has stipulated exact boundaries in m/s for twelve different terms; among them ‘calm’ (0–0.2 m/s), ‘gentle breeze’, ‘gale’, ‘storm’, and ‘hurricane’ (32.7–40.8 m/s). Still, of course, there is the epistemological problem of how to know when there are actual instances of quantitative measures such as 0.2 m/s.

We are now in a position to return to the distinction between ‘epithelial cells’ and ‘neural cells’. In Chapter 11.1.1 this was mentioned as an example of a distinction that is necessarily language-constituted. But this example contains a special complexity. It combines a natural kind term (‘cell’) with property terms (‘epithelial’ and ‘neural’, respectively). The distinction between ‘epithelial cells’ and ‘neural cells’ is in this sense doubly language constituted. Furthermore, since the distinction between nominal and prototypical terms is applicable both to natural kind terms and property terms, a complex term such as ‘epithelial cell’ can be given four different readings: (i) both the terms are nominal, (ii) both are prototypical, (iii) the kind term is prototypical but the property term nominal, and (iv) the kind term is nominal but the property term prototypical. Case (iii) is displayed in Figure 5.

	Nominal term	Prototypical term
Natural kind term		‘cell’
Property term	‘epithelial’	

Figure 5: *Illustration of a possible combination of nominal and prototypical terms.*

In engineering, a prototype is either a purely mental construct or an initial model, for example a simple scale model. Metrology (the science of measurement), is also using the term ‘prototype’. Basic measuring units can be either theoretically defined repeatables or defined by means of a particular material thing; when the latter is the case there is a prototype. This dual possibility is easily seen in relation to the International System of Units (‘the SI-system’), which stipulates what measuring units the natural-

scientific community should use (for length ‘meter’, for time ‘second’, for mass ‘kilogram’, for electric current ‘ampere’, and so on). For quite a period of time, a real material standard meter in Paris (a platinum-iridium bar with two scratch marks) was the prototype, which meant that all length instances which were exactly similar to this rod were regarded as being 1 meter long too. But in 1983 this real thing was exchanged for a theoretically defined construct: ‘1 meter $=_{def.}$ the distance covered by the speed of light in 1/299792458 of a second’. All instances of this ‘1-meter repeatable’ are now by definition 1 meter long.

The SI-system contains seven basic measuring units, and today all of them except the prototype for ‘1 kilogram’ have been theoretically defined, but it is expected that sooner or later even this prototype (a solid platinum-iridium cylinder) will be exchanged for a theoretical construct. Such a switching between a particular (a prototype) and a repeatable (a theoretical definition) is possible since, as we have made clear, all instances of a property repeatable are exactly similar. This distinction between mental or theoretical prototypes/constructs, on the one hand, and real spatiotemporally located instances, on the other, must not be conflated with the following distinction between two kinds of prototypical terms:

- real-prototypical terms – prototypical terms that *can have* spatiotemporal referents
- ideal-prototypical terms – prototypical terms that *necessarily lack* spatiotemporal referents.

Terms for engineering prototypes (‘the new Ford car’, etc.) are real-prototypical terms; so are the terms for the basic units in the SI-system, and so are terms such as ‘typical lion’, ‘typical cell’, and ‘typical red’. All the prototypical terms that children first learn without pictures have of course to be real-prototypical terms, since what is talked about has to be perceivable. For the same reason, all the terms in old non-evolutionary plant and animal classifications are real-prototypical; it was taken for granted that there exist prototypical exemplars of all species. And, to end the list, the same goes for medical anatomies; when this feature is being stressed, one talks of ‘canonical anatomy’.

Ideal-prototypical terms, on the other hand, are such that the spatiotemporal world cannot possibly contain anything that directly corresponds to the term in question. The referent is a kind of fiction; in science it is called an ‘ideal type’ or an ‘idealization’. When modern physics entered the scene, it brought with it this kind of conceptual construction. Galileo made important thought experiments in which entities thought to be impossible were referred to by terms such as ‘frictionless planes’ and ‘vacuum’. And later physics has followed suit with notions such as ‘ideal gas’ and ‘absolutely elastic collision’. In microeconomics, the so-called ‘economic man’ is an ideal type that cannot possibly have any prototypical instances. In the time span required, no human being can handle and make calculations with all the information that such a man is assumed to have. In analogy with our earlier figures, Figure 6 illustrates what it means to be an ideal-prototypical term. Just as we have a capacity to apply real-prototypical terms to non-prototypical instances, so we also have a contrary linguistic capacity to move from what we perceive (somewhat rational behavior) to something we cannot possibly perceive (‘economic man’).

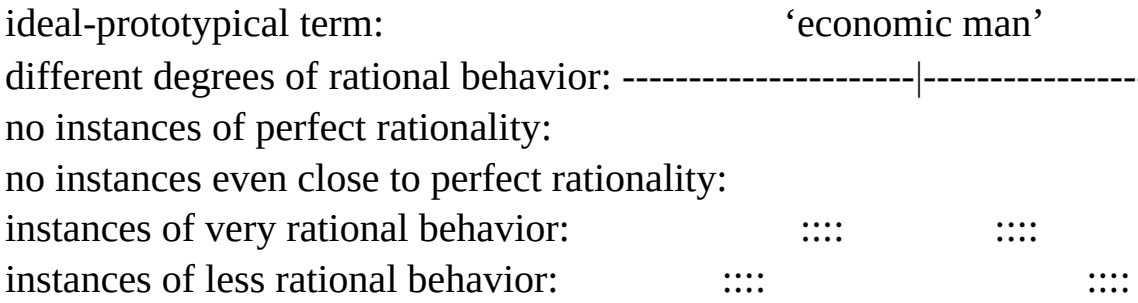


Figure 6: *Illustration of what it means to be an ideal-prototypical term.*

There is nothing odd in calling economic man a prototype, since we can, despite his fictional existence, measure real human behavior in relation to this benchmark in a way similar to the way we measure physical properties in relation to the relevant measuring units (1 m, 1 kg, etc.).

In itself, neither the idea of real-prototypicality nor the idea of ideal-prototypicality need to involve any thoughts about what is right or normal in a normative sense. In the history of biology, however, fusions of prototypicality and normativity have been rather the rule than the

exception. Linnaeus, for instance, who thought that God had created all the species, considered non-prototypical exemplars as deviant and imperfect. They were not as they should be. Other reasons for the common fusion of prototypicality and normativity might be the fact that prototypes in engineering mostly represent something planned or wished for, and the fact that it would be odd (but not logically impossible) to take sick human beings as measuring units in medicine.

Often, the old fusion of prototypicality and normativity in biology is still present in the view that what is prototypical is ‘natural’, and what is natural is as it ought to be. However, even putting the normative aspect aside, it can be difficult to see what terms such as ‘natural’, ‘natural kind’ and ‘nature-given’ mean exactly, and in what way they are contrasted with what is ‘artificial’. The next section is intended to clarify some of these semantic issues.

11.1.6 The contrast between the natural and the artificial

Many kinds of property-bearers are simply found in nature, but others are due to human invention and intervention; tools and machines are invented, hybridized plants and animal breeds are products of intervention, whereas genetically modified animals are so to speak both inventions and interventions. The former group of property-bearers can be called ‘*natural* objects’ and the latter ‘*artificial* objects’ or ‘artifacts’.

On this division, even classes such as (i) ‘the class of lions that weigh 200 kg or more’, (ii) ‘the class of mammals whose adults are on average taller than 1.23 m’, and (iii) ‘the class of animals that have red fur’ come out as classes of natural objects, since the members of these classes have been found in nature. The *boundaries* for the latter three classes, however, are a matter of human arbitrary decree; for example, lions that weigh 200 kg are more similar to those that weigh 199.9 kg (which belong to a different class) than those who weigh 250 kg (which belong to the same class). The traditional boundaries between lions and non-lions, mammals and non-mammals, and animals and non-animals, on the other hand, rely on clusters of similarity relations between phenotypic characters that make these traditional boundaries look much more natural than a boundary such as (i) ‘weighing 200 kg or more’. Therefore, it is useful to make a further distinction between artificial and natural classifications of natural objects,

and distinguish between classes of natural objects with a *natural* boundary and classes with an *artificial* boundary, respectively.

Among the classes with a natural boundary, species and clones stand out as being special in the way earlier explained; they are *nature*-given classes in contradistinction to language-constituted (*artificial*) classes. We hope that the exposition in Figure 7 will make visible the ambiguities latent in a context-free term ‘artificial’.

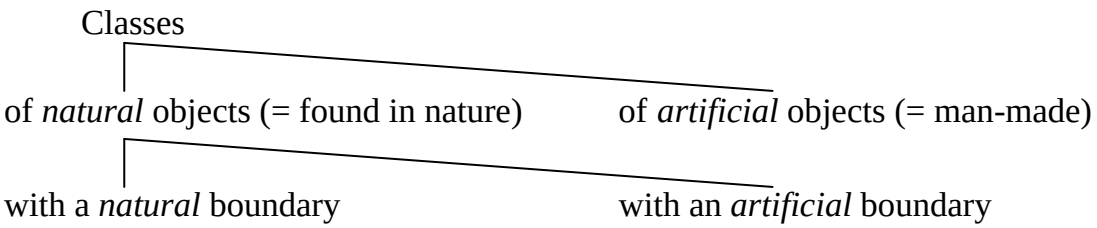


Figure 7: *The term ‘artificial’ and its opposites.*

11.1.7 Partonomy

The anatomy of the human body relates the class of hearts, the class of livers, and so on, to each other. But even though class terms are used, it is not a classification of classes in a taxonomic tree of classes, subclasses, and sub-subclasses of the kind we have spoken about earlier in this chapter. The proper overarching label is ‘partonomy’; and anatomy is only one among several partonomies in science. In anatomy the spatial and functional parts of the human body are distinguished and classified; in chemistry the spatial parts of molecules are distinguished and classified. When classes are systematically related to each other by means of subsumption, we arrive at a taxonomy; and when a whole of a certain kind is systematically and exhaustively partitioned, we arrive at a partonomy. Normally, the term ‘anatomy’ (from Greek ‘anatemnein’, cut up or open up) is reserved for partitions of biological entities. That is, there are animal anatomies, plant anatomies, and cell anatomies, but no ‘molecule anatomies’ or ‘atom anatomies’. Therefore, the term ‘partonomy’ is needed; in informatics, the parthood relation is often written ‘A part _of B’.

There are not only spatial partonomies. In relation to processes, temporal partonomies are possible. Phylogenetic systematics is wholly devoted to such a study, the process of speciation. In medicine, one often finds simple temporal partonomies such as the distinction between the diastolic and systolic phases of the beating of the heart. Some diseases have a development that is commonly divided into temporal parts or stages. For instance, syphilis is divided into primary, secondary, and tertiary syphilis, and the initial course of multiple sclerosis is divided into relapses and remissions. In the study of the DNA molecule and its role in reproduction, both a spatial partonomy of the molecule and temporal partonomies of the processes these parts are involved in are constructed. We will in this subsection restrict our philosophical remarks to anatomy. The distinctions between (i) particulars and repeatables, (ii) nature-given and language-constituted classes, and (iii) nominal and prototypical terms will again be useful.

Whereas it is possible to subsume repeatables and the corresponding classes in taxonomies without bothering about their instances in space and time (e.g., ‘red is a color’, ‘the lion is a mammal’), it is impossible to partition repeatables as such – only their instances in space can be partitioned. However, this is of minor importance since a repeatable is in all its instances identical. At once, let us put completely fiat spatial partitions aside. Of course, one can draw an arbitrary boundary around any possible area or volume within a human body, whatever curious shape the boundary may have and whatever organs it may dissect, but this is uninteresting from a purely scientific point of view, even though it might be extremely important in specific surgical situations. When human bodies, cells, or molecules are divided into parts, researchers have at least one of their eyes on what we have called nature-given classes.

Every instance of a repeatable must have a spatial boundary. Where there is a nature-given boundary for such an instance, i.e., a bona fide boundary, there is a discontinuity in nature. Both in complete spatial homogeneities (such as a sea with the same kind of water) and in spatial property continua (such as the rainbow) there can, by definition, only be fiat boundaries. Where there is a bona fide boundary, there is one kind of thing on one side of the boundary and another kind on the other side. This means that almost all anatomical entities are given a *partly* fiat boundary,

since even if for example the heart is regarded as mainly having quite a distinct nature-given boundary, the boundaries between the heart and the aorta and the pulmonary artery cannot possibly be regarded as constituted by a discontinuity between the heart muscle and the arteries. The heart is an organ with a partly nature-given and a partly fiat (language-constituted) boundary. And a similar observation can easily be made in relation to the other organs; all of them have some kind of ingoing and/or outgoing vessels. The fiat boundary component is of course larger and even more obvious in partitions such as ‘the upper part of the right ventricle’ and ‘the lower part of the right ventricle’.

Now, as we all know, looked at from an airplane there seems to be a stable discontinuity between sea and shore; but if we come close to such a boundary, there is because of the continuously moving water no distinct line between very wet sand and the sea. Analogously, when we are looking at a forest from a distance, we see a distinct discontinuity where the forest begins, but coming close we see only a number of trees with spaces between them. And what goes for a forest when approaching it in ordinary life goes for anatomical organs when approached by means of a microscope. The boundaries we normally see disappear. Such facts, however, do not make the distinction between nature-given and conventional boundaries inapplicable, but it makes the existence of nature-given spatial boundaries dependent on another distinction, one between levels or strata of reality. We will briefly comment on this problem at the end of Chapter 11.1.8.

We have already described how both nominal and prototypical terms such as ‘red’ and ‘typical red’ can be vague in the sense that the semantic content of the terms do not contain any definite boundaries, and that therefore there may be application problems at the semantic fringes. Another kind of vagueness can appear in relation to instances that surely fall under a certain term. Think of the term ‘mountain’. Even if one knows that something for sure is a mountain, one may nonetheless not know exactly where to draw the boundary between the mountain and the valley. Such vagueness afflicts many anatomical terms. Where does an organ end and its environment start? Mostly, no definite boundaries have been stipulated by the medical community. Nonetheless, physicians can easily communicate about an organ even if they have drawn the fiat part of the

boundary in question somewhat differently. The bona fide part of it seems to be enough. If one physician says to a colleague ‘Now I can see the whole jejunum’, this assertion can be true for both of them even if they draw the exact boundary for the jejunum somewhat differently, or simply leave open the question where the borders to the duodenum and the ileum are situated. Quite generally, in most conversations people do not care much about exact spatial boundaries. Surgeons, however, have to work with strictly definite boundaries. But in the same way that normal language users can move back and forth between more or less specific terms such as ‘very tall’, ‘1.96–1.98 m’, and ‘1.97 m’, surgeons can move between terms referring to spatial delimitations of the jejunum with different degrees of precision.

Traditionally, gross anatomy has used real-prototypical terms when talking about the organs of the body. A prototypical human hand has five fingers, but there are several (non-prototypical) humans with hands that have four, four and a half, five and a half, and six fingers. And similar remarks apply to many other anatomical parts. To repeat: a prototypical term refers to one repeatable (the prototype) and the corresponding class directly and some other repeatables (that are more or less similar to the prototype) and their classes indirectly. As a whole, the anatomical prototype comes with what in physics would have been called two ‘isotopes’. Isotopes of atoms differ with regard to some of their parts, and so do males and females.

It can be noted that biological prototypes are relatively independent of size. If the partonomy is constructed only by means of shapes and spatial patterns, then size can be disregarded, since shapes and spatial patterns are in principle size-independent. For instance, a circle is a circle, and zigzag pattern is a zigzag pattern quite independently of whether their size is small or large. Therefore, the classical anatomies of the human body can be used for both adults and children.

Anatomies are not just prototypical; they are in a definite sense at least *doubly* prototypical. First the body as a whole has a prototypical number and kinds of parts, and then each kind of parts has a specific prototypicality; first the prototypical human body has two hands, and then the prototypical hand has five fingers.

Prototypical terms delineate classes of particulars, and partonomies partition classes even when they are constructed by means of prototypes. To be a heart, a lung, a liver, etc., is to be a member of a certain class. Trivially, the part-terms spoken of in partonomies for classes can be used to classify particulars; as in ‘this is a heart’, ‘this is a lung’, and ‘this is a liver’. They can even be used in subsumption classifications. That is, taxonomies may well make use of partonomies. This is not just a logical possibility; it is in fact often the case in zoology and botany as well as in chemistry and physics. For instance, the subphylum ‘vertebrata’ (of the phylum ‘chordata’) is delineated by means of the fact that its members have a backbone or spinal column as part. The Linnean so-called ‘sexual classification’ of plants relies to a very large extent on what ‘sexual organs’ the different kinds of plants have. And, to continue, isotopes of atoms are not classified according to their chemical properties, but according to what parts (subatomic particles) they contain. The first isotope of hydrogen (protium; ordinary water) has a nucleus with one proton and no neutrons, the second isotope (deuterium; heavy water) has a nucleus with one proton and one neutron, and the third isotope (tritium) has one proton and two neutrons.

The subsumption relation of taxonomies (i.e., all the members of one class are members of another class) is a relation between classes (or repeatables), not between the members of the classes. This is one feature that makes the subsumption relation distinct from the part-whole relation, which primarily is a relation between two particulars (the term ‘part’ is here used in its ordinary sense, i.e., the part can by definition not be identical with the whole of which it is a part; in formal mereology such parts are called ‘proper parts’). However, as the existence of partonomies shows, there can, secondarily, be part-whole relations also between classes. There are three kinds of such relations:

1. Each member of class A is part of a member of class B, but all members of B do not have a member of A as a part. *Example*: all human eyes (A) are parts of human organisms (B), but there are people who have lost their eyes. Call it ‘part-to-whole parthood’ (or ‘part_of’).

2. Each member of B has a member of A as a part, but all members of A are not parts of a member of B. *Example*: all water molecules (B) have an oxygen atom (A) as a part, but oxygen atoms are parts also of many other kinds of molecules. Call it ‘whole-to-part parthood’ (or ‘has_part’).
3. Each member of A is part of a member of B, and each member of B has a member of A as a part. *Example*: the human circulatory system (A) and the human organism (B) are always coexisting in this way. Call it ‘mutual parthood’.

All three relations can figure in definitions. This requires some further remarks. Let us start with whole-to-part parthood (2): if the water molecule is *defined* in such a way that it has to contain an oxygen atom, one might say that a water molecule inherits part of its identity from one of its parts. Next part-to-whole parthood (1): if the human eye is *defined* in such a way that (i) a human eye outside the body is not a human eye, and (ii) an animal eye transplanted into a human body should be called ‘human eye’, then one might say that a human eye inherits part of its identity from a whole of which it is a part. If one wants to claim that the human circulatory system and the human organism are *necessarily* interdependent (mutual parthood; 3), then one has to claim either that it is in principle impossible to transplant the circulatory system of an animal into a human being, or that such a transplant should be labeled ‘human circulatory system’.

When there is a part-whole relation between only some members of A and some members of B, i.e., neither class has all its members involved, we arrive at a fourth case:

4. Some but not all members of A are parts of members of B, and some but not all members of B has a member of A as a part. *Example*: some pig hearts (A) are now, after these transplantations, parts of human organisms (B). Call it ‘accidental parthood’.

The prototypical (canonical) anatomies of traditional medical textbooks have mainly used the relations part-to-whole parthood (1) and mutual parthood (3) in their constructions, but the development of medical technology will certainly bring accidental parthood (4), more to the fore.

Famously, Aristotle said that a hand cut off from the body is a hand in name only; meaning that it is no longer a real hand (part-to-whole parthood). He could just as well have used the kidney as his example. But, today, there are many very real kidneys outside a body; they are just on their way from one body to another. The relation whole-to-part parthood (2) comes to the fore in the next section.

11.1.8 Beyond taxonomies and partonomies: ontological pluralism?

Philosophy asks ultimate questions such as 'where and how does space *as a whole* end?' and 'where and how does *all* moral justification come to an end?' (see Chapter 9). Analogous questions can be asked in relation to taxonomic and partonomic trees:

- where (when we move 'upward' towards the top of the tree) do taxonomies and partonomies *necessarily* end?

In both cases, as we will show, there is an answer that seems rather trivial but can be questioned.

Let us start with taxonomies. Linnaeus stopped at the 'kingdoms' of inert matter, plants, and animals. Philosophers go further. Some of them distinguish, as we have already done, between kinds of property-bearers (in what follows: 'substances') and properties (in what follows: 'qualities'), and can then claim that all the three kingdoms mentioned are subclasses of the class of substances, not of the class of qualities. The taxonomic trees mentioned contain classifications of non-processual entities. So far, we have only mentioned processes in relation to the notion of 'temporal *partonomies*' and some of its instances such as phylogenetic systematics, but processes can be *taxonomically* ordered as well (see Chapter 11.2.1). Now we will philosophize a little about the process–non-process distinction.

Instances of substances and qualities can retain their identity over time; instances of substances can retain their identity even when they undergo changes of properties. Living beings, for instance, retain their identity not only through the changes implied in the process of natural growth, but also through other, sometimes drastic changes; not even the amputation of a limb takes away the identity of an animal. At each and every temporal

point, human beings, lions, and cells exist with their so to speak full individual identity. Just as instances of substances are bearers of qualities, so also can such instances be bearers of processes, both long-lasting and brief ones. But this fact does not mean that instances of substances *are* processes – just as the fact that they are property-bearers does not mean that they are merely a bundle of properties.

At no temporal point does a certain spatiotemporal process contain its full individual identity. This is because processes exist by unfolding their successive *temporal parts*. Plants and animals undergo life processes, which have temporal parts such as youth or old age, but the bearers of these processes can retain their identity both despite and because of the processes. Some philosophers think that individual property-bearers can have their identity quite independently of any natural kind (substance) they instantiate, i.e., these philosophers claim that there are 'bare particulars'. We do not. This means that we find it odd to think that if there is continuously in a temporal interval a certain natural kind, then it is not exactly one and the same enduring instance of that kind all the time. Of course, it makes good sense to talk about fiat (language-constituted) entities such as 'the lion during the summer', 'the cell between 7:52 and 7:54 pm', and 'the red color instance at x between t_1 and t_2', but this does not mean that the identity of the lion, the cell, and the color instance are divided into temporal parts. To the contrary, the grammar of these sentences ('*the* lion … during …' and '*the* cell/color-instance … between …')' indicates that the particulars in question are assumed to retain their identity even outside the temporal interval spoken about.

These remarks seem to imply that we can take another step upwards on our taxonomic ladder. First, we bring substances and qualities together under one taxon, 'enduring entities'; and we bring processes and other kinds of entities that necessarily have temporal parts (e.g., activities and changes) under another taxon, 'occurring (perduring) entities'. The nineteenth century medical distinction between anatomy and physiology implies in this terminology, that anatomists deal with enduring entities and physiologists with occurring entities. In accounting, 'stocks' should be classified as enduring entities and 'flows' as occurring entities. Second, we bring the last two taxa together under the heading 'temporal entities', contrast them with 'abstract entities' such as numbers, and at last reach the

absolute end with ‘entities’. Is there any problem with the classification in Figure 8? Yes, there is.

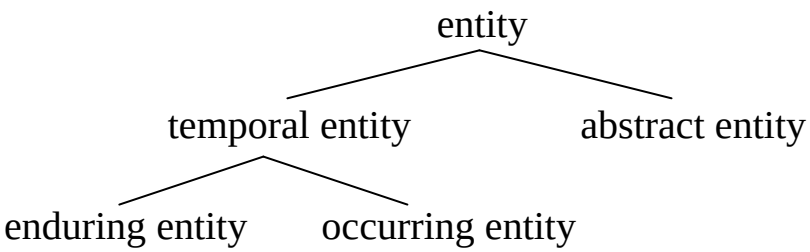


Figure 8: *A top-level taxonomy* (a ‘philosophical ontology’; Chapter 11.3) *which is problematic*.

It seems as though everything that empirico-scientific taxonomies classify ought to exist in the same way, i.e., for any given point in time and any given existing class in a taxonomy, either the class should have a member or not. But since processes necessarily have temporal parts and are extended in time, they cannot possibly exist *at* a single point of time; taking a certain ‘now’ as given, they always bring in either the past or the future or both. According to common sense, however, since neither the past nor the future exists, at least not in the same strong way as the present, processes cannot possibly exist in the same way as substances and qualities can. Therefore, perhaps we should end with at least two top-level nodes, one for enduring entities (with their specific way of existing) and one for occurring entities (with their specific way of existing).

As indicated by the crucial role played by the term ‘temporal part’ in the above, this problem involves in a profound sense the question of what time really is. We do not intend to enter into the philosophical issue of ‘presentism’ (which holds that only ‘the now’ truly exists) and ‘four-dimensionalism’ (time is in its essence like space, and the apparent priority of ‘the now’ is an illusion), but we wanted to mention it, since it may perhaps one day become necessary even for information scientists to go beyond taxonomies with their classifications of *kinds* of existing things and take into account also *ways* of existing – and bring in *ontologico-taxonomic pluralism*.

Now that we have addressed the question of where taxonomies stop when we move upwards, let us turn to the problem of the limits of

partonomies. Our question is: ‘where do we find something that cannot be a part in a larger partonomy?’ The seemingly trivial answer is: the whole universe. But this answer is beside the point. There is (we take it for granted) only one universe, but we are discussing partonomies for classes of entities. Therefore, to start with, we have to ask what it can mean to relate one partonomy to another partonomy. When answering this question, it is important not to conflate, or not to connect too strongly to each other, the following four relations: (spatial) *whole-to-part* parthood, *later-to-earlier* (in evolution), *larger-to-smaller*, and one-sided existential dependence or *dependent-to-independent*.

Big Bang cosmology and evolutionary biology delivers a temporal picture of the development of the Universe that to some extent has a spatial counterpart. According to science, the world has evolved as follows: first subatomic particles, then atoms, then molecules, then stars and planets, then cells, and then organisms. At the end of the last subchapter, we distinguished between four kinds of parthood relations. One of these, the spatial whole-to-part parthood, does sometimes run parallel with the relation of later-to-earlier in the evolutionary picture:

- organisms always have cells as parts
  (and came in evolution later than the first cells)
- cells always have molecules as parts
  (and came in evolution later than the first molecules)
- molecules always have atoms as parts
  (and came in evolution later than the first atoms)
- atoms always have subatomic particles as parts
  (and came in evolution later than the first subatomic particles).

This means that the partonomies of the different atoms (hydrogen, helium, lithium, etc.) can be connected to the partonomies of all the different molecules – and so on upwards to all the different cells and organisms. Note that the converse partonomic relation of *part-to-whole* cannot be used to create such a hierarchy of connections. Atoms are not always parts of molecules. They can form aggregates of their own; pieces of metals are aggregates of atoms, not of molecules. Similarly, molecules are not always parts of cells, as is amply testified to by all inert matter.

And cells are not always *parts* of organisms; there are unicellular organisms, too.

Note also that the list proceeding down from organisms to the level of subatomic particles does not in general mirror a relation of larger-to-smaller. Organisms consist of cells, but some kinds of single cells are larger than some multi-cellular organisms; a one-cell ostrich ovum can have a diameter of 120 mm. All molecules consist of atoms, but many atoms are larger in size than the smallest molecule, the diatomic hydrogen molecule. Subatomic particles can be momentarily connected (‘entangled’) over astronomical distances, but atoms and molecules cannot. In the same vein, it can be observed that single scientific partonomies can posit parts whose size varies considerably. The partonomy of the circulatory system of the human body contains as parts both the aorta and the capillaries. Our solar system contains as parts both the planet Jupiter, which has an equatorial diameter of ca 140,000 km, and asteroids with a diameter of less than 1 km.

The whole-to-part parthood relations that obtain between organisms, cells, molecules, atoms, and subatomic particles might be seen as merely expressions of natural laws that state that certain kinds of wholes of physical necessity require certain kinds of parts, but some philosophers think there is more to be said. They think that instances of the wholes in question can also be bearers of so-called ‘emergent properties’, i.e., property *dimensions* that the parts of the whole cannot be bearers of. Let us give examples. Organisms, cells, molecules, and atoms can all have determinate values of property dimensions such as size, mass, shape, and electric charge, but only organisms and cells can have the properties of engaging in metabolism and reproduction. The mass (weight) of an organism is simply the sum of the mass of its parts, and the same goes for the mass of cells, molecules, and atoms in relation to their parts; mass is *not* an emergent property. Metabolism and reproduction, on the other hand, do simply not exist as properties, capacities, or qualities of molecules and atoms.

Both the determinate mass (non-emergent property) and the reproductive capacity (emergent property) of an individual cell have in common the fact that they are dependent for their existence on the molecules of the cell. But there is also a difference. In the mass case, the dependence in question is a

relation between determinates of the same property dimension (the value of the mass of the cell depends on the values of the masses of the molecules), but in the reproduction case it is a relation of existential dependence between a property dimension (e.g., the capacity of metabolism) and some other property dimensions (those inhering in the molecules). Therefore, it is argued, organisms and cells have to be regarded as existing on another stratum or *ontological level* than molecules and atoms do. Such a view implies that there is more to the old distinction between living beings and inert matter than merely a number of complicated spatial parthood relations.

Similarly, the claim that the molecule-atom distinction is a distinction between ontological levels amounts to claiming that some chemical properties or chemical bonds (or both) are not merely the sum of the properties and forces that are assumed to exist on the atomic and subatomic levels. The mainstream view today is that all known chemical bonds can be explained by quantum mechanics, but since approximations and simplification rules are used in the derivations of these chemical bonds, some philosophers of chemistry have argued that, first theoretical appearances notwithstanding, some chemical bonds are emergent properties, i.e., they are not reducible to a number of more fundamental forces.

Moreover, in Chapter 7.5 we presented arguments around the question whether or not mental phenomena (qualia and intentional states) can be regarded as identical with brain states. If they cannot, they are examples of emergent properties, too. Such a view does not imply that mental phenomena can exist without organisms with brains. It implies only that there is a relation of (one-sided) existential dependence of the mental on some biological substrate. That is, the mental (the upper ontological level) cannot possibly exist without a brain (the lower level), even though there might be brains that do not give rise to any mental phenomena. Now, how is this dependent-to-independent relation related to the spatial whole-to-part relation? (Note that ‘independent’ only means ‘relatively independent’, as is clear from the list below.) It is hard to find any general spatial whole-to-part relation between mental states and brain states. Are dreams larger or smaller than the brain? Are veridical perceptions larger or smaller than the brain? Are these questions meaningful at all? Therefore, it

seems as if relations of existential dependence need not necessarily run parallel with a corresponding spatial whole-to-part relation, even though this is so in the cell-to-molecule case. Leaving all parthood relations aside, we might claim that:

- mental phenomena are existentially dependent on organisms, but not vice versa
- organisms are existentially dependent on cells, but not vice versa
- cells are existentially dependent on molecules, but not vice versa
- molecules are existentially dependent on atoms, but not vice versa
- atoms are existentially dependent on subatomic particles, but not vice versa.

In such a philosophico-ontological hierarchy of irreducible strata, an upper ontological level cannot start to exist before the lower ones exist; either it has to start to exist simultaneously with the lower ones, or come after them in time. According to contemporary evolutionary theories, the latter is in fact the case: first subatomic particles, then atoms, then molecules, then simple forms of life, then more complex forms of life, and then mental phenomena.

The saying ‘he can’t see the forest for the trees’ relates to everyday perceptual life; in the present context it can be exchanged for ‘there are philosophers and scientists who can’t see the upper level for the entities on the lower level’. One thing that is taken to tell in favor of the view that a forest is no more than an aggregate of trees is the fact, that if we come very close to where a forest begins, we see no forest boundary but only a number of trees with spaces between them. Similarly, under a powerful microscope the cell boundary disappears in favor of molecules with spaces between them. How to look upon this boundary problem? One way is to stop taking it for granted that there is a theory-free and level-free way of identifying spatial boundaries *of specific entities*. The fact that we can delimit spatial regions by means of arbitrarily thin and arbitrarily curved boundaries needs not imply that the same is true of the boundaries for specific kinds of entities. Perhaps boundaries-for-entities have to be regarded as level-relative in the sense that there is one way of drawing boundaries for trees and another for forests; one way of drawing

boundaries for cells and another for molecules; and so on for other ontological levels. Perhaps, where there are emergent properties, there has to be an emergent kind of spatial boundaries, too.

It is hard to predict the future of science, but it is not unconceivable that one day some information science will have reason to bring in what might be called '*ontologico-partonomic pluralism*', i.e., produce systematic accounts of the way the world is, according to science, divided into strata or ontological levels.

After these speculations about what might be close to but nonetheless outside taxonomies and partonomies proper, let us return to taxonomy.

11.2 Taxonomy and the philosophy of science

As we said at the beginning of this very chapter, classifications surround us all the time, and science takes informal and scattered classifications and develops them into taxonomies. But these scientific taxonomies have not been given much attention by philosophers. Taxonomy is basic to science, but neglected in the twentieth century philosophy of science. Why? Probably because taxonomic work was regarded as a proto-scientific activity that is not in need of philosophical elucidation. What, on the other hand, was regarded as being a truly mature scientific activity in need of philosophical elucidation was the empirical testing of both paradigms and normal-scientific hypotheses. Since modern science has amply shown how hard it is to know that a theory is completely true, it is by no means odd that philosophers have focused on the epistemological question of how to judge theories in the light of empirical evidence. It is quite natural that the philosophy of science has been much concerned with problems around induction, abduction, and hypothetico-deductive reasoning – nonetheless, taxonomy has to be given its due, too.

11.2.1 Subsumptions, orderings, and specializations

In the empirical sciences, taxonomy brackets epistemological questions and asks how various kinds of entities are related to each other on the presupposition that what is regarded as being true of them really is true. Taxonomical questions are ontological questions. A superficial look at taxonomy, however, may give rise to *the false impression that taxonomies are only matters of rather trivial convention*. If animals are classified

according to their type of habitat and their mode of locomotion, then whales and fishes should be lumped together; if they are classified according to their mode of reproduction and the source of nourishment for their offspring, then they should be kept apart. Whales are then mammals not fishes. Does this not show that taxonomic work is a rather simple enterprise, where entities are lumped together or kept apart depending on what similarity relations the taxonomists choose to focus on? No, it does not.

As the ‘whales-as-fish’ and ‘whales-as-mammal’ example is presented above, the problem seems to be only in what respects whales are like fishes and in what respects they are like mammals; and since they are similar to mammals in one respect and to fishes in another, it is merely to choose one or the other alternative. However, this is too narrow a perspective. The problem is to relate all animals in such a way that we obtain a class-subsumption hierarchy of some sort; either like the one presented in Table 1 below, or in some other way. It is only for the sake of expositional simplicity that in Table 1 there are four ranks and that each class is divided into exactly two subclasses (as Plato would have it).

Rank 1	class A(1)							
Rank 2	class A(2)				class B(2)			
Rank 3	class A(3)		class B(3)		class C(3)		class D(3)	
Rank 4	class A(4)	class B(4)	class C(4)	class D(4)	class E(4)	class F(4)	class G(4)	class H(4)

Table 1: *The formal structure of one possible class taxonomic schema based on subsumption.*

Taxonomists should strive to make all the classes on one rank exhaust the class on the next over-lying rank. Otherwise, there is a great danger that they will have to revise the definition of the upper class if they later want to incorporate a newly discovered class that belongs to the lower rank. However, this danger cannot be completely taken away. Fallibilism has to be kept in mind even here. Science progresses not only through theoretical changes, but also through taxonomical changes. If a new natural

class is suddenly discovered, then this event may have more or less radical repercussions on the existing taxonomy. And such discoveries have been made a number of times in the history of biology. For instance, when it was discovered that there are some non-mammals that also feed their offspring with milk, the traditional definition of mammals (as animals where the females feed the offspring with milk) had to be substituted by one that says that the females have mammary glands.

Nature-given classes are by definition mutually exclusive; otherwise they would not be classes given by nature. Language-constituted classes, on the other hand, can be constructed as being either overlapping or not, but from a communicative point of view it is important that even such classes are mutually exclusive when they belong to the same rank. Systems with mutually exclusive classes contain more information and are therefore from a communicative point of view more efficient. This point can easily be seen if one remembers the interplay between *classifications of classes* and *classifications of particulars*. If the classes are constructed as being mutually exclusive, one knows for sure that if one person says ‘this is an A(3)’ and another person says of the same thing ‘this is a B(3)’ both cannot be right (see Table 1); if, on the other hand, the classes are overlapping, then both might be right.

When taxonomists create a taxonomy that contains several classes and ranks, and which fulfills the requirements that (i) all classes belonging to the same rank are mutually exclusive and (ii) together exhausting the rank above, then the constraints on what properties they can use as classifying criteria become rather restricted. They are then far away from the isolated question ‘should we lump whales together with fishes or with mammals?’ If ‘living-in-water-and-moving-by-swimming’ is used as a taxonomic criterion, then ‘living-on-land-and-moving-by-walking’ would be natural as another criterion in the same rank, but this gives too broad a class to fit a taxonomic hierarchy.

To construct a taxonomy is very much like solving a jigsaw puzzle. There are many pieces that should fit into each other. If one has only three pieces, and one of the pieces (‘whale’) fits both of the other pieces (‘fish’ and ‘mammal’), then it is a matter of taste what two pieces to put together, but if the three pieces are part of a much larger puzzle, then things are more complicated. All pieces of the puzzle should be fitted together and,

therefore, one cannot be sure where the whale-piece really fits until the whole puzzle is solved; see Table 2. However, as it is *in principle* possible that the pieces of a jigsaw puzzle might be put together in more than one way, it is in principle possible that a certain domain of the world may be exhaustively classified in more than one way (also: in one way by different criteria). Whether or not this is the case or not is for nature to tell us is in each individual case.

Phylum	Chordata							
Subphylum	Vertebrata				class B(2)			
Class	Mammals		Fishes		class C(3)		class D(3)	
Genus	Mammals a – m	Whales n	Fishes a – g	Fishes h – n	class E(4)	class F(4)	class G(4)	class H(4)

Table 2: *A simplified class-subsumption schema with whales and fishes.*

The upshot of our remarks so far in this section is that taxonomies are neither purely conventional constructs nor easily constructed. They are partly based on pre-existing similarity relations, and to construct a good taxonomy for a whole domain can be as hard as to invent a mathematical model for some unobservable entity in physics.

Part of the scientific enterprise is to find out what kinds of parts various entities contain. In this vein it has been shown that the human body consists of bones, muscles, veins, and organs, that organs consist of cells, cells consist of molecules, molecules consist of atoms, and so on into the subatomic mysteries. Now, for example, when molecules are shown to contain atoms, it may seem as if the taxonomy of molecules becomes an easy affair. When one has found out what atoms constitute different kinds of molecules, one can use the resultant molecule partonomies instead of traditional chemical properties as a basis for a taxonomy of molecules. But then one needs a taxonomy for atoms. And, to go one step further, if atoms are classified according to what number and kind of sub-atomic particles they consist of, then a taxonomy for sub-atomic particles is needed. This regress must end at some point where the taxonomy is not based on a partonomy; for logical reasons, science has at every point in its evolution

to contain (at least) one taxonomy that is not based on a partonomy. Normally, science seems to rely on many such ground-taxonomies. Furthermore, even when a taxonomy based on a partonomy is constructed (e.g., molecules on atoms), the old properties (chemical properties) and the corresponding taxonomy may still be useful.

Taxonomies that are not based on partonomies rest on similarity relations. All the various mammals are similar in that the females (to connect to the old definition) feed the offspring with milk, or (to take the modern one) in that they have mammary glands as parts of the body. But this does not mean that all mammals feed their offspring in exactly the same way, or that the mammary glands of different mammals look exactly the same. Similarity relations can be more or less specific. Sometimes similarity in a certain respect even allows for degrees (*a* is more similar to *b*, than *b* is similar to *c*) in the strong sense of a linear ordering. It is then possible to construct numerical scales such as the length scale, the mass scale, and the temperature scale.

When we say things such as ‘this room is warm’ or ‘the temperature of this room is 22 ^{o}C’, we are making a *classification of a particular* (the room), but both the terms ‘warm’ and ‘22 ^{o}C’ have in isolation a *repeatable* and a whole *class* as referents. Furthermore, the term ‘22 ^{o}C’ belong to a very systematic classification of temperature classes, namely a temperature scale. From a broad perspective, scales might be said to constitute a kind of taxonomies, since they systematically relate repeatables and classes to each other by means of similarity relations. Mostly, however, scales have been discussed as taxonomies neither in theoretical taxonomy nor in the philosophy of science. For historical reasons, the term ‘taxonomy’ has mostly been applied only to systematic classifications of natural kinds. Perhaps, this restriction of the term was also due to the fact that, unlike scales, classifications of natural kinds are non-quantitative. That is, there is a real difference between subsumption schemas and orderings, and we will now say some more words about it. Let us continue to use temperature as our example. Look at the taxonomy and the accompanying – partly subjective! – scale in Table 3 below.

Level 1	Temperature							
Level 2	minus temperature				plus temperature			
Level 3	too cold		acceptably cold		acceptably warm		too warm	
Level 4	extremely cold	very cold	cold	very cool	cool	warm	very warm	hot

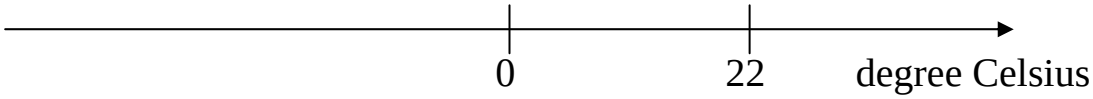


Table 3: *A formal class-subsumption schema and an accompanying temperature scale*.

Table 3 presents a class-subsumption schema that is formally similar to those in Tables 1 and 2, and an accompanying scale; the latter is impossible to construct in relation to Table 2 (Chordata). The scale represents the fact that the most determinate temperature repeatables can be ordered on a line as being lower and higher; temperatures represented by points to the right are higher than those represented by points to the left. Chordata can not be so ordered. The horizontally represented similarity relations between temperatures can very well, as the table shows, be used for constructing the vertical taxonomy of temperatures with its subsumption relations. In this taxonomy it holds true, to take one example, that ‘very cold’ is a subclass of ‘too cold’, that ‘too cold’ is a subclass of ‘minus temperature’, and ‘minus temperature’ is a subclass of ‘temperature’. In traditional taxonomies of substances and kinds, such (‘vertical’) subsumption relations are the only relations that make up the classificatory pattern, but in Table 3 there is also the lower-higher or left-right relation of the scale.

That there is a *continuous* scale beneath the subsumption schema means that all the slots in the schema above are drawn by fiat. Where there is a bona fide boundary there is a discontinuity, but in the case of temperature (and of many other physical dimensions) reality allows a continuum of repeatables. Therefore, we have to create, when needed, such discontinuities ourselves. It is like creating boundaries between countries

in the sea or in the desert, even if it is now a matter of keeping repeatables and classes apart, not of individual spatial regions.

Subsumption taxonomies as well as orderings (scales) are human constructions concerned with relations between repeatables and between classes. They encapsulate knowledge and make communication more efficient. When a particular is classified by means of a term that belongs to a subsumption taxonomy or a scale, one receives for free an incredible number of relations to other particulars that are classified by means of the same taxonomy or scale. If one knows that something's temperature is '22 ^{o}C', then one knows its temperature relations to all other temperature instances that are described by means of the Celsius scale. If one is shown a kind of animal that one has never seen before, and is told that it is a mammal, one knows at once about similarities and differences between this animal and others that one knows of.

In Chapter 11.1.8, we highlighted the distinction between kinds/properties and processes; the term 'process' takes in also changes and activities. Classes of processes allow for classifications and subsumptions that do not fit the subsumption schemas described above; neither are they orderings of the kind described. We will call some of them 'subsumption-by-specialization', or 'specialization' for short; their natural representational device is graph constructions. One possible formal skeleton of a construction that brings in three levels and four classes is to be found in Figure 9 (compare Table 1, which of course can be transformed into a graph, too).

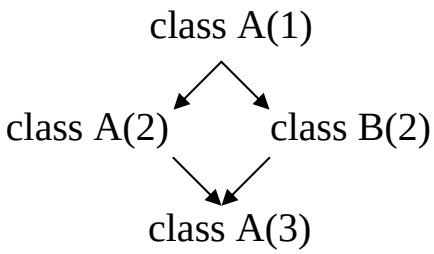


Figure 9: *The formal structure of a graph with a specialization 'diamond'.*

We will start with an everyday example before we turn to biological processes. Painting is an activity and thus a process. Let us call the class of painting processes 'class A(1)'. A process of painting is always directed at

certain kinds of objects or parts of objects. As we normally use the term ‘specialization’, we can say that one painter has specialized in painting houses and another in painting chairs, one in painting the exteriors of houses and another in painting the interiors. In order to give substance to the formal schema of Figure 9, we can let the class of painting-houses activities be class A(2), the class of painting-the-exteriors (of things) be class B(2), and the doubly specialized class of painting-the-exteriors-of-houses be class A(3). We then arrive at Figure 10. A similar structure has no proper place in a scale ordering or traditional subsumption tree, even though it contains subsumption relations. The class A(3) is s subclass of both A(2) and B(2), and both are subclasses of the class A(1). But here such a diamond structure is completely adequate, since processes belonging to the same class can be specialized in many directions simultaneously.

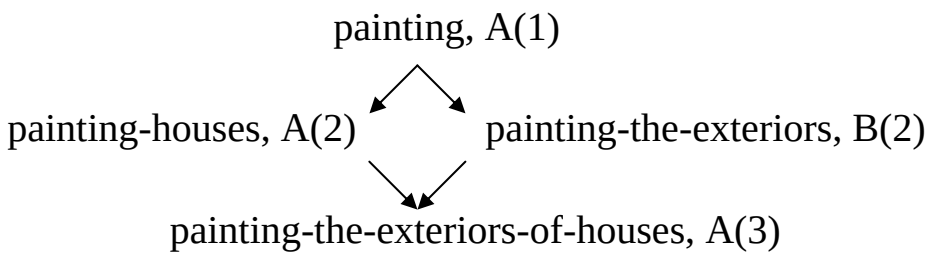


Figure 10: *Example of a graph with a specialization ‘diamond’*.

Some activities are activities only of a subject (e.g., swimming and running), whereas others also involve objects that are acted *on* (e.g., painting one’s house and driving one’s car). Similarly, some chemical processes simply occur in an object (e.g., rusting and burning), whereas others (e.g., digesting food and printing papers) act on one or several objects. It is only in the ‘acting-on’ kind of processes that classes of processes can have the specialization relation. When there is talk about painting, driving, digesting, and printing *simpliciter*, everyone knows that there is an object that has been abstracted away. It is this taken-away object that re-enters when a specialization is described. Nothing like this occurs in subsumptions of natural kinds and properties.

In the Gene Ontology (GO) for molecular functions one finds (hyphens added) ‘endo-deoxyribo-nuclease activity’ (GO:0004520) as a subclass to

both ‘deoxyribo-nuclease activity’ (GO:0004536) and ‘endo-nuclease activity’ (GO:0004519); both the latter, in turn, are subclasses of ‘nuclease activity’ (GO:0004518). These specializations are presented in Figure 11.

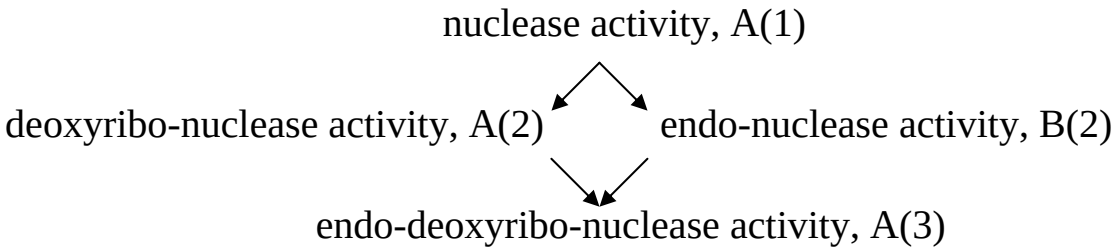


Figure 11: *A specialization schema with examples from the Gene Ontology.*

This is merely one of numerous examples of specializations that can be extracted from the GO. A nuclease activity is an activity (performed by an enzyme) that catalyzes hydrolysis of ester linkages within nucleic acids. Such activity can be specialized along at least two different directions: (i) according to *what* is acted on (deoxyribonucleic acid, DNA, or ribonucleic acid, RNA), and (ii) according to *where* the action takes place, i.e., cleaving a molecule from positions inside the molecule acted on (‘endo-’), and cleaving from the free ends of the molecule acted on (‘exo-’), respectively. Since nothing stops the specialization from going in both these directions at once, we arrive at the schema for ‘nuclease activity’ in Figure 11, which is completely analogous to the schema for the process of ‘painting’ in Figure 10; and both conform to Figure 9.

Specializations allow, and often require, the kind of diamond schema presented, which in its lower part contains so-called ‘multiple inheritance’, i.e., one subclass is presented as being a subclass of many (here two) classes. This ought not to be the case in taxonomies of natural kinds and properties, since (as noted in our remarks in relation to Table 1) this would reduce the informational content of the classification. In Figures 9-11 all subclass-class relations are specializations, but specialization relations can in graph constructions very well be combined with both ordinary subsumption and parthood relations. Even in such mixed cases can multiple inheritance be the normal and the required kind of subclass to class relation.

11.2.2 Taxonomy and epistemology

In direct taxonomical work epistemological questions are put aside; taxonomical work is ontological work. But this does not mean that there are no interesting relations at all between taxonomy and epistemology. Empirical science has two general epistemic sources, *observation* and *reason*. On the one hand there has to be some collecting of empirical data, be it in the form of direct observations, experiments, randomized control trials, or something else; on the other hand there has to be reasoning around the data, logical conclusions have to be drawn, and often some mathematical calculations are needed.

In some disciplines one makes an administrative distinction between scientists who are working theoretically and scientists who are working experimentally or empirically. One can then find sub-disciplines such as theoretical in contrast to experimental physics and theoretical in contrast to empirical sociology. Individual scientists on either line of such divisions may perhaps think that their work is independent of what transpires on the other side of the fence, but from a broader perspective this is not the case. As we tried to make clear in Chapters 2-4, all observations and experiments rely on some theoretical presuppositions, while at the same time no theory concerned with the world can be epistemically justified without empirical data. In cases like the ones just mentioned, *observation* and *reason* have only given rise to a division of labor between groups of scientists. One group works as if reason were the only general epistemic source, the other as if observation were the only such source, but it is quite clear that the scientific community as a whole needs both these sources. That being noted, two questions arise:

1. Does the scientific community *as a whole* need any other epistemic sources than reason and observation? – Answer: No.
2. Are there members *within* the scientific community that need any epistemic source beside reason and observation? – Answer: Yes.

Comment on point 1: It might be thought that tacit knowledge is a special source of knowledge, but this is wrong. Even though it is (as we explained it in Chapter 5) a kind of knowledge that is often important in

relation to both reasoning and making observations, it is not an epistemic source *besides* reason and observation. There is know-how *in* reasoning and *in* observing, but know-how is not itself an epistemic source in the sense now discussed.

Comment on point 2: The far-reaching division of labor in modern research gives rise to several kinds of dependencies. Research in the life sciences relies partly on knowledge that comes from chemistry, and research in chemistry, in turn, relies partly on knowledge that comes from physics. On the other hand, many experiments in physics rely on knowledge of chemical properties of various substances, and this knowledge comes from chemistry. Sub-disciplines within physics and chemistry take data from one another. Historical sciences such as archeology rely on natural-scientific methods such as radiocarbon dating ('C14-method'). But in its infancy, this method was verified by means of dating methods that had their origin in the humanities.

These observations show that modern research has for a long time been heavily dependent on *knowledge transmission* between various disciplines and sub-disciplines. Such transmission cannot work if the researchers do not trust each other. Therefore, most research groups within the scientific community have a third epistemic source, *trust in information* (from other researchers). As we said already when discussing arguments ad hominem in Chapter 4, even though the scientific community as a whole has only two epistemological sources, observation and reason, each individual scientist and research group has three:

- observation
- reason
- trust in information.

Let us now argue by analogy. In order to store information, retrieve information, and make information transmission efficient, good taxonomies and partonomies are needed. Of course, these latter do not in and of themselves create trust, but nor do microscopes and telescopes in themselves create observations, and nor do mathematical algorithms create reasoning. But they simplify observing and reasoning, respectively.

Similarly, good taxonomies and partonomies make the trusted information transmission simpler. We think the following three statements are true:

- Classifications are necessary for theory building, and taxonomies can make theory building easier; partonomies could even be regarded as kinds of theories.
- Taxonomic and partonomic work is not a trivial and merely proto-scientific kind of work.
- Good taxonomies and partonomies make it possible for the inter-scientific epistemic source ‘trust in information’ to function more efficiently.

11.3 Taxonomy in the computer age – ontology as science

Scientists have for a long time needed to store information outside of their brains. Without books and libraries science in the modern sense would not have existed. Up until the end of the nineteenth century, it might have been possible for some fast-learning individual scientists in disciplines such as physics, chemistry, and medicine to have grasped all the available knowledge in their field, but it would still been impossible for them to remember it all without the help of books. Today, however, in most disciplines the stock of knowledge is so great that it cannot be acquired by a single human individual during his lifetime; and it increases at such a rate that in many disciplines it is impossible to read in one day what is produced in one day.

To find both old and new knowledge in books and journals takes time. Both researchers and physicians have an efficiency problem well known from logistics in the world of trade: how to make the right thing (kind of knowledge) available in the right place (research group) at the right time? During the twentieth century, each university had a university library, and each department had its own disciplinary library. The latter saved the researchers the journey to the university library. Co-operation between the libraries made it possible to read – but with some delay – even books and journals that were not available in the place where the researcher or physician was working.

The computer revolution and internet have radically altered the time predicament for knowledge acquisition. Through the internet, information

stored in a computerized format can be made instantly accessible to individuals almost everywhere across the globe. Were it not for the simultaneous explosion of information, knowledge transmission would have become a simple enterprise. Today, sitting in their offices, researchers and physicians can gain more or less instant access to a whole universe of information; the only problem is to know how to find it. In informatics and the information sciences, this problem is mirrored by the question of how to organize knowledge so that (i) it becomes easy for the knowledge seekers to find, and (ii) it becomes easy to update.

There is no reason to think that the combined information increase and computerization within the sciences will soon come to an end. Rather, it is easy to see some possible lines of development. Thus, if the knowledge possessed by experts in the various sub-disciplines of biology and medicine could be organized and stored in *interconnected* computer databases, it would become even more easily accessible. Probably also updating in the light of new scientific and medical discoveries would become easier. Perhaps the information contained in such databases could also be used as a basis for various kinds of automated reasoning that would assist in furthering the goals of scientific research and in the diagnosis and treatment of patients. Such an accomplishment, however, implies yet another and more expanded division of labor within the scientific community – one which has been developing already for some decades. Informatics, information science, and ontology have entered the scene as new sciences. Let us take a quick historical look.

When scientific knowledge was stored only in journals and books, the latter were classified by librarians. In order to access the stored knowledge, people had to learn the classification systems of the libraries. To a very large extent, the librarians' classifications mirrored the taxonomies of the world that explicitly or implicitly were used by the scientists themselves. With the growth of the number of journals and books that presented research results, this classificatory enterprise became more complicated, and in the second half of the twentieth century the discipline 'library science' saw the light. Nonetheless, one can say that during the whole of this era there was a clear-cut boundary between the taxonomies of entities in the world produced within scientific disciplines and the taxonomies of books constructed by librarians. Library classifications had no

repercussions on the taxonomic work done by the scientists themselves. Today however things are different.

We have reached a stage where in most cases scientists are so specialized that it falls on people working in informatics, the information sciences, and ontology to put new knowledge bits in the right place in the larger knowledge picture. Aristotle had people send him plants and animals from all over the world, and then he tried to create taxonomies to fit. Today, information scientists annotate what they find in scientific journals and combine it together into taxonomic hierarchies stored in databases, which are then used by other scientists throughout the world. In this way, the classificatory work carried out in the information sciences actually influences how the world itself is divided by science into parts. That is, part of the taxonomic work needed in science is now performed by a special group of people that are working with databases and the interconnections between databases.

In this work, an old term from philosophy, ‘ontology’, has taken on a partly new and wider meaning. Traditionally, when a philosopher was said to have put forward an ontology, this meant that he had put forward a taxonomy of the most general kinds of entities (often called ‘categories’) that he thought there are or could be in the world. Classical examples of such categories are substance (which subsumes everything called ‘natural kinds’ earlier in this chapter), quality (subsumes everything earlier called ‘properties’), relation, quantity, place, time, and so on. Since such philosophical ontologies were about what exists in a very general sense, they can (from an information scientific perspective) be called ‘realist ultimate top-level taxonomies’.

The term ‘taxonomy’ was originally used only in relation to systematic classifications of biological entities, first plants and animals, later microorganisms. But the term has for a long time now been used also for the classifications of many other kinds of entities, natural as well as artificial.

As the term ‘ontology’ is used in today’s informatics, information science, and computer science, it can cover:

(i) computer representations of philosophical ontologies
(ii) computer representations of scientific taxonomies and partonomies
(iii) systematic computer classifications of diverse very specific domains of artificially created things and states of affairs, for example cars and films.

If (i) and (ii) are combined in one representation we may speak of ‘ontology as science’.

The process of regimenting information for building an ontology of a certain domain (i.e., building a low-level ontology) can be a very complex affair, but here are certain steps into which it can be broken down, at least ideally. First, domain information is gathered. Second, terms and a format for these terms are selected. Third, clear, scientifically accurate and logically coherent explanations of the terms are provided. Fourth, each of these terms is given one or more places in a hierarchical classification of the domain information. Fifth, the domain information in question is formalized and implemented in computers.

For hundreds of years scientists have sought consensus as to what classification schemas, nomenclatures, and measurement units to use. Standardization conferences have been an important part of the development of science. The fact that today part of the taxonomic work of science falls upon information scientists means that even much standardization work has to be done by informaticians; and probably this will be even more so in the future. Interaction between bioinformaticians and scientists within the life sciences seems to be required, and in a way that is at least implicitly accepted on both sides. With today’s rapid growth of knowledge, such interaction can be expected to become an almost continuous affair.

Builders of taxonomies and ontologies should try to make their constructions mirror the world, but computer ontologies are built also in order to make communication and knowledge transmission efficient. Therefore, the value of an ontology depends in part on the quality of the network for shared communication that it provides and on the number of users who agree to adopt this common network. Before proposing an ontology for a given scientific domain, the custodians of this ontology have

a duty to maximize the likelihood that it will provide for the needs of a maximally large number of potential users. That is, the ultimate objective of biomedical ontologies can be seen only when the ultimate objectives of these ontologies are brought into the picture. And these objectives are to enhance health care and make treatments of individual patients and patient populations more efficient.

In relation to all the different kinds of information science ontologies we have mentioned, we would now, approaching the end of the book, like to reconnect to Chapter 3.5 and to the central view there defended, namely realist fallibilism, which we have claimed is the golden mean between positivism and anti-scientific social constructivism. In informatics and the information sciences, there seems to be no positivism around, but, on the other hand, social constructivism is presently a seemingly widespread view. Many people working in these areas seem falsely to think *either* that the whole world is simply constituted by language, *or* that even though there is a world independent of language, we can nonetheless know nothing about it; and that this epistemological predicament makes it reasonable to behave as if the whole world were constituted by language. In both cases, it is thought that classifications, partonomies, taxonomies, and ontologies have no relation to anything outside the terms and concepts used in their respective construction, i.e., that the terms and concepts in question are not regarded as being able to have language-independent referents.

Both scientific theories and ontologies are man-made artifacts, but they are *representational* artifacts, i.e., they are *about* something. Even though both theories and ontologies are *made of* terms or concepts, they are *about* what the concepts refer to. For instance, theories about cancer are not about ‘the concept of cancer’. They are about cancer. Or to take two examples from bioinformatics: the Gene Ontology is about genes, not about ‘the concept of gene’, and the Foundational Model of Anatomy is about the anatomical parts of the human body, not about ‘the concept of anatomical part’. The real issue of philosophico-ontological realism versus social constructivism is whether the objects referred to in artifacts such as those mentioned can be regarded as existing independently of language. If they cannot, then they are fictions created by our concepts, but they are even then not identical with such concepts since they are then what the concepts

refer to. In relation to ‘cancer’ the question is whether this concept refers to some mere fiction or to spatiotemporally real cellular processes. We hope that the clarifications made in Section 1 of this chapter, has shown what the realist alternative really amounts to, and in what way realism allows partly language-constituted classes within its framework, too.

Let us repeat our point in other words. If it turns out that a proposed scientific theory or a proposed philosophical ontology has used terms and concepts that lack a real referent – as is the case in our favorite examples of ‘phlogiston’ and the Galenic ‘spirits’ – these terms and concepts do nonetheless not refer to themselves. They are not looked at and talked about in the way they can be talked about in linguistics (‘phlogiston’ is a noun) and in works on the history of chemistry (‘phlogiston’ is a concept once wrongly used in chemistry to explain combustion). *Retrospectively*, these terms and concepts have to be described as referring to fictions in something like the way in which words in novels can describe and refer to fictions. But, as explained in Chapter 3.5, this is no reason to think that all presumed referents are purely fictional referents. In that chapter we discussed and dismissed the challenge that the facts of the history of science create for realism, but in the information sciences there is also another fact that seems to create a tendency towards social constructivism: taxonomic *class hierarchies* often run parallel with corresponding *term hierarchies*.

Take the three terms ‘color’, ‘red’, and ‘dark red’. We can immediately see that the class of referents to ‘dark red’ is a subclass to the class of referents to ‘red’, which, in turn, is a subclass to the class of referents to ‘color’. At the same time it seems as if the meaning of the term ‘dark red’ already contains the meaning of ‘red’, which, in turn, already contains the meaning of the term ‘color’. Here, there is a clear parallel between ‘class subsumption’ and ‘meaning inclusion’; when the class of A (red things) is subsumed by the class B (colored things), then conversely the meaning of the term ‘A’ (‘red’) includes the meaning of the term ‘B’ (‘colored’). So why, in taxonomic work, should we bother about realism and the referents of terms? Isn’t the meaning of the terms all that taxonomists need? Can they not just (in the terminology of Chapter 3.5) *look at* the terms and forget about the world? To see why this is not the case, let us once again bring in our lion example; let us see how it looks when the classes and

terms are made to run parallel in the way we had class subsumption and meaning inclusion run in parallel in the color example above. We then obtain Table 4.

Classes (the lower classes are *sub-classes* of the higher classes):	**Terms (the meaning of the upper terms are *meaning-parts* in the lower):**
Animalia	'Animals'
Chordata	'Chordata' (having spinal chord)
Mammalia	'Mammalia' (feeding new-born with milk)
Carnivora	'Carnivora' (meat eating)
Felidae	'Felidae' (long tails, moves like a cat)
Pantera	'Pantera' (being big)
Pantera-leo (lion)	'Pantera-leo' ('lion')

Table 4: *Illustration of how classes and corresponding terms may run in parallel.*

On the left hand side of Table 4, we have the subsumption hierarchy of classes; on the right hand there is the corresponding hierarchy of terms and their meanings (concepts). We have made explicit the more common meaning of the Latin terms. This makes clear the fact that the larger the class of entities that falls under a given term is, the less meaning the term has, and vice versa. There is an inverse relationship between class membership and semantic content of the term that names the class. The problem at hand is whether the class hierarchy can be obtained only by means of the term hierarchy or not. Our answer is both 'yes' and 'no', and the importance of both answers have to be seen.

If a person who knows very little about biology were to read through (over and again) the modern literature on lions, but no book that made explicit any hierarchy like that in Table 4, then he would probably nonetheless be able to construe it. In this sense, the hierarchy can be obtained only by means of the terms in the right column. This knowledge of lions, however, has not been with mankind since lions were first discovered and named. It has required much study of real lions in the world to create the corresponding body of knowledge that we today have; and in

this sense the hierarchy can by no means be attained only through the organization of terms in a language.

Let us deepen this consideration with a fictional story. We hope it will function as a thought experiment that further stresses the fact that the information sciences cannot always rest content with studying terms and concepts, but need to interact with empirical science.

Once upon a time there was a knowledge engineer called KE. He was given three tasks: to construct a taxonomy of (a) all the presently existing mammals on earth, (b) all dinosaurs, and (c) all humanoids in the movie Star Wars. How did he proceed?

In the case of the earthly mammals, KE started to read relevant zoological literature as well as to interview taxonomically interested zoologists, i.e., the domain experts at hand. Since KE was a very good knowledge engineer, he quickly understood the contemporary scientific mammal classification, and turned it into a good 'knowledge model' in his computer. In the second case, he did the same with dinosaur literature, learned about the two main orders of Saurischia and Ornithischia, and all the subclasses; and, again, he produced a good computer model. In the third case, KE started immediately to watch the Star Wars movies, to read about them, and to check on the internet what the domain experts now at hand, the real Star Wars fans, had to say. Immediately, he found on the internet a list of the humanoid types that occur in Star Wars. The Abyssian species was said to be 'a swarthy-faced humanoid with a single slit-pupilled eye, a thirsty representative of the cyclopean species'; the Aqualish species was 'a bellicose humanoid species with some superficial properties both aquatic and arachnid; they have hair-lined faces with large, glassy eyes and mouths dominated with inwardly-pointing tusks'; and so on. After some work, as expected, KE had classified the humanoids into genera, families, and orders; and put also this taxonomy on the internet. So far, so good, and there is no essential difference between the three cases.

After a while, however, KE discovered that he had worked a bit too fast. In each of the three taxonomies there was one group of animals that did not really fit into the respective taxonomy at hand. He went carefully through all his material once again, but these three problems remained. What to do? His reputation as a good knowledge engineer seemed to be at stake. He acted as follows.

He first tried to solve the problem in the mammal taxonomy. In this case KE asked a zoologist to make some completely new observations related to his problem. Since the mammals to be classified were really existing species, this was merely a practical problem. As things turned out, the new observations were made and they completely solved the taxonomic problem at hand. KE was quite happy.

In the dinosaur case, he behaved in the same way. Couldn't some researchers try to find some new dinosaur traces that could solve the problem? But the answer was disappointing. No one had any idea at all where to find new dinosaur traces. The situation was epistemologically closed. But his domain experts promised to contact KE if something relevant should turn up in the future. To his annoyance, KE had to leave one group of dinosaurs outside his taxonomy – at least for the foreseeable future.

But how should he behave with respect to the problematic humanoids? Since they do not exist, no new observations could be made; and since they have not even existed in the past, there was no hope, either, of finding new traces. Furthermore, no more movies would be made. KE's problem came to an end when one day he realized that fictional entities necessarily have what might be called 'spots of indeterminacy' (the term is taken from the Polish philosopher Roman Ingarden, 1893-1970). What is not described by the creators of a fictional entity is indeterminate in the strongest possible sense: it does simply not exist. Therefore, there was nothing more that KE could look for or do in relation to the problematic humanoids. The situation was *ontologically* closed. Nothing could and nothing should be done by a good knowledge engineer. Without any annoyance at all, KE left one group of humanoids outside his taxonomy – for ever.

When a knowledge engineer meets domain experts that specialize on some areas or aspects of the real world, then he has, whether he is aware of it or not, a choice. *Either* he behaves as if what the experts tell him is all there is to say; that is, he treats them as authors of a story about fictional objects that may contain indeterminacies, perhaps even inconsistencies. *Or* he takes a realist perspective and keeps an eye open to the possibility of interaction with the world itself; if the domain experts contradict each other, then he asks them to reconsider their views, and if they lack knowledge, he asks them to try to acquire it.

The default position for every ontology creator concerned with science should be that of realist fallibilism. Terms with substantive content are primarily links to the world; be it directly as in nominal or real-prototypical terms, or indirectly as in ideal-prototypical terms. With new knowledge the ontology has to be upgraded.

*****

In language we often abbreviate without any misunderstandings occurring. ‘Ingrid Bergman is Ilsa Lund’ is short for ‘Ingrid Bergman *is playing* Ilsa Lund’, and everyone who knows the movie *Casablanca* understands what this means. In our opinion, analogously, the statement ‘the colors we perceive are electromagnetic waves’ is short for ‘the colors we perceive are *caused by* electromagnetic waves’; but here the so-called ‘reductive materialists’ and ‘identity theorists’ (described in Chapter 6.1) falsely take the shorter statement to be literally true. Many information scientists agree with an unclear characterization of ‘ontology’ that has become (January 2008) the view of Wikipedia’s entry ‘Ontology (computer science)’. It says: “In both computer science and information science, an ontology is a data model that represents a set of concepts within a domain and the relationships between these concepts.” From the perspective of this book, such a statement is true only if it is understood as an abbreviation of the following sentence:

- an ontology is a data model that represents a taxonomy (or partonomy) of classes, which, however, runs parallel with a set of concepts within a domain and the relationships between these concepts.

Earlier in this chapter, we claimed that the realist fallibilism we favor does not halt in front of taxonomy, informatics, and the information sciences. We end by claiming that the Wikipedia-sentence quoted must by no means be turned into the view that the information sciences are concerned only with terms and concepts.

Reference list

Bunge M. *Scientific Research*, 2 vol. Springer-Verlag. New York 1967.

Donnelly M, Bittner T, Rosse C. A Formal Theory for Spatial Representation and Reasoning in *Biomedical Ontologies. Artificial Intelligence in Medicine* 2006; 36: 1-27.

Dupré J. In Defense of Classification. *Studies in the History and Philosophy of Biology and the Biomedical Sciences* 2001; 32: 203-219.

Ghiselin MT. *Metaphysics and the Origin of Species*. State University of New York Press. Albany 1997.

Grenon P, Smith B. Persistence and Ontological Pluralism. In Kanzian C. (ed.). *Persistence*. ontos verlag. Frankfurt 2007.

Hand DJ. *Measurement Theory and Practice. The World Through Quantification.* Arnold. London 2004.

Jansen L, Smith B (eds.). *Biomedizinische Ontologie. Wissen strukturieren für den Informatik-Einsatz*. vdf Hochschulverlag. Zürich 2008.

Johansson I. Hartmann's Nonreductive Materialism, Superimposition, and Supervenience. *Axiomathes. An International Journal in Ontology and Cognitive Systems* 2001; 12: 195-215.

Johansson I. Determinables as Universals. *The Monist* 2000; 83: 101-121.

Johansson I. Functions, Function Concepts, and Scales. *The Monist* 2004; 87: 96-114.

Johansson I. Roman Ingarden and the Problem of Universals. In Lapointe S, Wolenski J, et al (eds.), *The Golden Age of Polish Philosophy. Kazimierz Twardowski's philosophical legacy*. Springer. Berlin (forthcoming).

Mahner M, Bunge M. *Foundations of Biophilosophy*. Springer. Berlin 1997.

Mayr E. *Principles of Systematic Zoology*. McGraw-Hill. New York 1969.

Mayr E. *Toward a New Philosophy of Biology*. Harvard University Press. Cambridge Mass. 1988.

Munn K, Smith B (eds.). *Applied Ontology: An Introduction*. ontos verlag. Frankfurt 2008.

Neuhaus F, Smith B. Modeling Principles and Methodologies – Relations in Anatomical Ontologies. In Burger A. et al. (eds.). *Anatomy Ontologies for Bioinformatics: Principles and Practice*. Springer. Berlin 2007.

Robert JS. *Embryology, Epigenesis and Evolution: Taking Development Seriously*. Cambridge University Press. Cambridge 2006.

Rosch E. 1983. Prototype Classification and Logical Classification: The Two Systems. In Scholnick, E., *New Trends in Cognitive Representation: Challenges to Piaget's Theory*. Hillsdale, NJ: Lawrence Erlbaum Associates: 73-86

Rosse C, Mejino JLF. The Foundational Model of Anatomy Ontology. In Burger A. et al. (eds.). *Anatomy Ontologies for Bioinformatics: Principles and Practice*. Springer. Berlin 2007.

Schulz S, Johansson I. Continua in Biological Systems. *The Monist* 2007; 90 (no. 4).

Smith B, Varzi AC. Fiat and Bona Fide Boundaries. *Philosophy and Phenomenological Research* 2000; 60(2): 401–420.

Smith B. Ontology. In Floridi L (ed.). *Blackwell Guide to the Philosophy of Computing and Information*. Blackwell. Oxford 2003.

Smith B. The Logic of Biological Classification and the Foundations of Biomedical Ontology. In Westerståhl D. (ed.). *Invited Papers from the 10th International Conference in Logic Methodology and Philosophy of Science, Oviedo, Spain, 2003*. Elsevier-North-Holland 2004.

Smith B. Beyond Concepts: Ontology as Reality Representation. In Varzi A, Vieu L. (eds.). *Proceedings of FOIS 2004. International Conference on Formal Ontology and Information Systems*, Turin, 4-6 November 2004.

Wiley EO. Phylogenetics. *The Theory and Practice of Phylogenetic Systematics*. John Wiley & Sons. New York 1981.

Wilson J. *Biological Individuality*. Cambridge University Press. Cambridge 1999.

Index of Names

(philosophers, physicians, scientists, and science journalists)

Index of Subjects

Picture Acknowledgements

For allowances to use pictures, we would like to thank:
Wellcome Trust Medical Photographic Library (London),
Chapter 2: Figures 1, 2, 3, 5, 6, 7, 8, 9.
Chapter 3: Figures 1, 2, 5.
Chapter 4: Figures 3, 4.
Chapter 6: Figure 5 (right).
The Nobel Committee for Physiology or Medicine (Stockholm),
Chapter 7: Figure 3.
National Library of Medicine (Maryland),
Chapter 6: Figure 5 (left).
John W. Kimball's Biology Pages (http://biology-pages.info),
Chapter 4: Figure 2.
Urs Broby Johansen,
Chapter 6: Figure 4 (drawing by Ebbe Sunesen).
Jon-Kar Zubieta,
Chapter 7: Figure 2.
Thomas Boren (Jani O'Rourke, Adrian Lee, and 'Scientific American Science & Medicine'), Helicobacter pylori on the front cover.
(Possible copyrights to some of the pictures have been impossible to track down.)