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).

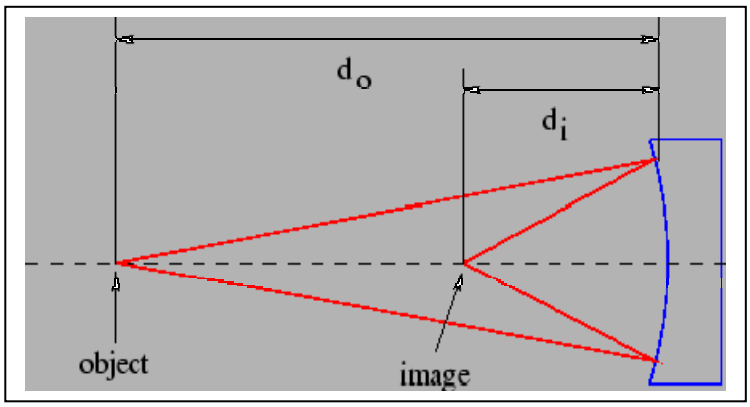

Figure 2: *The mirror formula exemplified.*

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.

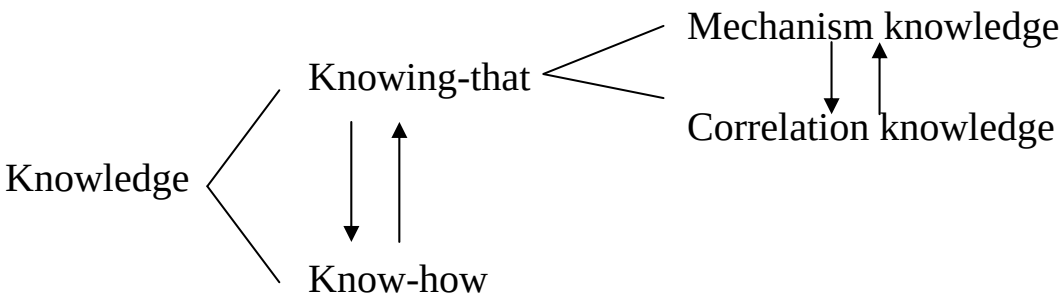

Figure 3: *Interactions between different forms of knowledge.*

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).

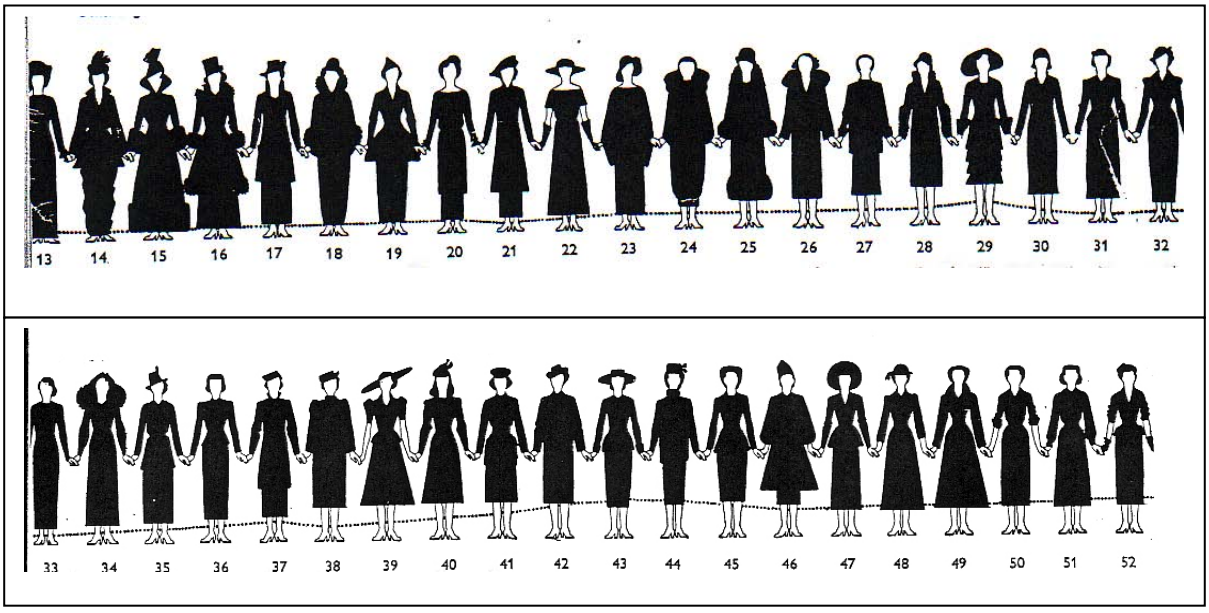

Figure 4: *Picture from Broby-Johansen R. Body and Clothes.*

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.

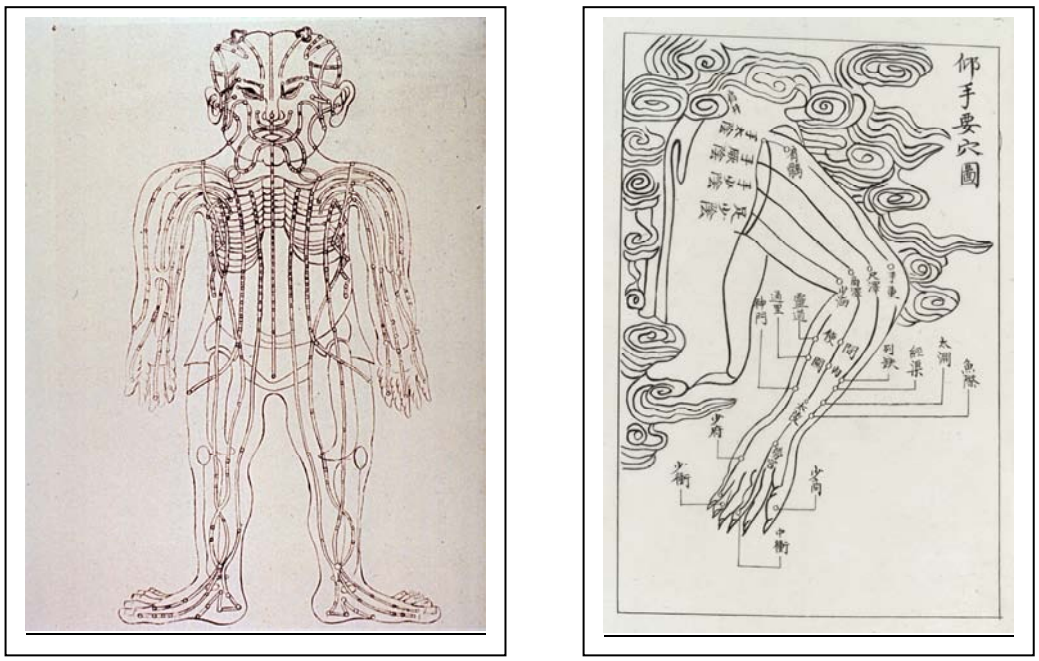

Figure 5: *Acupuncture meridians in old Chinese drawings.*

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.

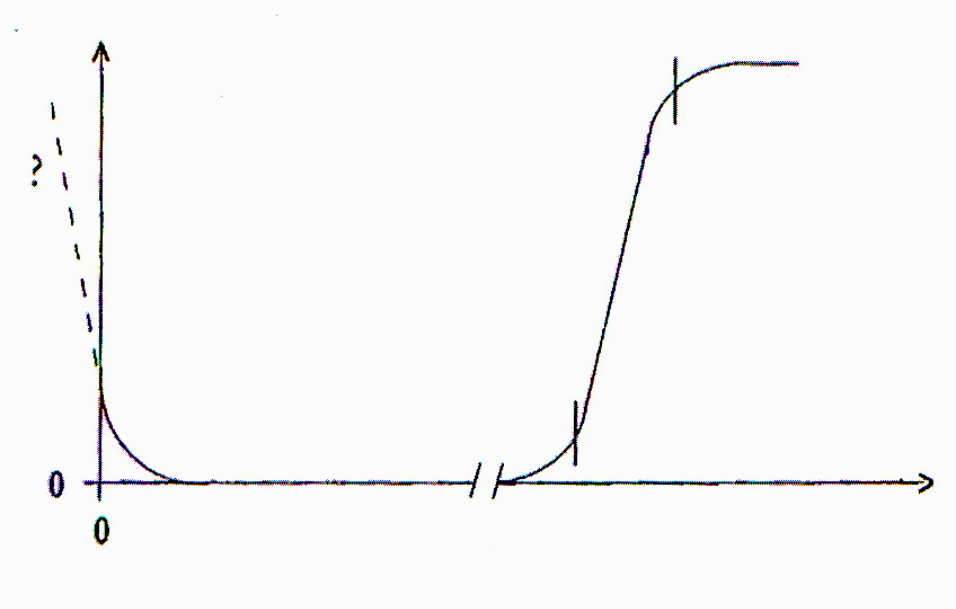

Figure 6: *To the right is a normal dose-response curve; doses on the x-axis and responses on the y-axis. Within a certain range there is a rather linear relationship where a higher dose gives a higher response. To the left is an attempt of ours to illustrate homeopathic theory. Only small doses give 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.