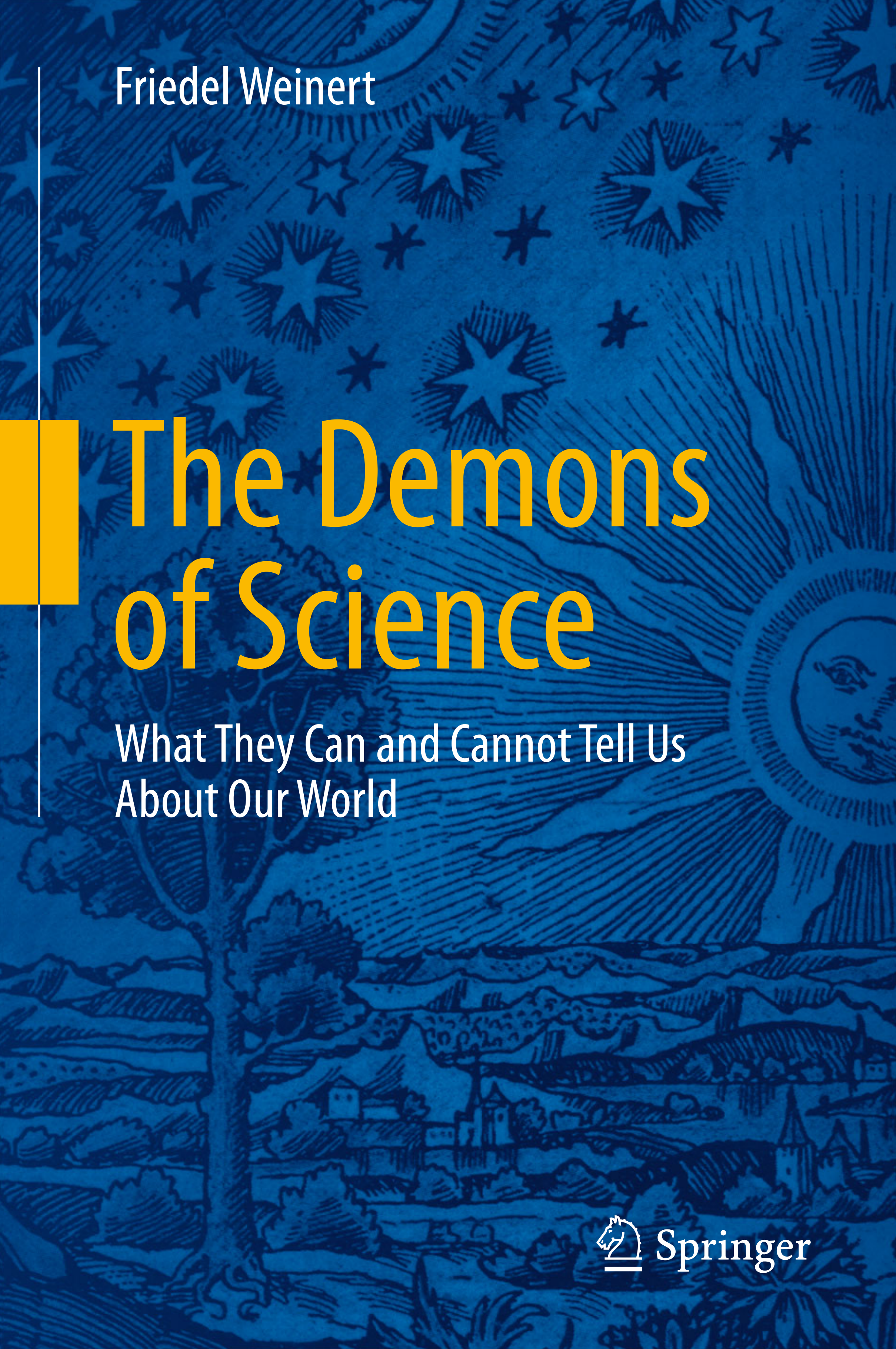


Friedel Weinert

The Demons of Science

What They Can and Cannot Tell Us About Our World

Springer

The Demons of Science

Friedel Weinert

The Demons of Science

What They Can and Cannot Tell Us About Our World

Springer

Friedel Weinert
Faculty of Social Sciences
University of Bradford
Bradford
UK

ISBN 978-3-319-31707-6 ISBN 978-3-319-31708-3 (eBook)
DOI 10.1007/978-3-319-31708-3

Library of Congress Control Number: 2016936279

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

The title *The Demons of Science* may at first appear like a contradiction in terms. Demons are associated with the forces of darkness; science represents the power of light. One could assume, therefore, that science has no time for demons. This book aims to destroy this assumption. Science opens its gates to demons as long as they play a rational rather than an evil part. They are put to work. Demons are figures of thought: they belong to the category of thought experiments, which are routinely employed in science and philosophy. As they are cast as agents with superhuman abilities, we may expect that demons provide us with valuable—albeit non-empirical—clues about the constitution of the physical world. But I am interested in exploring not only what the demons tell us but also what they do not tell us about our world. They are cast as superhuman actors but even demons have their limitations. The following chapters contain, I believe, the first systematic study of the role of demons in scientific and philosophical reasoning about the external world.

I have to thank a number of people for helping me along the way: Roger Fellows (Senior Research Fellow at the University of Bradford), Roman Frigg (Professor of Philosophy at the London School of Economics) and Robert Nola (Professor of Philosophy at the University of Auckland) who either read all or part of the manuscript and have given me valuable advice. An invitation to give a talk on the cosmological arrow of time at the Sigma Club of the Department of Philosophy at the London School of Economics (January 2016) has helped me clarify some uncertainties about the powers of Loschmidt's Demon. I thank the members of the audience for a stimulating discussion. I was granted sabbatical leave in the summer of 2015 and I would like to thank the Faculty of Social Sciences at the University of Bradford for granting me the time to finalise the manuscript. I spent the 3 months of the sabbatical at the *Center for Mathematical Philosophy* at the University of Munich. I would like to thank its Director, Stephan Hartmann, for the invitation, the stimulating atmosphere and the warm welcome. I take this opportunity to thank Angela Lahee, not only for her enthusiasm for the Demons of Science, but also for her unfailing support over the years.

I can confirm that no demons had a hand in writing this book. But I hope that the reader will enjoy reading it as much as I enjoyed writing it.

Friedel Weinert

Contents

Part I Thought Experiments

Part II Laplace's Demon

Part III Maxwell's Demon

Part IV Nietzsche's Demon

Part V Conclusion

Chapter 1
Introduction

(...) physicists have become demons.

Lewis (1930: 40)

This is a book about demons, not the scary demons of fiction but the reasonable demons of science. Demons are figures of thought who are used as argument patterns in philosophical and scientific reasoning. In his wonderful book *From Here to Eternity* the cosmologist, Sean Carroll, exclaims at one point: 'What is it with all the demons, anyway?' (Carroll 2010: 400, Footnote 167). The occasion for this outburst is his discussion of Nietzsche's thesis of an eternal recurrence of all events. Nietzsche employs a demon to convey his message of the 'wheel of the cosmic process.' As Carroll rightly implies, demons are frequently employed as thought experiments in the history of philosophical and scientific reasoning about the world.

The present book aims to answer Sean Carroll's rhetorical question. Scientists—and philosophers alike—seem to be fond of **demons**: references to metaphorical demons abound in their thought experiments. Descartes, Laplace, Maxwell, Loschmidt, Landsberg, Nietzsche and Freud conjured up their own demons. Even the genetic work of the humble Augustinian friar Gregor Mendel has been associated with a demon.

So what about demons? Demons are supernatural beings. In the scientist's reasoning repertoire they fulfil an important function. Their job is to explore the coherence, limits and the potential of human knowledge about the natural world. They may also propose bold new hypotheses and challenge existing knowledge claims. But scientific knowledge also has philosophical consequences. Often wide-ranging philosophical claims are made in the name of the demons of science.

The French astronomer Pierre Laplace used his eponymous Demon to claim that the world is completely deterministic, like a clockwork universe. If the universe is a deterministic chain of events it seems to follow that the passage of time and our cherished free will are mere illusions. Where does this leave our human impression of the flow of time and the exercise of free will?

Maxwell's Demon cast a shadow of doubt over the 19th century view that the universe was inexorably on a trajectory, from order to disorder, towards an unavoidable 'heat death', providing us with a cosmic arrow of time. According to

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_1

Maxwell's Demon the transition from order to disorder—the increase in entropy—is only probabilistic, not deterministic. Where does this leave the cosmic arrow of time?

Finally, Nietzsche's Demon claims that the events in the universe repeat themselves over and over again. The Demon announces the eternal return of events. But do we actually live in such a cyclic universe? Not according to Landsberg's Demon who casts his eyes not just on the history of our universe but on the multiverse.

These are momentous claims and one of the **aims** of this investigation is to evaluate their validity. The investigation hopes to draw the 'true' boundaries of what, in the name of demons, *science tells us and does not tell us about our world.* It is undeniable that science plays a major role in the explanation, control and understanding of the natural world and the universe. But the overall *thesis* of our investigation is that the demons of scientific thinking do not show that humans have no free will, that the flow of time is a human illusion, that the universe is like a massive stack of cards, on which all events—past, present and future—are already inscribed as if frozen in a timeless universe. Such claims are *philosophical* consequences, which do not follow deductively from the scientific theories. There is disagreement amongst the demons. Maxwell's Demon opposes Laplace's Demon. Landsberg's Demon contradicts Nietzsche's Demon. Others demand their say. Demons have limitations, which make them less powerful than they appear to be. The *first* aim of our investigation is therefore to establish what philosophical consequences can really be drawn from an investigation of the role of demons in scientific thinking. In the process this project pursues a *second* aim: to investigate the shared conceptual structure of science and philosophy, to explore their common conceptual toolbox. It proposes to probe the numerous connections between thought experiments in science and wider philosophical notions, which often underlie the description of nature. There are of course many thought experiments, which work without the services of demons. But it is equally true, as we shall see, that thought experiments would often benefit from the helping hand of a demon.

The employment of demons invariably has wider philosophical implications since these thought experiments—these demons—involve notions, which form a shared conceptual platform where scientific and philosophical thought meet. As we will discuss, Laplace's and Loschmidt's Demon address issues like determinism and causality, free will and fatalism, reversibility and predictability; Maxwell's Demon is concerned with indeterminism and irreversibility, probability and the Second law of thermodynamics; Nietzsche's and Landsberg's Demons are pre-occupied with cosmic evolution, our universe and the multiverse as well as the cosmic arrow of time. The demons pull together the strings of some of the important notions and their consequences, which underpin the work of science and philosophy in an endeavour to understand the surrounding cosmos. To investigate the demons means to investigate these notions and the philosophical consequences of scientific thinking.

The study's focus on the demons of science leads to a natural **coherence** of the topics to be discussed. It consists of four parts, each with individual chapters. The chapters spell out the conceptual ramifications of the overall themes in each part.

The first task, in **Part I**, will be to evaluate the role of **thought experiments** in science and philosophy. Although thought experiments only happen in the workshop of the mind, rather than in real laboratories, they have played a decisive role in the history of rational thinking, from the Greeks to the present day. What is their function? A number of philosophical accounts of thought experiments have been proposed in the literature, but after a consideration of their strengths and weaknesses this part will settle on the view that they are a particular type of model—they are conceptual models. Hence demons, too, are conceptual models. As models generally are of great importance in science, thought experiments fit into a typology of models, which will be proposed. Like all models, thought experiments make use of abstractions, idealizations and the interrelations between the modelled parameters. They employ counterfactual and hypothetical reasoning and test the non-empirical values of scientific theories. They do not enrich the store of empirical knowledge but they contribute to our *understanding* of the world around us. As conceptual models, demons are particularly well equipped to address counterfactual questions: What if a demon could travel to the edge of space or through the interior of the Earth? What if a demon could manipulate molecules at will? What if a demon could survey the whole universe or even the multiverse? The subsequent parts and chapters will focus on demons who stand at the crossroad of physical science and philosophy—such as Laplace's, Maxwell's, Loschmidt's, Nietzsche's and Landsberg's Demons. But many more demons populate the pages of scientific and philosophical volumes. Part I will conclude with a brief consideration of Freud's, Descartes's, and Mendel's Demons.

Part II is devoted to **Laplace's Demon**. Laplace's Demon is a denizen of a deterministic world, of the clockwork universe. He is a determinist, not a fatalist. He sees the whole universe as an interlocking chain of events, stretched out from past to future. Laplace's Demon can be interpreted as a representative of different versions of determinism (causal, metaphysical or scientific). His determinism naturally points to a discussion of the nature of fundamental laws. The fundamental laws of physics make no distinction between past and future. They are t-invariant. It would appear, then, that Laplace's Demon recognizes no arrow of time because to his superhuman gaze all events—past and future—have already occurred. But if every event has a prior cause, the Demon is led to deny the existence of free will. As will emerge in this part, Laplace's determinism has its limits, even in the classical realm in which his Demon operates. The Demon's mistakes tell us not to confuse determinism and causality and that his determinism can be made compatible with the arrow of time. But if the classical world is not as rigid as Laplacean determinism would suggest, it is unlikely that classical theories imply fatalism, i.e. the belief that the die of existence has already been cast and cannot be changed. If determinism is limited in its grip over the world, there seems to be room for chance and free will. Some arguments in favour of free will be reviewed. The concluding chapter will explain that Laplace's Demon does not tell us that the world is completely deterministic or even fatalistic; that there is no passage of time or free will. The discussion of chance and indeterminism gives rise to a consideration of statistical notions, which leads naturally to Maxwell's Demon.

Part III will therefore focus on **Maxwell's Demon**. Maxwell's Demon was originally concerned with the refutation of a particular reading of the Second law of thermodynamics, which roughly is a statement of the universal transition from order to disorder in the natural world. Some leading scientists of the day used this increase in disorder—which is the 19th century understanding of the notion of entropy—to *identify* entropy with the arrows of time. It turned out to be a mistake; it is better to use entropy as an *indicator* of the arrows of time. It is also necessary to introduce a distinction between *local* and *cosmic* arrows of time: humans would experience a lapse of time even in a universe, which curls back on itself. On a local level, time would go forward but this limited experience does not reveal whether the universe itself displays an arrow of time. The focus in this part will be on local arrows of time. How are they recognized? The chapters (in this part) introduce two further readings of the notion of entropy: one in terms of information loss and the other in terms of phase-space volumes. Especially the latter reading gives rise to the question whether the trajectories of physical systems are reversible or irreversible. In order to answer this question the services of a new demon: Loschmidt's Demon are required. In the textbooks of physics, Loschmidt's Demon is usually tasked with making trajectories of mechanical systems reversible. But as it turns out even Loschmidt's Demon cannot reverse the trajectories to achieve a reversal of time. If trajectories of systems are often irreversible, in practice if not in theory, the world is to a certain degree indeterministic, that is, the present leaves open alternative future histories. Hence Maxwell's Demon disagrees with Laplace's Demon. If they disagree, indeterminism requires a re-examination of the notions of causality and the role of the mind in the material world. Such a re-examination has several consequences. One consequence is the introduction of a 'conditional' notion of causality, a probabilistic *ersatz* for deterministic causation. Although Maxwell's Demon demotes the Second law of thermodynamics from the place of pride it once held in classical physics, the Demon allows the notion of entropy to be used as a criterion, amongst others, for the discussion of the direction of causality, the past-future distinction and the local arrows of time. Another consequence of this re-examination is a reconsideration of the Darwinian research programme to locate the mind in the material world. According to the Darwinian programme, the brain is an indeterministic system and the mind 'emerges' from the brain. Two modern 'solutions' to the problem of the mind are discussed, one in terms of physics and the other in terms of evolutionary biology. Unfortunately, both fail in their attempt to complete the Darwinian research programme. Maxwell's Demon introduces the world to statistical notions. Ludwig Boltzmann—the Austrian physicist who made major contributions to our understanding of the notion of entropy—dubbed the 19th century the 'statistical age', adding that it could also be known as Darwin's century. Darwin's theory of evolution is in fact a statistical theory. It reveals an interesting connection between Maxwell's Demon, entropy, the evolution of life and the universe. The Maxwellian Demon points to an entropic arrow of time.

Part IV is mainly concerned with the cosmic arrow of time and how it relates to other temporal arrows. It starts with a discussion of **Nietzsche's Demon**. Nietzsche's claim of the eternal return of events stands in a long tradition of scenarios of cyclic universes. But cyclic universe models are philosophically incoherent. In order to consider a cosmic arrow of time **Landsberg's Demon** is a better guide. For Landsberg's Demon realizes that the universe is no longer Newtonian in character and that it is necessary to move beyond Laplace's Demon. Laplace's Demon only focussed on our universe—the Milky Way—but Landsberg's Demon is a denizen of the multiverse. He perceives the panorama of the whole multiverse, and how it gives birth to individual universes. The multiverse can be conceived in a number of ways. It may be represented as an eternally existing cosmic landscape, a succession of oscillating universes or as the cosmic mother of 'baby universes'. Each case throws up the question of the cosmic arrow of time, both for the multiverse and its galactic offspring. Contrary to received views, space-time models of the universe are compatible with a physical arrow of time since they are 'time-orientable'. A physical arrow of time is derived from the fundamental connection between time and dynamic change. Given the existence of physical time, the question arises how physical time is related to phenomenal time, i.e. our subjective impression of the flow of time and a universal Now. Does mental time presuppose physical time? If physical time exists, it manifests itself both on a local and a cosmic level. There are in fact many arrows of time and the question imposes itself whether there is a master arrow of time. On the strength of an evolutionary view, this part will argue against the existence of a master arrow of time—like entropy—, from which all other arrows could be derived. The various arrows of time are explained in analogy with Darwin's evolutionary tree image. Just as various species evolved along the evolutionary tree, so various arrows of time have emerged since the Big Bang.

Nietzsche's Demon does not show that humans are locked in a nightmare scenario of an everlasting return of events, which they are forced to re-live. The universe is not cyclic in nature. Landsberg's Demon informs us that dynamic changes, from the smallest to the largest scale, provide criteria for inferences to the many arrows of time. Although there is no master arrow this part will conclude that time and its arrows are multi-fingered.

Science has not killed the demons. They serve their purpose as thought experiments in order to explore, test and investigate our knowledge claims about the world. If the demons teach us what they can and cannot tell us about the world, they will have done their job!

Part I
Thought Experiments

In a thought experiment, one strives to uncover general principles from the mere mental consideration of experiments that one might *perform.*

Penrose, *The Emperor's New Mind* (1986: 466)

A thought experiment is generally a conceptual model, in which an unrealized or unrealizable situation is depicted, whose conceptual or logical consequences are then investigated in the laboratory of the mind. The purpose of a scientific thought experiment is to probe the consistency and rationality of accepted scientific arguments, to test the limits of scientific theories, to formulate new questions and hypotheses about the natural world and to simulate natural phenomena. Thought experiments may lead to a change or even abandonment of accepted theories. This part will provide a general discussion of thought experiments in science and introduce demons as a special case. The subsequent parts will shift the focus to the role of demons in thought experiments and will discuss, amongst others, Laplace's Demon, Maxwell's Demon and Nietzsche's Demon. Demons command superhuman powers and are well-equipped to expose and test the limits of our knowledge about the natural world. But demons also shine a light on the many conceptual links between science and philosophy, and the philosophical claims that are made in their names.

The *notion* of thought experiments has a long history, which harks back to the 18th century and various writers have used different terms—imaginary experiment, *Gedankenexperiment*, thought experiment—to describe this armchair activity. The German philosopher Immanuel Kant spoke of experiments of pure reason and so did the physicist and philosopher G.C. Lichtenberg (see Part I, Sect. 3.5). The Danish physicist and philosopher Hans Christian Ørsted became the first thinker to explicitly write about thought experimentation (1811); he was also the first to use the term explicitly (1812). But Ørsted's efforts remained largely unknown. The practice of thought experimentation entered academic discourse only with the work of the Austrian physicist and philosopher Ernst Mach. (Kühne 2005: 21-2) But the *use* of thought experimentation goes back to the Greeks and flourished after the Scientific Revolution of the seventeenth century.

Chapter 2
Thought Experiments in Ancient Greece

> *It is not so much the particular form that scientific theories have now taken – the conclusions which we believe we have proved – as the movement of thought behind them that concerns the philosopher.*
>
> Eddington, *The Nature of the Physical World* (1932: 353)

Image an ancient Greek who is exercised by questions of cosmic import: Is the universe finite or infinite? Is the Earth spherical or flat? Is the Earth the centre of the universe or does it rotate around a different hub, say the sun?

To some of these questions the answers are known today, thanks to the theoretical and observational work of our predecessors. But even in the absence of observational evidence the ancient Greeks, driven as we are today by theoretical curiosity, sought solutions. How do you satisfy this theoretical curiosity when observation fails as a guide and theory is uncertain? One possibility is to investigate the logical and conceptual consequences of an adopted view with the aim of establishing whether it provides an answer. If one proposition claims that the universe is finite, another that the Earth is flat, and yet another that the Earth moves, in each case an investigation must be launched in order to ascertain the consequences, which follow from each hypothesis. In the absence of real experimentation or actual observation, an investigation of conceptual and logical consequences amounts to experimentation in thought. Just as in real experiments, thought experiments introduce a number of parameters, which depict the imaginary scenario in a mental laboratory, in order to investigate their consequences. This is precisely the procedure, which some of the ancient Greeks adopted.

To illustrate, consider the conundrum of whether the universe is finite or infinite, a question to which even today no definitive answer is known. The Greek mathematician and philosopher Archytas of Tarentum introduced a thought experiment, with the help of which he hoped to obtain an answer to the question (see Huggett 2010: 33–34; LePoidevin 2003: Chap. 6; Genz 2005: 205–206). As Archytas's life coincided with the lifetimes of Plato and Aristotle, he must have been aware of the Greek geocentric worldview. The geocentric worldview was the dominant paradigm until it was displaced by the heliocentric worldview of Nicolaus Copernicus in

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_2

1543 (Weinert 2009: Part I). According to the geocentric worldview the Earth sits motionless—bereft of both a daily and an annual rotation—at the centre of a closed universe. In the Aristotelian version of this model concentric shells carry the planets in perfect circles around the central Earth. The sun itself is regarded as a planet, which occupies the sphere which, in the later heliocentric worldview, will be occupied by the Earth. The geocentric model harbours a closed universe, because the 'fixed' stars mark its boundary, beyond which resides a Deity, described by Aristotle as the 'Unmoved' Mover. The Unmoved Mover remains outside the bounded sphere, which constitutes the universe. But this Deity is ultimately responsible for all the motions below the outer sphere because it provides the energy, which keeps the spheres spinning around the centre. The Greek geocentric worldview therefore assumed a finite cosmos because the universe of planets and spheres reaches its limit at the boundary of the 'fixed' stars.

Humans cannot physically travel to the 'edge' of space but the flight of fantasy is less fettered. Archytas's imagination saw a space traveller flying to the boundary of the cosmic sphere: he might as well have imagined a demon. He asked whether the space traveller could penetrate the outer layer.

> If I am at the extremity of the heaven of the fixed stars, can I stretch outwards my hand or staff? It is absurd to suppose that I could not; and if I can, what is outside must be either body or space. We may then in the same way get to the outside of that again, and so on; and if there is always a new place to which the staff may be held out, this clearly involves extension without limit. (Quoted in Grant 1981: 106; see Fig. 2.1)

Archytas concludes that the universe has no edge and must therefore be infinite. How reliable is this conclusion, given that it was reached without access to empirical data? Can thought experiments teach us something about the external world?

Fig. 2.1 Archytas's traveller reaches the end of the universe and extends his spear through the canopy of the fixed stars. *Source*: Wikimedia Commons

A preliminary answer to these questions emerges from a consideration of two thought experiments, both due to Aristotle, which address two further issues regarding the shape of the world.

As mentioned before, the Greeks also faced the question of whether the Earth was spherical or flat. There is no doubt that throughout the ages a number of scholars were led to the conclusion that the Earth is flat (see Hannam 2009: 35–38). But the great authorities of the ancient geocentric worldview—cosmologists like Aristotle and astronomers like Claudius Ptolemy were convinced that the Earth was spherical. There was, first, empirical evidence for the sphericity of the Earth. As Aristotle says, the 'evidence of the senses' corroborates the assumption of the spherical shape of the Earth. He refers to the eclipses of the moon, which show a 'curved outline' of the Earth on the surface of our satellite,

> (…) and, since it is the interposition of the earth that makes the eclipses, the form of this line will be caused by the form of the earth's surface, which is therefore spherical. (Aristotle 1952b: Book II, Chapter 14, 297^{a})

The Greeks were also aware that the view of the night sky changes, as an observer on Earth moves from north to south.

> There is much change (…) in the stars overhead, and the stars seen are different, as one moves northward or southward. Indeed there are some stars seen in Egypt and in the neighbourhood of Cyprus which are not seen in the northerly regions and stars, which in the north are never beyond the range of observation, in those regions rise and set. All of which goes to show not only that the earth is circular in shape, but also that it is a sphere of no great size: for otherwise the effect of so slight a change of place would not be so quickly apparent. (Aristotle 1952b: Book II, Chapter 14, 298^{a})

Centuries later Ptolemy would point out that an observer, moving in an eastern direction from Greece, would notice that the sun rises earlier in eastern than in western parts of the globe. If the Earth were a flat disc, all observers would experience a simultaneous rising of the sun in the east and a simultaneous setting in the west. As this is not the case the Earth must be a sphere or at least, it cannot be a disc.

It is interesting to note that Aristotle is not content with the observational evidence of the spherical shape of the Earth. He feels the need to prove that 'its shape must necessarily be spherical' (Aristotle 1952b: 297^{a}9). In his attempt to provide a proof he employs a thought experiment: he considers how the Earth could have acquired its spherical shape (Aristotle 1952b: 297^{a}13–30). He assumes that every portion of the Earth has weight, endowed with a downward movement towards the centre of the universe. Aristotle here appeals to his theory of motion. According to it material objects 'strive' to where they naturally belong, i.e. the geometric centre of the universe (Weinert 2009: 7–9). Hence the reason for the downward motion of 'every portion of the earth' is that an object, which possesses weight—as pieces of earth do—'is naturally endowed with a centripetal movement' (Aristotle 1952b: 297^{a}15–20). And if an equal amount of such material chunks 'strive' towards the centre, they will form a mass with a spherical shape.

Whilst the empirical observations show that the Earth must be spherical, it is the task of the thought experiment, relying on Aristotle's theory of motion, to 'prove' that the Earth is spherical by necessity.

Aristotle's theory of motion, with its central doctrine—that there is no motion without a mover (Aristotle 1952a: BKVII, VIII)—played a central part both in his cosmology and his 'proof' that the Earth occupies the 'centre' of the universe, where it was neither endowed with a daily nor with an annual rotation. Did the Greeks have any 'observational evidence' that the Earth does not move? They believed themselves to be in possession of such evidence: for if it moved, buildings would crumble under the impact of the motion, and such strong easterly winds would blow that birds would never be seen flying from west to east (Ptolemy 1984: §1.7). Aristotle even produced a thought experiment—the so-called tower thought experiment (Fig. 2.2)—to this effect. A consideration of the fall of an object from the height of a tower seemed to show that the Earth cannot possibly perform a daily rotation on its own axis from west to east.

Imagine an object is released, like a stone, from a tower, which sits on a rotating Earth. Would the object fall in a straight line down to the bottom of the tower? A modern physicist would answer in the affirmative but Aristotle came to a different conclusion. According to Aristotle's theory of motion, when the object is dropped from the height of the tower, it 'strives' back to its natural place near the centre of the universe, which is occupied by the Earth. But whilst the body is in free fall, the Earth moves in an eastward direction beneath it. An orbiting Earth would leave the falling object behind. However, no such observations are ever made, from which Aristotle concludes that the Earth must sit motionless at the centre of the universe.

In order to make Aristotle's demonstration move convincing, it can be retold with the insight of modern physics in mind. An object, which is dropped from a height of, say, 50 cm, will descend to the ground in 0.3 s (Fig. 2.2). During this time the Earth will travel 140 m eastward, at a speed of 464 m/s, with respect to a point on the equator. Hence if an object were released even from such a moderate height, it should land 140 m to the west of the bottom of the tower, on the assumption of a

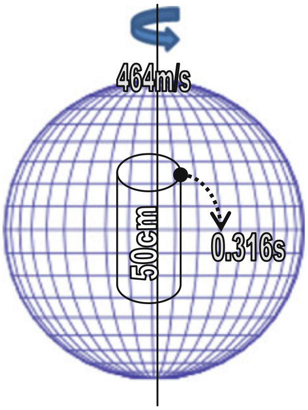


Fig. 2.2 Aristotle's Tower Argument. Although the argument was meant to show that the Earth is stationary, the argument is not valid, because it is based on mistaken premises. *Source* (of sphere): Wikimedia Commons

rotating Earth. A falling object would trail the small tower, which rotates with the Earth—like a person on a spinning wheel—by an impressive gap of 140 m. As such occurrences are not observed, even a 'modernized' Aristotle would conclude that the Earth must be motionless.

The Aristotelian theory of motion, which leads to the stipulation of a motionless Earth, looks as if it were able to explain the appearances: the Earth *seems* to be at rest with respect to the sun, which glides across the horizon from east to west; the released object *seems* to fall straight down towards the centre of the Earth; it *seems* to be eager to return to its natural place. What can be inferred from such examples?

2.1 Some Preliminary Lessons

From the consideration of these thought experiments some preliminary conclusions can be drawn.

1. Thought experiments can be inconclusive.
2. Thought experiments can be misleading.
3. Thought experiments can lead to alternative conclusions.

Ad (1) *Thought experiments are inconclusive*. However appealing Archytas's thought experiment about the infinity of the universe appears to be, it is hardly conclusive. It is an attempt to highlight the logical inconsistency of the Aristotelian assumption that the universe has a boundary. But it has no empirical force, which could disprove the assumption. Archytas does not take into account that an unbounded surface is not the same as an infinite surface. If the universe were like the surface of a sphere it would be finite but without a boundary. The British cosmologist Stephen Hawking indeed made the suggestion that space-time could be finite and yet unbounded if it were described in imaginary time. In imaginary time the universe would have zero size at both the beginning (Big Bang) and the end of time (Big Crunch). The Big Bang starts in a smooth condition, an ordered state, but the Big Crunch corresponds to a collapse into a black hole (Fig. 2.3).

The French physicist and mathematician Henri Poincaré proposed a different response, by way of a thought experiment, which also assumes that the universe is a sphere, but subject to some unusual laws (Poincaré 1952a: 85–86; cf. LePoidevin 2003: 98–99; Huggett 2010: 34–35). In the sphere temperature is not uniform but diminishes towards the edge. It reaches absolute zero at the edge, which constitutes the boundary of the imaginary universe. The temperature, T, varies in such a way that absolute temperature is proportional to $R^2 - r^2$ (where R is the radius of the sphere and r is the distance of a point on the sphere to the centre) (Fig. 2.4). Furthermore, in this world all objects shrink in proportion to their change in temperature as they move away from the centre. 'A moving object will become smaller and smaller as it approaches the circumference of the sphere' (Poincaré 1952a: 65). This world will appear infinite to its inhabitants, since their bodies and measuring

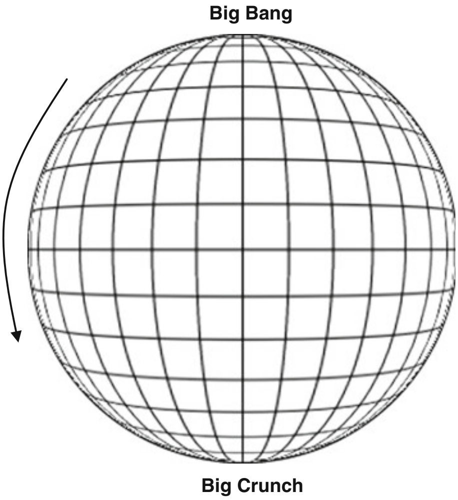
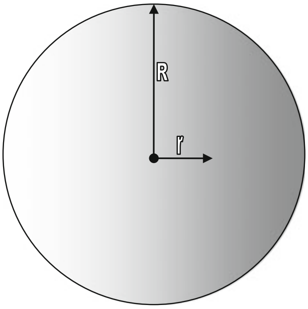


Fig. 2.3 *Hawking's no-boundary proposal* on the analogy of a globe with lines of latitude. The size of the universe increases with increase in imaginary time, as indicated by the *downward arrow*. Note that this cosmological model is asymmetric with respect to time, since the beginning is characterized by smooth conditions, whilst at the end the universe collapses into black holes (Hawking 1988: 138; see Penrose 2005: §25.8)

Fig. 2.4 Poincaré's Imaginary World, in which temperature varies with distance to the edge and objects shrink accordingly



tapes will become colder and smaller as they approach the boundary of the sphere. Even their steps will shrink in such a manner that they will never reach the edge.

Yet another response can be drawn from Leibniz's relational view of space. According to the German mathematician, physicist and philosopher G.W. Leibniz, space is the order of coexisting things, i.e. material objects are constitutive of space. A Leibnizian could argue that as long as there is matter—any kind of matter, even radiation—there is space. On such a view space may be unbounded but still finite since one could always add further material to expand the existing space, as it were. It would be an expanding space, although Archytas may well ask the question: Does the material not expand into a pre-existing space?

Such thought experiments are inconclusive because they are empirically underdetermined. They do not muster enough empirical evidence to secure the conclusion. Aristotle, for instance, could have defended his view of a closed cosmos by pointing out that the fixed stars form indeed the boundary of the material universe, and accept that Archytas's spear-wielding space-travelling demon could have penetrated it. His spear would have travelled through the layer of the fixed stars but not entered the vacuum beyond. This ether-like vacuum constitutes the habitat of the Deity—the Unmoved Mover—but it was no longer to be regarded as physical space. The postulation of an ether, beyond the boundary of fixed stars, would have allowed Aristotle to escape Archytas's conclusion.

Ad (2) *Thought experiments can be misleading*. Aristotle employed his tower argument to 'prove' that the Earth must be stationary. Instead it is the celestial objects —the sun, the planets and the 'fixed' stars—, which circle around the central Earth. The sun occupies the place of the Earth in the heliocentric view. The Greeks generally underestimated the distances of the planets from the 'centre'. Such miscalculations can lead to certain inconsistencies: the 'fixed' stars were said to reside at a distance of 20,000 Earth radii, which is less than today's Earth-sun distance of 150,000,000 km (Zeilik 1988: 29–31). Nevertheless the whole canopy of the fixed stars was supposed to rotate, from east to west, in a 24-h rhythm whilst a planet, like Saturn, which orbits below the sphere of the fixed stars, completes its journey in 30 years.

But the main inconsistency in Aristotle's 'proof' of a motionless Earth derives from his theory of motion. According to Aristotle's theory, every motion needs a mover and objects possess 'natural' places. A stone dropped from the height of the tower 'strives' back to Earth where it naturally belongs. By contrast, smoke rises to the sky, that is, to its natural place. In the thought experiment the tower is attached to the surface of the Earth. It moves with the spinning Earth. But what would be the source of the falling stone's motion? The air, it must be assumed, is not strong enough to give it a push in the horizontal direction of its motion. As it only has a vertical component, the source of its motion is its 'desire' to return to its natural place on Earth. It follows from Aristotle's reasoning that the tower, on the assumption of a spinning Earth, would have a centrifugal motion but not the falling stone. If the Earth turned on its axis the stone should land to the west of the tower because it does not partake of the centrifugal motion of the Earth. But as this displacement is not observed, it must be concluded that the Earth does not spin. So Aristotle reasoned. However, already Nicolaus Copernicus—the first modern proponent of heliocentrism—was able to parry the force of the Aristotelian argument by adopting the medieval impetus theory of motion. According to the impetus theory of motion a projector impresses a certain impetus—a motive force—onto the moving body, which equips it with motion. Applied to the Earth, this means that the motion of the Earth is not a violent but a natural motion. As Copernicus explained, 'the clouds and the other things floating in the air or rising up' take part in this natural motion of the Earth (Copernicus 1543: Bk. I, §8). Equally for the tower argument. The tower, the stone and the experimenter are part of the rotating reference frame and hence take part in the motion of the Earth. The stone falls straight down to the bottom of the tower, not because the Earth stands still, but because it is part of the reference frame, in which the experiment takes place. This phenomenon is well known to every traveller. A train moves at a constant speed in a straight line so that writing, reading, coffee drinking and dropping objects happens in the same way in a moving train as on a stationary platform. Physicists no longer accept the impetus theory but explain the phenomenon by reference to the principle of inertia. An object, if undisturbed by an external force, will either remain at rest or in rectilinear motion. Any object, which is part of the reference frame, will partake of this motion. Therefore an insect in a moving car will buzz around in the same way as in a room of a house. Just by following the erratic flight of the insect an observer will not be able to tell whether the insect is in a reference frame, which is at rest or in uniform motion. Hence the Aristotelian theory of motion is misleading because it is based on a mistaken premise: his theory of motion. As his theory of motion is mistaken, his thought experiment remains inconclusive.

Ad (3) *Thought experiments can lead to alternative conclusions*. Thought experiments can be retold from a different perspective, which may lead to an alternative interpretation of the phenomenon. They are not logically compelling (cf. Gendler 1998; Bishop 1999). Aristotle, Ptolemy and the Greek tradition provide what looks like compelling arguments against the motion of the Earth. But even during Greek antiquity there were some dissenting voices. Hiketas of Syracuse, and Heraclides Ponticus both taught the diurnal (daily) motion of the Earth. Aristarchus of Samos is reported to have taught both the daily and annual rotation of the Earth. But to the Greeks the evidence seemed to weigh so heavily in favour of a stationary Earth that it took some 1400 years before Copernicus was able to resurrect the ancient ideas and put them in a coherent framework. In his heliocentric model, Nicolaus Copernicus displaced the Earth from the centre of the universe. He bestowed on the Earth a dual motion: a daily rotation on its own axis and an annual rotation, from west to east, like the other planets, around the 'central' sun (Weinert 2009: Chap. I). Although Copernicus's work was largely based on the astronomical observations provided by his Greek predecessors, he arrived at a different conclusion, based on the impetus theory of motion. The impetus theory of motion was itself the result of a medieval thought experiment (see Fig. 3.1), whose purpose was to disprove the Aristotelian theory of motion. Such alternative conclusions are possible because thought experiments are inconclusive and empirically underdetermined. They do not replace real experiments. Yet, as the subsequent Chapters on the demons of science will show they are of considerable importance in the history of ideas. Many leading scientists grant them a leading role in scientific thinking.

Given the somewhat uncertain nature of thought experiments, it is not surprising that views differ on how to characterize such mental activities.

Chapter 3
What Thought Experiments Represent

> *Is not the solution now apparent? The demon is simply the complication which arises when we force the world into a flat Euclidean space-time frame into which it does not fit without distortion. It does not fit the frame, because it is not a Euclidean or flat world. Add a curvature of the world and the mysterious disturbance disappears. Einstein has exorcized the demon.*
>
> Eddington, *The Theory of Relativity and its Influence on Scientific Thought* (1922: 28)

An extensive literature on thought experiments exists.[1] The authors try to define or at least to characterize 'what thought experiments are' or to assimilate them to methods and argument patterns familiar in the natural and social sciences. It is probably fair to say that due to the large number of thought experiments in the history of ideas and rational thinking about the world any simple classification is bound to fail. Their real interest lies in understanding their epistemic functions. Their fascination derives from their paradoxical nature: they are examples of 'armchair philosophy', yet seemingly offer the enticing prospect of teaching us new knowledge about the world. Reflecting on their functions in reasoning will help to dissolve this paradox. But in order to identify their functions it will be useful to present a brief summary of the various models of thought experiments, which have been discussed in the literature.

3.1 The Experimentalist View

A natural proposal is to treat thought experiments as extensions, or limiting cases, of real experiments (McAllister 1996). They purport to achieve their aims 'without the benefit of execution' (Sorenson 1992: Chaps. I, VIII). As thought experiments are then 'offshoots' of real experiments, this view implies a continuity thesis. Thought

---

[1] For overviews, see Brown (1991, 2014), Cooper (2005), Genz (2005), Kühne (2005), Sorensen (1992).

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_3

experiments, like real experiments, establish claims in the ‘light of evidence about the world’ (McAllister 1996: 233). On one version of this view, the evidential import is not an intrinsic feature of thought experiments but the outcome of historical accomplishments. That is, the evidence appears as a consequence of accepting certain metaphysical assumptions about the world. One such assumption, according to McAllister, is the distinction between the ‘phenomena’ and the particular circumstances—or natural occurrences—, in which the phenomena manifest themselves. Much of Greek thought, as reflected in the ancient thought experiments, introduced above, was preoccupied with ‘saving the appearances’. That is, the Greeks faced the problem that their theoretical convictions often clashed with the observations. According to most Greek cosmologists, for instance, the planets move in perfect circles around the central Earth. The Greeks were aware, however, that the planets’ motions appear to be subject to certain irregularities: at certain periods they move faster than at other times and even abandon their normal west-to-east motion to ‘retrograde’ for a few weeks in a east-to-west movement before resuming their normal trajectory. Rather than abandoning their assumptions—the centrality of the Earth and the circular motion of all celestial objects—the Greeks designed complicated models, whose purpose was to make the apparent observations compatible with the fundamental assumptions: hence the expression ‘saving the phenomena’. Theoretical presuppositions and observations do not need to clash in this way. In the case of Aristotle’s tower argument, a fundamental assumption—his theory of motion—and the appearances seem to go hand in hand.

Nevertheless in both cases there is an underlying regular process—the phenomenon of motion—and the concrete manifestation of this process in the material world, i.e. the real orbit of a planet or the actual fall of a stone. Thus the observable events—or what the Greeks called the ‘appearances’—seem to be composed of an underlying regularity and the boundary conditions, which render the event possible. McAllister calls the underlying, not-directly-observable regularities, ‘phenomena’ and the observable event ‘a natural occurrence’. The phenomena underlie the natural circumstances, under which the phenomena appear. To mention an example: Newton discovered the inverse-square relationship, which governs the gravitational attraction between any two bodies:

$$F_g = g\frac{m_1 m_2}{r^2}.$$

The expression captures the underlying regularity. But in order to compute the actual gravitational attraction between two given bodies in the solar system, both their masses (m_1, m_2) and their distance, r, must be known numerically. The phenomena are the underlying invariant laws—like Newton’s law of gravity—or other regular processes. The ‘natural occurrences’ are the variable, particular circumstances, in which the phenomena appear. On the basis of this distinction McAllister formulates, with respect to Galileo, the thesis that thought experiments are a source of evidence about phenomena, when it is impossible to reduce the influence of boundary conditions sufficiently to exhibit the phenomena. Thought experiments display phenomena in accident-free form (McAllister 2004: 1168). Thought experiments exhibit non-actual occurrences of phenomena, which concrete appearances may fail to do (McAllister 1996: 245).

Although the experimentalist view sees thought experiments as the continuation, in extreme form, of real experiments, the distinction between ‘phenomena’ and ‘natural occurrences’ undermines this view. Real experiments deal with ‘natural occurrences’ in order to detect phenomena. Thought experiments rely on hypothetical and counterfactual thinking. Thought experiments employ degrees of abstraction and idealization, like models, which real experiments cannot achieve. They may fail to detect real phenomena, as Aristotle’s tower argument shows.

Furthermore the experimentalist view ignores the contestable and indeterminate outcome of thought experiments. Thought experiments are indeterminate because they are empirically underdetermined. Aristotle’s empirical arguments in favour of the spherical shape of the Earth are much stronger than his conceptual arguments. Real experiments, which are often repeated many times with varying boundary conditions, are far less contestable than thought experiments.

Several prominent scientists have stressed the discontinuity between *real* and *imaginary* experiments. Thus Ernst Mach, who is credited with having reintroduced the term ‘thought experiments’ into philosophical discussions about science (cf. Kühne 2005: 165), emphasized that a thought experiment need not materialize in order to serve a purpose. A thought experiment only renders explicit ‘instinctive knowledge’. It does not provide proofs but it furnishes idealizations (Mach 1883: Chap. I). Thought experiments, however, are not open invitations to flights of fantasy. Even a thought experimenter must sail close to the coastline of empirical facts, as both Ernst Mach and the German physicist Max Planck recognized. Planck rejects the view that a thought experiment only acquires significance if it can be realized through measurement (or displays a phenomenon).

First, says Planck, thought experiments employ ‘abstractions’ but abstractions (and idealizations) are as important in science as the empirical findings of laboratory experimentation.

> Nothing is more mistaken than the claim that a *Gedankenexperiment* only has importance insofar as it can always be realized through measurements. If this were true, there would be no exact geometric proof. For every stroke of a pen on a piece of paper is in reality not a line but a more or less small stripe, and every drawn point is in reality a more or less small spot. But we do not doubt the rigid proof of geometric constructions.

Planck is very critical of an experimentalist view but he defends the place of thought experimentation in science:

> (…) thought experiments lift the spirit of the researcher above real measurement tools. They help them formulate hypotheses and new questions, the testing of which through real instruments opens up insights into new lawlike connections, even connections which are beyond the grasp of real instruments. A thought experiment is not tied to precision limits (…). The successful conduct of a thought experiment only depends on the existence of the validity of non-contradictory lawful relationships between the observed events. What cannot exist cannot possibly be found.

Planck then continues to point out the abstract nature of thought experiments:

> It is true that a thought experiment is an abstraction. But the (experimental and theoretical) physicist needs this abstraction in his research as much as the assumption of a real external world. In particular, the great minds and pioneers of physics – men like Kepler, Newton, Leibniz and Faraday – were motivated by their belief in the reality of the external world on the one hand, and the prevalence of a higher reason in and above reality on the other. (Planck 1948: 294; translated by the author; cf. Kühne 2005: 190, 261; Sorensen 1992: Chap. 3)

The experimentalist view loses its appeal if it is not necessary for a thought experiment to be 'continuous' with a real experiment. Even in the absence of continuity thought experiments play an important part in scientific thinking. Perhaps, then, a view at the opposite end of the spectrum is closer to the mark. The Platonic model of thought experiments stands in stark contrast to the experimentalist view.

3.2 The Platonic View

According to the Platonist view thought experiments contribute genuine knowledge about the empirical world. In the terminology of Kantian philosophy thought experiments provide synthetic a priori knowledge—a priori because the knowledge derives from thought processes and *synthetic* because the knowledge is genuinely new, if conjectural knowledge (Brown 1991, 2004). Although E. Mach, M. Planck and A. Einstein all affirmed the power of thought experiments in scientific reasoning, the Platonic view, if correct, would provide a powerful justification for the place of thought experimentation in scientific reasoning. It is therefore worth examining.

Its proponent seems to agree with the tradition, which holds that the function of thought experiments is to test the consequences of theories. Brown proposes a taxonomy: the role of *destructive* thought experiments is to highlight problems with an established theory. The thought experiments of the Scholastics, as explained below, highlight the problems with the Aristotelian theory of motion. So did Galileo's thought experiment about falling objects. According to Aristotle, heavy object fall faster than lighter ones. But what happens, asks Galileo, if a light and a heavy object are tied together? Together they form a heavy object, which should fall faster than the two objects separately. But the lighter object should also slow the fall of the heavier object. *Constructive* thought experiments aim at establishing a positive result, for instance Einstein's famous elevator thought experiment, which establishes the numerical equivalence between motion, due to accelerating forces, and motion due to gravitational forces.

Brown's Platonic view only embraces a small number of experiments, which presumably furnish a priori knowledge of nature. Platonic thought experiments, like Galileo's thought experiment of free fall, are both destructive and constructive. In Galileo's case the thought experiment destroyed the Aristotelian claim that heavier objects fall faster than lighter ones, but established the Galilean view that all objects near the surface of the Earth fall at the same rate of 9.81 m/s^2, independently of their mass. As the American astronaut Neil Armstrong—the first man on the moon demonstrated visibly, on the surface of the moon, which lacks an atmosphere, a hammer and a feather dropped from the same height, hit the ground at the same time. In McAllister's terms, Galileo's thought experiment would have revealed a new phenomenon—the invariant rate of falling objects—which is masked on Earth by the presence of the atmosphere. According to Brown a thought experiment of the Platonic type establishes 'fairly conclusive evidence' for a new 'theory', like Galileo's fall law (Brown 1991: Chap. 2, §2). It does so in an a priori manner because 'it is not based on new evidence nor derived from old evidence' Brown (1991: Chap. 4). In particular Brown favours the view that thought experiments permit scientists to perceive abstract laws of nature.[2] As thought experiments happen only in the laboratory of the mind, the perception of abstract laws of nature cannot be due to ordinary sense perception. Brown, in fact, endorses a special kind of perception:

- Real experiments carry us from *sense* perceptions to a proposition.
- Thought experiments take us from *intellectual* perception to a proposition (Brown 2004: §3).

As is to be expected, many commentators have rejected the appeal to intellectual perception as mysterious (cf. Norton 2004). Even though the claims of Brown's Platonic view apply only to a 'small number of thought experiments', and hence his is a very limited view, Brown's thesis would perhaps have gained a better press if he had appealed to a rationalistic attitude amongst scientists. The language of Platonism recalls Plato's theory of forms—for instance the ideal form of a triangle—which can only be grasped intellectually. As M. Planck observed, every line drawn on a piece of paper is not really a mathematical line but a more or less regular stripe (Planck 1948: 294). Idealization is therefore essential for scientific thinking.

Brown could have struck a careful balance between rationalism and empiricism as in Einstein's attitude to scientific reasoning. Albert Einstein—a master of thought experimentation—rejects the inductive view, according to which scientific principles are derived from experience. Scientific principles and theories, according to Einstein, are free inventions of the human mind. Of his theory of gravitation Einstein said:

> No ever so inductive collection of empirical facts can ever lead to the setting up of such complicated equations (Einstein 1949: 89; cf. Weinert 2006).

In this way, rational thinking can arrive at the formulation of fundamental mathematical relationships—$E = mc^2$—which govern the empirical world. But

---

[2] Brown defends a Necessitarian view of laws of nature as abstract relations between universals. See Weinert (1995: Introduction) for a discussion of various approaches to laws of nature.

such fundamental laws remain conjectures, since experience is needed to confirm the accuracy of the theoretical conjecture. Einstein states that

> (i)n science the logical foundations of physics are always in (…) peril from new experiences or new knowledge. (Einstein 1940: 920)

And furthermore,

> (e)xperience alone can decide on truth. (Einstein 1950: 355)

Brown's Platonism differs from the philosophical tradition by treating a priori knowledge of nature as conjectural, rather than certain knowledge. Nevertheless he believes that Platonic thought experiments allow us to 'grasp relevant abstract universals, which have an existence of their own' (Brown 1991: Chap. 4). In order to grasp the abstract universals—the laws of nature—a special kind of intellectual perception is required, which remains unexplained (Norton 1996, 2004; cf. Clatterbuck 2013). Just as Plato believed the philosopher had privileged access to the world of forms, Brown requires a privileged kind of introspection, which is not guaranteed to be objective and intersubjective. Furthermore, even if a special kind of introspection would allow some privileged minds to grasp the 'laws of nature'—conceived as relations between universals—this metaphysical insight would be of little help, since in the world of empirical science the laws of nature appear as mathematical relationships between well-defined, quantifiable parameters (see Weinert 1993, 1995). Brown's view suffers from a hidden tension. Let us say that a priori—or rationalistic—knowledge of nature is conjectural. Then experience, as Einstein saw, must be the ultimate arbiter of this type of knowledge. Hence it is very much based in the empirical world. As the great thought experimentalists have shown, thought experiments can be of great help in order to arrive at conjectural hypotheses about the world. But then the postulation of a Platonic heaven, in which universals lead an independent existence and reveal themselves only to intellectual perception, is redundant. Hence Brown's Platonic view suffers from three defects:

- If it is acceptable at all it only applies to a small number of cases—how small, how large?—and cannot claim to be a general, unified account of thought experiments.
- The approach requires a mysterious kind of perception and it is unclear who is blessed with this special gift.
- The aim of the intellectual perception is to grasp the 'laws of nature' in an abstract realm. However, reflecting back on the thought experiments introduced above, neither Archytas nor Aristotle seems to appeal to intellectual introspection or even seek knowledge of the laws of nature. Rather their thought experiments seem to take the form of arguments, which have the purpose of critically examining accepted views. Any rational person can follow the argument, without a need for intellectual perception.

As mentioned above, E. Mach and M. Planck both stress the importance, in Planck's words, 'to soar above the world of real measuring instruments' and explore the consequences of scientific theories. Thought experiments involve abstractions and idealizations, which are familiar types of reasoning.

- In *abstraction*, the human mind deliberately factors out certain parameters, which may have a measurable effect on the system under consideration. Thus Newton's inverse square law of gravitation allows the computation of the gravitational attraction between two particular bodies (say the Earth and the moon) but the gravitational influence of all other celestial bodies on the Earth-moon system is deliberately neglected, even though it exists.
- In *idealization*, inaccuracies and small deviations are 'straightened out' to arrive at a pure type, which may be easier to describe or compute. In many models of the solar system, the orbit of planets is depicted as circular, even though it is elliptical, because a circular orbit is easier to calculate than an elliptical one.

In their use of abstraction and idealization thought experiments resemble scientific models. A thought experiment can neglect the messy details of the empirical world and focus on the argument under consideration. Idealizations and abstractions help scientists to respect the empirical constraints, under which science must operate, in a way that introspection does not. Thought experiments help scientists to explore the existence of invariant relationships, which are the objective of scientific work (cf. Mach 1883).

Perhaps the pendulum needs to swing back to a more moderate position, closer to scientific practice. According to one such view, thought experiments involve arguments in an essential way.

3.3 The Argument View

According to the argument view, thought experiments are simply a type of argument. They do not provide a priori knowledge of the natural world. They are not Kantian a priori synthetic principles (Genz 2005; Norton 1991, 1996, 2004; cf. Hempel 1952). Rather they infer, from postulated premises, consequences, which in principle can be tested. But an empirical confirmation of the conclusion does not provide proof of the postulated premises. Recall the tower argument: the object falls straight to the bottom of the tower but this provides no proof of the Aristotelian theory of motion. However, a thought experiment, which highlights logical contradictions, constitutes some scientific progress. Even though it has a destructive role, Galileo's thought experiment shows that a particular theory—like Aristotle's theory of motion—is mistaken or at least contains inconsistencies. It should be kept in mind that thought experiments are inconclusive, hence they cannot provide logically compelling proofs. They furnish insight rather than decisive refutation (Genz 2005: Chap. 1). Consider, for instance, the criticism, which some leading natural philosophers at the University of Paris made of the Aristotelian theory of motion during the late Middle Ages. As a result of this criticism, Nicolaus of Oresme and Johann Buridan adopted the afore-mentioned impetus theory of motion. Buridan and Oresme considered the logical consequences of Aristotle's theory of motion and found it inconsistent. According to Aristotle, every motion needs a mover. Consider then two objects, which are thrown through the air on similar trajectories. Let one be the stone from Aristotle's tower experiments. Its flight path will be a parabola, on Aristotle's theory, because it is subject to two forces. Its 'natural tendency' is to return to the Earth but it disturbs the air, which pushes it forward. Eventually its 'gravity'—or natural tendency—prevails and it returns to the Earth. But now consider the flight of, say, Archytas's spear. In this case the spear has its inherent tendency to return to Earth but it offers the disturbed air far less surface area to push it along; hence it should return to Earth sooner than the stone (Fig. 3.1).

The Aristotelian view can also lead to the opposite result. The stone is heavier than the spear so that it should fall much faster back to Earth than the much lighter spear. Although the disturbed air pushes harder on the stone than the spear the effect of their respective 'gravity' should be taken into account: the 'heaviness' of the stone should return it sooner to the Earth than the spear.

Such logical inconsistencies may not have persuaded the Aristotelians to abandon their theory of motion but they posed sufficient difficulty for the Parisian philosophers to adopt the alternative impetus theory of motion, which became an essential prerequisite for the Copernican revolution.

These considerations seem to show that thought experiments require both human insight into lawlike generalities (whether true or false) and imagination into the possible consequences of thought experimentation (Genz 2005: 60).

But if insight and imagination are to be taken into account where does this leave the argument view? According to the unadorned argument view 'thought experiments are really just dressed-up arguments' (Cooper 2005: 331). A pure argument view has been defended by John Norton in a series of papers (Norton 1991, 1996, 2004). According to Norton's view thought experiments are picturesque adornments, which can explicitly be reconstructed as sober arguments. The main function of the *reconstruction* thesis is to make explicit what are only implicit or tacit assumptions. The arguments must satisfy two necessary conditions:

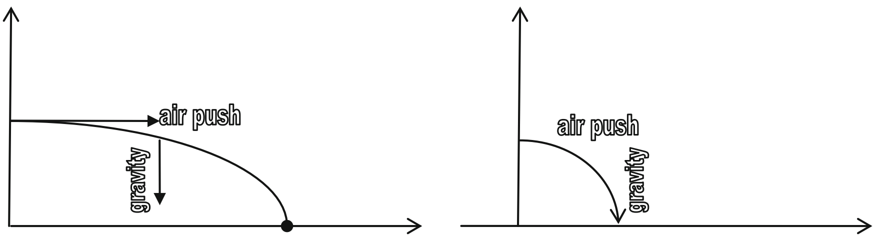


Fig. 3.1 Comparison of the trajectory of stone (*left*) and spear (*right*). According to Aristotle's view the light spear should return to Earth sooner than the bulkier stone, because less air pushes it along, and yet both experience the same trajectories, if thrown with equal force

(i) They posit hypothetical or counterfactual state of affairs.
(ii) They invoke particulars, which are irrelevant to the generality of the conclusion. These particulars, according to the *elimination* thesis, can always be eliminated.

The information gained about the physical world from the conclusion is already contained within the premises, which themselves contain information taken from areas like physics or philosophy. The conclusions are either deductive or inductive inferences with a certain degree of probability. On Norton's view thought experiments are 'inferential devices' (Norton 1996: 335). Thought experiments can fail either because they are based on false assumptions or because the inferences are fallacious. But according to the *reliability* thesis they are nevertheless trustworthy devices because thought experiments are governed by the usual deductive or probabilistic inferences (Norton 2004: 1140). As we have observed in Aristotle's case the reliability of thought experimentation is increased if the premises are based on empirical evidence, like the spherical shape of the Earth. Norton summarizes his argument view in the following statement:

> Thought experiments are just picturesque argumentation of a hypothetical or counterfactual nature. Essentially all that is needed is that the science admits hypothetical or counterfactual reasoning for it to admit thought experimentation. (…) if the science supports counterfactuals, they are admissible. (Norton 2004: 1150)

A good illustration of the reconstruction thesis is Einstein's celebrated 'elevator thought experiment', with which he sought to establish the equivalence of inertial and gravitational mass. Einstein imagined a 'spacious chest' suspended in a 'large portion of empty space' far removed from gravitational influences (Einstein 1920: 60–70). A rope is attached to this space lab, which is home to an observer, equipped with various instruments. An unspecified 'being' then begins to pull the lab 'upwards' so that it acquires a 'uniformly accelerated motion.' Departing from Einstein's original version of the thought experiment we can imagine that this being is a demon, since only a superhuman being could exercise the force in empty space to impart a uniform acceleration to the lab. To the observer inside, this slight amendment makes no difference but to us it is an indication that demons serve useful functions in thought experiments. In the present case, this advantage remains hidden, since the focus of Einstein's thought experiment is not the demon's abilities which the thought experiments of subsequent chapters will explore—but the observer. For the demon the upward motion of the chest is a uniform acceleration, due to the exerted pull. But the observer inside the chest has no idea of the demon's activities and concludes that the lab finds itself in a gravitational field, since objects fall to its floor. Einstein infers from this thought experiment the 'law of the equality of inertial and gravitational mass' (Einstein 1920: 68). That is, a uniformly accelerated frame is physically equivalent with an inertial frame in a homogeneous gravitational field. Norton reconstructs this thought experiment as an explicit argument:

1. In an opaque chest, an observer will see free bodies move identically in case the box is uniformly accelerated in gravitation-free space and in case the box is at rest in a homogeneous gravitational field.
2. Inductive step: (a) the case is typical and will hold for all observable phenomena and (b) the presence of the chest and the observer are inessential to the equivalence. Therefore:
3. A uniformly accelerating frame in gravitation-free space and a frame at rest in a homogeneous gravitational field are observationally identical, but theoretically distinguished, which contradicts a rule for theory construction, i.e.
4. States of affairs which are not observationally distinct should not be distinguished by the theory. Therefore:
5. A uniformly accelerating frame in gravitation-free space and a frame at rest in a homogeneous gravitational field are the same thing (which becomes a postulate of a new theory) (Norton 1991: 137).

Norton clearly holds that if a science permits hypothetical and counterfactual reasoning, it admits thought experimentation. However a study of G.C. Lichtenberg's use of such reasoning will reveal that not all hypothetical or counterfactual reasoning actually amounts to thought experimentation (see Part I, Sect. 3.5). Another reservation about Norton's argument view is that some 'picturesque' thought experiments cannot be reduced to logical argument patterns (Cooper 2005: 332). For instance, a thought experiment may appeal to our imagination—as Archytas's spear-bearing traveller demands—and invite us to imagine what it would be like to 'see' the edge of the world. Einstein, at the age of sixteen, asked himself what could be seen if one rode on a beam of light. In both cases a logical argument could be constructed so that Norton could dismiss the imaginary scenes as 'chaff', i.e. redundant, unnecessary details. But then the argument view deprives thought experiments of their imaginary quality, which itself has a persuasive value. Norton focuses on famous thought experiments (due to figures like Einstein, Galileo, Newton, Stevin) in which conclusions can be deduced explicitly and in which intuition and imagination play indeed a minor role. It is for this reason that Einstein can say that the nature of the 'being' is immaterial. In these thought experiments the premises can be stated explicitly and are uncontroversial. But this does not apply to more intuitive thought experiments, which leave more room to the imagination (cf. Hempel 1965: 164). In such imaginary thought experiments the focus may shift to, say, the actions of a demon who is employed for the purpose of exploring the conceptual coherence of theories. This is the case in Archytas's space-traveller, as well as Laplace's and Maxwell's Demons. In Poincaré's imaginary world a demon's strides towards the edge may not be subject to the temperature dependence of all other objects (Fig. 2.4). In all such cases the job of the demon is to invoke previously unimagined situations in order to explore the implications of the knowledge claims.

The argument view makes thought experiments look more conclusive than they actually are (or ought to be). On the argument view the conclusions of thought experiments are either true, highly plausible or simply false but not indeterminate. Yet two scientists may analyze the same thought experiment and draw different conclusions from it. Although they agree on the type of thought experiment they are discussing, they use two different arguments to arrive at opposite results (Bishop 1999). Thought experiments can be 'rethought' from different perspectives and retooled for different purposes (Bokulich 2001). Thought experiments often operate with non-empirical premises and have a discursive flavour, which the argument view fails to capture. The argument view does not appreciate that the particularities involved in thought experiments—which Norton dismisses as unnecessary bling—can have some epistemic force. They help to persuade (Gendler 2004). The function of thought experiments is also to show how conceptual schemes can be modified or maintained and therein lies their persuasiveness (Gendler 1998, 2004; Kuhn 1964; cf. Norton 2004).

But persuasion can be achieved in two ways. According to the argument view, thought experiments persuade through the logical force of their reconstructed arguments. According to the constructivist view thought experiments persuade through their discursive force, through a 'reconfiguration of internal conceptual space' (Gendler 1998: 420). This reconfiguration may require a 'gestalt switch', a new way of looking at the old world. A *gestalt switch* happens beyond the force of logic. The presence of a mental image may play a crucial role in the formation of a new belief (Gendler 2004: 1162). It may even 'be sufficiently reliable as a source of justification' (Gendler 2004: 1154).

If thought experimentation involves mental modelling, the thought experimenter will be able to mobilize cognitive resources—intuition and imagination, implicit background information, prior beliefs, judgement—which cannot be captured in the argument view (Miščevic 1992). However, if we want to forgo appeal to psychological factors—just as we did with intellectual perception—thought experiments may be best conceived as conceptual models (Part I, Sect. 3.4).

Thought experiments enjoy heuristic fruitfulness. They may lead to new insights but their reliability remains dependent on the trustworthiness of the material out of which they are built. For almost two thousand years Aristotle's tower argument convinced astronomers and natural philosophers alike that the Earth did not turn on its own axis. Yet Aristotle was mistaken because his argument did not take the notion of inertia into account.

The considerations so far have led us to the view that thought experiments cannot be real experiments because they rely on hypothetical and counterfactual reasoning and employ both abstraction and idealization to a large extent. They cannot be Platonic entities: neither do they need a special kind of perception to perform their task, nor do they capture relations between abstract universals. Not all thought experiments can be reduced to deductive or inductive reasoning because insight and imagination may play a heuristic part; and persuasion can be achieved either by appeal to the head or appeal to the heart. Thought experiments may sometimes be instruments of rational persuasion (Sorensen 1992: Chap. 2). But not all forms of hypothetical and counterfactual reasoning amount to thought experiments. What thought experiments do, however, is to provide understanding.

All accounts considered so far capture some aspects of thought experimentation but do not offer a unified view of thought experimentation. A hint of a unified, comprehensive account nevertheless comes from the suggestion that thought experiments may involve ‘modelling in the head’; or more generally that they are conceptual models.

3.4 A Model-Based Account

According to the model-based account thought experiments are attempts to construct models of possible worlds (Cooper 2005).

> To conduct a thought experiment is to make a judgement about what would be the case if the particular state of affairs described in some imaginary scenario were actual. (Cooper 2005: 328–329; cf. Gendler 1998: 338)

Thought experiments address hypothetical or counterfactual scenarios by posing ‘what-if’ questions[3]:

- What would happen if we dropped a stone from the top of the mast of a moving ship?
- What would we see if we observed insects flying around in a closed cabin of a ship at sea?
- What would we see if we travelled to the edge of the universe?
- What would we see if the world were flat?
- How would an observer in a closed lab experience the upward acceleration caused by a demon’s pull on the rope?

In answering such ‘what-if’ questions the thought experimenter tries to construct a ‘coherent model’ of the imaginary scenario under consideration and to evaluate all the relevant consequences. The rigour with which thought experimenters try to answer ‘what-if’ questions differentiates them from daydreams and fiction (Cooper 2005). The result of such considerations can be an ‘internally consistent model of a possible world’ or a template of possible worlds, which may refer to logical or physical possibilities. But in some cases no internally consistent model can be produced, in which case the hypothetical situation is deemed to be impossible. In both cases the thought experiment will have taught us some lessons about the world.

It is worth emphasizing that this particular version of a model-based account does not restrict model-building to mental processes. A thought experimenter is allowed to reason using either a diagram, ‘a set of propositions, a mental picture or even plasticine characters’ (Cooper 2005: 341). As this account employs a broad

---

[3] The first set of questions is of a hypothetical nature, the second set is of a counterfactual nature. But the counterfactual nature of thought experiments should not be exaggerated since some thought experiments have become real experiments (Irvine 1991: 151). A good example of a hypothetical thought experiment, which has turned into a real experiment, is the two-slit experiment in quantum mechanics (see Fig. 17.1).

notion of model it is best, as will be argued below, to think of thought experiments as conceptual models. On this account thought experiments can fail in two ways:

(a) The thought experimenter is unable to answer the 'what-if' question correctly;
(b) the thought experimenter may be mistaken about whether an internally consistent model has been constructed or may be wrong about the consequences, which follow from the thought experiment.

In the tower experiment Aristotle was mistaken about the fall of the stone because he was not aware of the notion of inertia. And Archytas was mistaken about the 'edge' of the universe because he overlooked the fact that the surface of a sphere can be finite and unbounded.

The strength of a thought experiment therefore depends on the reliability of the data, which enter into it. Nevertheless thought experiments are important tools because they help us explore the consequences of our knowledge about the world, both for possible and impossible worlds.

Several authors have stressed that in the exploration of the consequences, the use of 'what-if' questions is important. But what precisely is the relationship between hypothetical and counterfactual reasoning on the one hand and the use of thought experiments on the other? And if thought experiments are conceptual models, how do they fit into the raft of models used in science? It will help to answer these questions if we turn our attention to the work of G.C. Lichtenberg.

3.5 G.C. Lichtenberg's Aphorisms

Georg Christian Lichtenberg was one of the foremost experimental physicists of his age—he designed some 600 experiments (Schöne 1982: §6). He was an Enlightenment philosopher and unique in his use of hypothetical and counterfactual reasoning in his campaign to promote enlightened thinking. The anglophile Lichtenberg is famous for his witty and thought-provoking aphorisms, which are collected in his notebooks—*Sudelbücher*, a word, which he translated himself as waste books (Bd. I, Heft E, §46).[4] Lichtenberg's thought experimentation had a clear purpose, i.e. to investigate alternative ways of thinking and to promote a critical and rational approach to the exploration of the natural and social world. Thought experimentation was Lichtenberg's way of contributing to the Enlightenment project. He fully subscribed to Kant's motto of the Enlightenment: *sapere aude* (I, D121, 425, 434, 536; F441, 860). His own liberal translation (in English) of Kant's motto—'Have the courage to use one's own reason'—reads:

---

[4] This reference refers to Volume I, Notebook E, §46; I will abbreviate references to I, E46 etc. All translations, unless otherwise indicated, are my own.

> *Much pain is taken and time bestowed to teach us what to think; but little or none of either to instruct us how to think.* (I, F432; italics in original)

Lichtenberg admonishes his contemporaries to reason on the basis of facts rather than wallowing in mere opinion (I, D19). In the spirit of the Enlightenment he calls for the critical examination of all doctrines, ideas, thoughts (I, B285, E137). The main rule of philosophy is 'to be attentive (…), to measure and compare' (I, A130), not to trust one's instincts and not to postulate a *deus ex machina* (I, E17). Lichtenberg is highly critical of what he calls 'system dogmatism' (I, F431), which imposes shackles on the progress of science (I, C9, 209, 278). He does grant, however, that thought systems have the advantage of encouraging thinking and providing guidance (I, E497). In order to prevent thought systems from stifling reflection, Lichtenberg ponders whether one should encourage every 100 years 'a general revolution in the minds of people' (I, C78). The purpose of his aphorisms and thought experimentation, so he declares, is to encourage 'cautiousness', not of a general kind but sceptical caution towards dogma and unexamined claims (I, F802).

> All evil in the world can be attributed to unreflective esteem for old laws, old customs, old religions. (I, D369; cf. II, J1634)

In praise of doubt he says, ironically, that 'happy' are those who 'believe everything they wish to believe' (II, G79; K50). Put more positively:

> Doubt everything at least once, even if it is the proposition '2×2 = 4'. (II, K303)
>
> The main point everywhere is to doubt things, which are believed without further examination. (II, J1276)

Doubting everything in the Cartesian sense of a methodical doubt encourages new way of thinking. We can learn from our own mistakes since they teach us 'that everything could be different' (I, J942). But why wait for mistakes to happen, we should 'invent new errors' (II, L886; cf. H73). Lichtenberg encourages his audience to look,

> in everything for something that nobody has yet seen and nobody has yet thought about (II, J1363, 1770)

and he welcomes 'new conjectures' (I, D484) for

> one has to make new things in order to see new things. (II, J1770, cf. 1341, 1352, 1708)

Lichtenberg invents a new device in the service of systematic doubt and alternative ways of thinking: *thought experimentation*, which is his way of contributing to the Enlightenment (cf. Schöne 1982).

> One has to experiment with ideas. (II, K308; cf. H149, KA310, L735)

And the best way of experimenting with ideas is to invent and envisage hypothetical and counterfactual scenarios.

> In all the sciences it can be useful to suppose cases which, as far as we know, do not exist in nature. (II, H20; cf. H178)

Let us think of Lichtenberg's new device—thought experimentation—as a tool box with an assortment of means to experiment with ideas:

1. Lichtenberg asks counterfactual questions.
2. He envisages hypothetical and counterfactual situations.
3. He considers deviations from the habitual rules.
4. He investigates irregular things in nature.
5. He formulates alternative hypotheses and encourages alternative analyses.

1. *Lichtenberg asks counterfactual questions*:

> If a human, having reached an age of 100 years, could be turned over, like an hour glass, to become younger again – always with the usual danger of dying – what would the world look like? (II, K277; cf. II, J1355, 2139, II K289, II L883; I, J547)
>
> Which motion would a planet perform if the gravitational centre changed its position according to a certain law? (II, A201; cf. II, J1284, 1314, 1674, 1874; II K330)

2. *He envisages hypothetical and counterfactual situations*:

> If a tunnel were driven through the centre of the Earth, one could comfortably jump into it and achieve a velocity at the centre (if one were not killed by the air) thanks to which one could reach the other end and arrive comfortably. (I, A200; cf. II, J1355)
>
> If one grafted alien roots onto the trees, what consequences would it have? (II, J1340)

3. *He considers deviations from the habitual rules*:

> One must try not only to investigate nature but to try completely different methods. (II, J1991; cf. J1781, 1329)
>
> Habit ruins our philosophy. (II, H21)

4. *He declares it useful to rescale things and consider them in different dimensions*:

> If one could imagine the Mediterranean in miniature, one would run the risk of finding it dry on a warm day. (II, J1719; cf. J1488, J1645)
>
> A fruitful mother of new ideas is the rule: to increase everything in order to see what would happen if the biggest things could grow properties. (II, J1644)
>
> Look for everything on a large scale what one observes on a small scale and vice versa. (II, J1666; cf. II J1821; K301; L732)
>
> A good method of discovery is to think away certain parts of a system and to discover how the rest would behave. (II, J1571)

5. *Lichtenberg engages in alternative analyses*:

> One has to try everything (II, L735, 861), [since even] monstrous thoughts have their use. (II, J1380)
>
> The general rule is indicated in the question: If certain circumstances were changed, what deviations would they suffer? (II, KA329, D765)

Such alternative analyses lead to an exploration of new hypotheses. One should consider what is known under the aspect of the unknown (II, KA299, 295, 340; cf. II, J1363).

> If light could push away transparent bodies, how would it push them away? And what would happen to a glass bowl if it were exposed to light? (II, J1569)
>
> Even artificial, false and unlikely hypotheses have their use. (II, J1360; cf. J1520, J1521)
>
> These theories are artificial systems which in the absence of natural systems have their use. (II, J1774)

There is much to learn from Lichtenberg's thought experimentation, since it makes ample use of hypothetical and counterfactual reasoning. According to the model-based account, all thought experiments can be formulated as 'as-if' questions. But, as we must conclude from Lichtenberg's aphorisms, not all 'as-if' questions lead to genuine thought experiments. Consider the following examples of 'as-if' scenarios:

> If all people were petrified in the afternoon at around 3pm … (II, E207)
>
> If dogs, wasps and hornets had the gift of human reason, they could perhaps conquer the world. (I, J360)
>
> If there were only beetroots and potatoes in the world, someone would perhaps express regret that plants stood upside down. (I, C272; cf. IA39)
>
> If humans could change their bodies like clothes, what would happen to them? (I, F292; cf. I, J1151)

These 'as-if' questions clearly display hypothetical and counterfactual credentials but Lichtenberg does not pursue the consequences of such 'as-if' scenarios. In the language of the model-based account, Lichtenberg constructs 'possible worlds' but he does not investigate the conceptual consequences of the 'as-if' scenarios. Lichtenberg praises the virtue of thought experimentation by asking hypothetical and counterfactual questions about the world—'thought games, to which nothing objective may correspond' (II, H149)—but he does not build model worlds to investigate their consequences. In brief, Lichtenberg practised imaginary experimentation, but not in the modern sense of thought experiments. Thus, if not all 'as-if' questions have the character of thought experiments, what is the function of thought experiments as they are understood today? Before that question can be answered, we should ask what kind of models thought experiments are. How do they compare to other models in science?

Chapter 4
Models and Thought Experiments

Thought experiments are models without the formal apparatus.
Frigg (2010: 123)

As the motto indicates an association exists between models and thought experiments. The role of models in scientific thinking has recently received much attention in the literature (Morgan/Morrison 1999; Bailer-Jones 2009). But the question of how thought experiments fit into these considerations still needs to be clarified. As it turns out they fit nicely into a well-chosen category of models. Recent discussions about models focus on the following questions:

1. What is the *role* of models in scientific thinking? How do models differ from scientific theories?
2. What *types* of models can be distinguished?
3. How do models *represent* reality?

Let us consider these questions in turn and ask how thought experiments compare to other models used in scientific reasoning.

4.1 Models as Mediators

The prevailing view in the literature seems to be that 'models are mediators' between theories and the empirical world (cf. Cartwright 1999; Morgan/Morrison 1999; Suárez 1999). Theories are very abstract and general entities, whose principles apply to a certain domain of the empirical world. Models are more concrete and particular entities, which represent specific systems in the domain of the theory. The function of models is manifold in science. They may help in the development and exploration of theories, as well as their testing. Models also fulfil an important representational function and, crucially, they provide understanding. Some sophisticated models, like a structural model, whose features are close to a scientific theory, may also lead to predictions (cf. Hartmann 1999; Bailer-Jones 2009:

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_4

Chap. VIII). A convenient catchphrase is that 'theories explain the phenomena' by providing the formal and mathematical framework, against which the functioning of the empirical world can be explained. Models, by contrast, provide understanding of the workings of particular systems. A false theory will fail to provide genuine explanations but a false model may still provide a coherent account of a system modelled, i.e. a model provides understanding. This catchphrase is misleading but it will serve as a useful starting-point.

What does **understanding** mean? As the ancient thought experiments indicate, 'understanding' provides a plausible account of the phenomena. A model tells a plausible story of a slice of reality or a range of data, which may otherwise make no sense to the observer (Hartmann 1999). Consider, again, the motion of the planets. The Greeks believed that the planets literally circled the central Earth, which itself experienced neither a daily nor an annual rotation (see Weinert 2009: Chap. I). The Earth occupied the centre of the universe, and all other celestial objects—the six known planets in antiquity, the sun and the 'fixed' stars—performed circular rotations around the stationary, motionless Earth. Since the work of Nicolas Copernicus (1543) it is known that this ancient geocentric worldview is mistaken. The Earth is not the centre of the solar system, let alone the universe. The sun lies at a focal point of the planetary orbits, around which the planets perform elliptical orbits. The daily rotation of the Earth on its own axis creates the impression that the stars rise daily in the East—like the sun—and set daily in the west. Although the geocentric theory was mistaken and failed to explain the workings of the solar system, the geocentric model made sense of the appearances.

Models do not have to be accurate to provide understanding of the observations. But there is a limit to their allowed degree of inaccuracy. In order to be useful models must, like thought experiments, sail close to the empirical facts. Aristotle, for instance, proposed a concentric model of the planetary system, according to which the planets move in concentric shells around the centre. This model, however, was soon discarded because the Greeks realized from observations that the planets move around the centre at different speeds and that they change their distance from the centre as they complete their annual orbits around it.

Models do not have to provide an accurate description of 'reality' but modelling must make sense of the observational or experimental data. In order to make sense of the data and to provide a 'plausible story' models employ a number of techniques: abstraction, idealization, factualization and systematization, which they partly share with thought experiments.

Abstraction means that certain parameters, which are known to have a negligible effect on the system modelled, are removed in the modelling process, although they are clearly part and parcel of the real system, which is the target of modelling. Scale models of the solar system, for instance, may leave out the moons of Jupiter and Saturn, since these small objects do not have any significant impact on the accuracy of the modelling process.

Idealization means that the properties of a given real system are simplified for the purposes of manipulating the parameters in the model. For instance, for the purposes of calculation it is often assumed in models of planetary motion that the orbit of the planet around the centre is circular, rather than elliptical, because it is easier to compute a circular than an elliptical orbit, with little loss of accuracy.

An ideal pendulum combines both techniques: the period of an ideal pendulum—$T = 2\pi\sqrt{l/g}$—is only dependent on the length of the string, l, and the gravitational constant, g, but not on its mass. The ideal pendulum, which is a model of a real pendulum, abstracts from the mass of the oscillating bob (and the string), and idealizes the string to 'a weightless, inextensible cord of length, l (see Bailer-Jones 2009: §6.6 for a more detailed discussion).

Factualization means that a model can be approximated to a real system by including abstracted factors and reducing the amount of idealization. But even a physical pendulum, described by the equation:

$$T = 2\pi\sqrt{J_A/mgs}$$

will only be valid for amplitudes smaller than 8°. (J_A is the moment of inertia about the point of suspension, m is the mass of the pendulum, g is the gravitational constant and s is the distance between the point of suspension and the centre of mass).

Apart from abstraction, idealization and factualization a model has another function: *systematization*. Systematization means that a model typically combines various factors into a coherent representation. Consider again the solar system. It is a perfect illustration of what a system comprises: a system consists of relata and relations. In other words, it consists of some constituents (for instance planets), which are held together by some relation (Newton's law, a statistical regularity, or a mechanism). Imagine a Laplacean Demon whose powers allow him to pluck the planets from the sky and place them, like billiard balls, on top of a table. The Demon would have destroyed the solar system, since the planets are now reduced to a row of isolated objects. Having performed this destructive act, the Demon can now proceed to re-assemble the 'planetary balls' into a system. In order to perform the constructive task, the Demon must do two things: (a) the planet objects must be arranged in a particular order. Since the Greeks different types of planetary orders have been envisaged by astronomers. They led to different planetary models. The Greeks adopted a geocentric order, with the Earth at the centre (Fig. 4.1a).

In 1543 Copernicus proposed a heliocentric order, which basically consisted in a swap between the positions of the Sun and the Earth (Fig. 4.1b). The Danish astronomer Tycho Brahe, however, sought a compromise system, according to which the Earth becomes again the centre of the then known universe, with the Moon and the Sun orbiting around it in circular orbits, whilst the other planets are made to circle around the Sun (Fig. 4.1c).

But having decided on the spatial arrangement of the planets, the Demon's work is not yet done. In order to recreate a *system*, the Demon must decide (b) how the

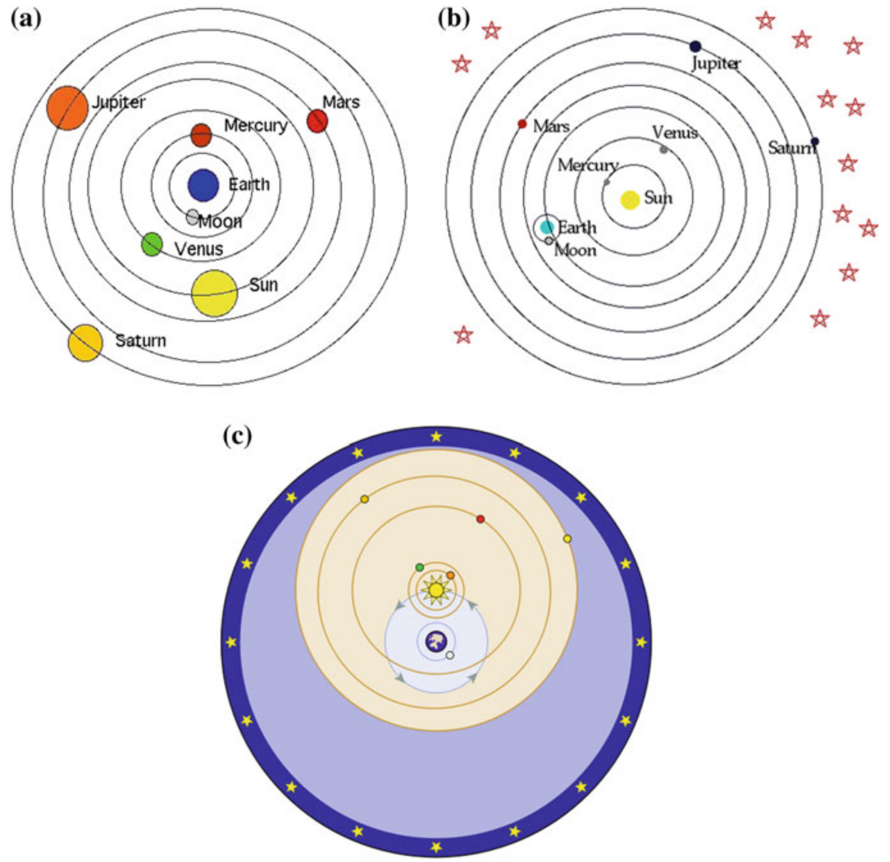


Fig. 4.1 **a** Geocentric arrangements of planets, according to Greek astronomy. **b** Heliocentric arrangement of planets. *Note* that the outer planets—Uranus, Neptune, and Plato—were discovered much later. **c** Tycho Brahe's compromise system, between geocentrism and heliocentrism. *Source*: Wikimedia Commons

planets are to be related to each other, in terms of some regular motion. Without investing the system with some motion, it is not a *planetary* system. This motion must be regular but in the history of astronomy two basic types of regular planetary motions were distinguished. A long line of outstanding astronomers from Ptolemy to Copernicus believed that the planets moved in perfect circles until Johannes Kepler stated, in his first law of planetary motion, that the orbits of planets around the sun were elliptical, not circular.

Having decided then on the spatial arrangement of the planets and how their motions are related to each other, the Demon has recreated a system, which can be returned to its rightful place in the heavens.

The thought experiment with the playful Demon illustrates that models often represent either a *topologic* structure—a spatial arrangement of the components—or an *algebraic* structure—the mathematical relations between the relata.[1] Some models—like structural models—combine both features (Weinert 1999, 2000). It is then natural to ask what types of models can be distinguished.

4.2 A Typology of Models

There is no agreed typology of models. One possibility is to distinguish between empirical and theoretical models or between representative and interpretative models (Cartwright 1999). Such distinctions fail to do justice to the full range of models, which exist in the natural and social sciences. In previous publications (Weinert 1999, 2000, 2006) the author has distinguished analogue models, scale models, functional models, hypothetical models and structural models. But this earlier classification did not include thought experiments. In line with the model-based account thought experiments count as conceptual models. Analogue models, on account of their proximity to thought experiments will also be classed as conceptual models. Like thought experiments, they employ analogies and metaphors. The revised classification, then, comprises conceptual models (including analogue models and thought experiments), scale models, functional models, hypothetical models and structural models. Conceptual models are employed frequently in philosophy as well as in the natural and social sciences.

- **Conceptual models** now include both analogue models and thought experiments.
  - *Analogue models* represent the unfamiliar or unobservable in terms of the familiar or observable. This type of model suggests that there is an analogy between certain elements of already known systems—say water and water molecules—and some elements of unknown systems—say electricity and electrons. Analogue models are based on formal or material similarity relations. Kepler wanted to understand the physical cause of planetary motion. He knew nothing of gravity so he used the analogy of a loadstone and its magnetic properties to rays coming from the sun, ensnaring the planets. The sun seemed to attract the planets like a magnet attracts a piece of metal. But as the planets do not fall into the sun, Kepler imagined that planets had 'friendly' and 'unfriendly' faces on which the sun's rays had different effects (attraction and repulsion):

> (…) the solar body has the force (*vim*) to attract the planet with respect to its friendly part and to repulse it with respect to its unfriendly part (…). (Kepler 1618–1621: 58)

---

[1] Kepler's third law of planetary motion establishes a relation between the speeds of planets in different orbits, P, and their average distance from the sun, A: $P^2 \cong A^3$.

  But the mere analogy does not ensure that the real systems will resemble the analogue model. The sun does not 'lead' the planets by magnetic rays; and planets do not display 'friendly' or 'unfriendly' faces. Analogue models are a useful, if limited tool in an attempt to achieve physical understanding. They suggest useful approaches to problem situations often exploiting visual resemblances between the models and the system modelled. However, if models are to represent natural systems and their structural features, more is needed than analogies. To achieve real physical understanding we need more sophisticated models.

- Many conceptual models occur in the form of *thought experiments*, both in philosophy and in the natural and social sciences. Such conceptual models are often of a qualitative nature, creating conceptual systems, whose job is to probe the world of facts or the world of ideas. In such conceptual models hypothetical scenarios are envisaged or counterfactual questions are asked. Thought experiments, more than analogue models, are the prime vehicle for conceptual models. Such conceptual models, as Lichtenberg has demonstrated, allow us to apply counterfactual reasoning in such diverse fields as history, political philosophy and the social sciences.

  - A historian may wonder, counterfactually, whether fascism would have taken root in Germany if Hitler had been killed by a bullet in a protest March in 1923 (Kershaw 1998). Historians may investigate, counterfactually, the economic effect traditional means of transport would have had in the United States, if the railways had not been developed (Fogel 1964). Max Weber, as discussed below, believed that counterfactual questions had a legitimate place in the social sciences.
  - Plato's construction of the Ideal State in his *Republic* is a political thought experiment. Such an ideal political state, in which philosophers become the rulers of the republic, does not exist in reality. But this political thought experiment enabled Plato to investigate the ideals of fairness and justice, which must characterize the ideal state. Although it is an ideal, real governments can be measured as approximations to it.
  - The 'state of nature' in Social Contract theories of the Enlightenment period is also a conceptual model, which takes the form of a thought experiment. Before the formation of the state, human agents are envisaged to live in an unsatisfactory state of nature. The state of nature is undesirable because it leads, in Hobbes's pessimistic view, to a war of all against all. The members of such an original position therefore come together to form a political community, according to the rules laid down in the social contract.
  - As discussed below (Part I, Sect. 7.1) Freud employed a Demon, similar to Maxwell's Demon, to increase understanding of his tripartite model of the mind.

Thought experiments employ hypothetical and counterfactual reasoning, whilst other types of models represent physical systems, to various degrees of factualization.

- **Hypothetical models**—or *as if* models—incorporate idealisations and abstractions. They claim to represent the system modelled *as if* it consisted only of the parameters and relationships stipulated in the model. Graphic representations of the solar system (Fig. 4.1a, b, c) are typical hypothetical models. They represent the solar system *as if* it consisted only of, say, six planets, without moons, and *as if* they orbited the sun in circular orbits. However, it is known that such idealised factors are mathematical simplifications and that abstracted factors are at work in the real systems.

  Hypothetical models also play an important part in the social sciences. Ideal types are hypothetical constructions of a socio-economic, political or historical nature which seek to delineate pure cases (Weinert 1996). They are abstracted from the empirical data but with complete disregard for their diversity. Max Weber did important work on the methodology of the social sciences. He was concerned with logically precise conceptions and not with their exact correspondence to empirical cases. His pure types of legitimate authority—which are subdivided into charismatic, traditional and rational forms—were not expected to be *descriptions* of historico-empirical realities. Rather, they were understood as ideal limits against which empirical cases could be gauged. Once such pure types were constructed, empirical occurrences of social action could be regarded as 'factors of deviation' from the ideal type. The ideal type is a conceptual construct (*Gedankenbild*), which is neither historical reality nor even the "true" reality. It is even less fitted to serve as a schema under which a real situation or action is to be subsumed as one *instance*. It has the significance of a purely ideal *limiting* concept with which the real situation or action is *compared* and surveyed for the explication of certain of its significant components (Weber 1904: 93).

  The social or historical reality will then appear as a departure or deviation from the ideal type construction, just as the behaviour of real economic agents comprises deviations from how the ideal-typical economic agent *would* behave. The deviations are brought about by 'irrational factors' (Weber 1968: 4–6).

  Given this function of ideal types, Weber distinguishes several tasks, which might be accomplished with their help: *First*, the ideal type allows the social scientist to conceptualise certain historical phenomena or movements to a maximum of conceptual clarity—examples are 'imperialism', 'feudalism', 'mercantilism' (Weber 1904: 92). The ideal type highlights the essential components of, say, feudalism, even though no such pure form of feudalism may ever have existed. *Second*, the ideal type serves as a limiting concept (*Grenzbegriff*), whereby historical individuals, like 'market economy', 'church' and 'sect' can be sharply distinguished by reference to ideal type constructions. But Weber warns against a characterisation of the ideal type as a representation of the essence of reality or of some underlying structure (Weber 1904: 94). *Third*, ideal types may be used to reconstruct the 'developmental sequences' in history where this reconstruction is, however, not to be identified with the actual course of history (Weber 1904: 101). Weber cites Marx as an example for this use of ideal types:

> All specifically Marxian "laws" and developmental constructs, in so far as they are theoretically sound, are ideal types. (Weber 1904: 103)

Fourth, ideal types help to formulate a notion of causation in the social sciences. The notion of adequate causation is Weber's attempt to introduce causal relationships into socio-historical studies. Counterfactual questions, he asserts, are not idle in the study of history. By considering what *would* have happened *if* certain conditions had either been absent or modified, Weber hopes to throw light on the 'historical significance' of the actual determinant factors in the emergence of some historical event. While there is an infinity of determining factors, the 'attribution of effects to causes take(s) place through a series of *abstractions*,' (Weber 1905: 171; italics in original) guided by the interest the historian has in the event. Thus through abstractions, isolations and generalisations, the historian is to construct a complex of possible causal relations 'in order to identify the conditions which are the likely cause of an actual historical event or events of that type' (Weber 1905: 184–185).

- **Scale models** represent real-life systems either in reduced size (the solar system) or in enlarged size (planetary models of atoms). Scale models are usually three-dimensional and require a fairly precise knowledge of the structure of the system. As indicated above, the history of astronomy shows that an accurate representation of the solar system was difficult to obtain.
- **Functional models**, as the name suggests, represent the functional dependence between several parameters. For instance, in economics, supply and demand curves show that the price of a commodity is fixed, where the supply and demand curve meet. There is no need to assign precise values to the symbols, which stand for the variables. What counts is the nature of the *functional* relationship between some parameters. We obtain a functional model, if the functional relationship between various parameters is represented in a diagram or graph. A functional relationship is captured in *Bode's law*. It was discovered by Johann Titius but it became better known through Johann Bode (1772). Bode's law states that the distance of the planets from the sun (measured in units of the Earth-sun distance, AU) follows the rule:

$$r_n = 0.4 + 0.3 \cdot 2^n (n = -1, 0, 1, 2 \ldots 8)$$

  Thus the distance, r, varies with the exponent n. When n = 1, for instance, we find $r_n = 1$, which is the distance between the sun and the Earth in the chosen units. When n = 4, $r_n = 5.2$ (AU), which is the distance of Jupiter from the sun. In these models, the basis of representation begins to shift from the topologic to the algebraic structure.
- **Structural models** typically combine algebraic and topologic structures in order to represent how some underlying structure or mechanism can account for some observable phenomenon. Structural models are very useful in the representation of macroscopic systems, like planetary systems, and microscopic systems, like atoms. Kepler's heliocentric model combines Copernicus's topologic structure

> of the solar system with an improved algebraic structure. Copernicus's geometric arrangement of the planets is structurally correct, since he places the sun at the centre of the solar system and arranges the then known planets in their correct order from the sun. The failure of his model lies in its algebraic structure, since Copernicus still assumed that planets move in circles. Kepler's achievement was to replace circular by elliptical orbits and to formulate his laws of motion. Once the topologic heliocentric structure is combined with Kepler's planetary laws and later Newton's theory of mechanics, a fairly accurate structural model of heliocentrism emerges.

But how are we to understand newspaper headlines, which tell us that scientists use worms as models to study, say, the development of animal embryos?[2] The soil worm, as a simple organism, represents in a simplified form embryo development in other species, but also sheds light on a number of diseases, like strokes and heart attacks. Cell division, the growth and death of cells are dynamic processes. The worm model is used as a hypothetical model because the development of animal embryos is simulated as if it followed the worm model.

This characterization of various kinds of models brings us to the third consideration in connection with models, namely the question how models represent. If models are, roughly, mediators between theories and phenomena, then models are used to connect abstract theories with the empirical world. Models represent particular systems; they concretize the abstract parameters (and relations) to be found in theories. The question then becomes how models represent the systems they model.

4.3 How Models Represent

Models have various functions, including representation and providing understanding of particular systems of the world. In building a model a kind of 'representative structure' is created (Morgan/Morrison 1999: 33, Chap. 2.2; Hartmann 1999: §2). The various kinds of models—from scale to structural models—use techniques like abstraction, idealization, factualization and systematization to perform their functions. Thought experiments can now be seen as conceptual models, which envisage hypothetical and counterfactual scenarios (although not all 'as-if' questions lead to thought experiments). Like genuine thought experiments, models must be subject to some constraints, if they want to perform their functions of representing aspect of the empirical world.

Representation can be thought of as a relationship between a symbolic construct (a model, a diagram, an equation) and a phenomenon:

---

[2] In 2002, H.R. Horwitz, S. Brunner and J.E. Sulston won the Nobel Prize in Physiology or Medicine for their discovery of a process called 'programmed cell death' (reported in the *New York Times* October 8, 2002).

> A model user, S, employs a representational device X (a model, diagram etc.) to represent W (aspect of the world) for particular purposes, P. (Cf. Bailer-Jones 2009: Chap. 8; Giere 2004; Suarez 2010: §1)

It may be tempting to think of representation as similarity, resemblance or even structural isomorphism between the model, as for instance in scale models of planetary systems. But this would not be generally true. In a thought experiment—as a conceptual model—a counterfactual scenario is enacted—like Archytas's space-traveller—but it is the experimenter who creates the possible world. Hence there is no basis of comparison between the model world and the real world.

In a *scale model* of, say, the solar system the spatial arrangement of the planets will resemble the order of the real solar system: *Sun, Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, Neptune and Pluto.* It represents the topologic structure. But even though the geometric order is correct, the model cannot truthfully reproduce the distances between the planets.

In a *structural model* the topologic and algebraic structures are combined but the algebraic structure does not resemble in any true sense the system modelled. Consider Kepler's third law: $P^2 \cong A^3$, which states that the average period of a planet's orbit squared is approximately equal to the average distance cubed. Thus the law describes the orbit of the planet in average terms but the actual orbit of the planet does not 'resemble' Kepler's third law.

A *functional model* does not display any similarity with the system modelled. Rather, a functional model stresses what J.S. Mill called 'concomitant variation', i.e. how several parameters vary with respect to each other.

Models only represent aspects of a system and they do so under the operations of abstraction and idealization. Hence terms like 'similarity' and 'resemblance' are ill-fitted to characterize the notion of representation. There may be no corresponding system to which the model may be compared; or there exists no resemblance or similarity between the model and the system modelled. Similarity and isomorphism are neither necessary nor sufficient for representation. Similarity cannot account for idealized representations or for counterfactual worlds, as they appear in thought experiments. Isomorphism cannot account for inaccurate or partial representation. Furthermore, while similarity and isomorphism are equivalence relations—they are symmetric, transitive and reflexive—representation is non-symmetric, non-transitive and non-reflexive (see Suárez 2003, 2004, 2010). A further problem with the similarity/isomorphism/resemblance account is that the same system can be modelled in different ways. And these ways of representing may contradict each other (Bailer-Jones 2009: Chap. VIII; Suárez 2010: §3). The history of science shows many incompatible models of the solar system, of the atom, of evolutionary processes, of the unconscious, with varying degrees of (in) accuracy.

There is therefore a need to find a conception of representation, which differs from the afore-mentioned views. If a model is to represent certain aspect of a system, it must satisfy criteria of accuracy or constraints so that it remains possible to differentiate fiction and reality. A model must, to a certain degree, 'fit' certain aspects of the system modelled. A plausible thesis about representation is: models represent, if they display an appropriate degree of 'fit', where 'fit' is explained in terms of satisfaction of constraints (Weinert 1999, 2000, 2006; cf. Giere 2004; Bailer-Jones 2009: Chap. VI, §§3–4, Chap. VIII).

Constraints can be understood as restrictive conditions on models and theories. Constraints limit the admissible input of data into the model. But it must be kept in mind that constraints can vary, they can be broken and that they may change their status. Consider a gatekeeper at a playground for young children. The gatekeeper's job is to admit only children up to the age of ten. They will typically be accompanied by their parents or some other adult known to the child. The gatekeeper imposes a restrictive condition on admissibility. Children under ten are admissible, including an authorized adult who accompanies them. Children above ten and adults without children are not admissible. This constraint may change if, for instance, the park authorities decide to raise the age of admissibility to, say, twelve years of age.

In the case of scientific models, constraints are imposed from both theories and phenomena (Bailer-Jones 2009: 152). It may therefore be appropriate to speak of a *constraint space*, in which models are embedded. If models are to perform their function of representation properly, it is important that they operate in such a constraint space.

First, the models are subject to constraints from the direction of theory. Theories are abstract structures, which impose on modelling theoretical constraints, like logical consistency, which the model must satisfy.[3] Second, models are also subject to empirical constraints, like agreement with observational data, without which a model would not 'conform' to the system modelled.

To illustrate, consider the case of astronomical models (whether geocentric or heliocentric) prior to Kepler's discovery of his first law of planetary motion, which states that planets move in elliptical orbits around the sun. All model builders prior to Kepler's discovery (1605) imposed the theoretical constraint that planets must move in circular orbits around the centre, which was usually taken to be the Earth. This theoretical constraint was not justified on empirical grounds—say as a result of observations—but was a metaphysical assumption, which had its roots in ancient Greek thinking. According to the Greeks the universe was divided into two spheres: a sublunary sphere, characterized by change and imperfection, and a superlunary sphere, between the moon and the 'fixed' stars, which displayed harmony and perfection. As the planets move in the superlunary sphere of perfection their motion must be harmonious and perfect; such characteristics can only be satisfied by the geocentric circle. Hence planets must move in circles (see Weinert 2009: Part I). The Greeks, however, were aware that planets did not move with circular regularity around the centre. They neither keep the same distance from the centre nor do they travel at the same speed around it. The Greeks needed to accommodate their

---

[3] As discussed below, Kuhn appeals to the values of accuracy, consistency, scope, simplicity and fruitfulness (Kuhn 1983).

observations to their theoretical constraints (their preconceptions) in order to 'save the appearances'.[4] The Greeks constructed models, which apparently accounted for the appearances of 'planetary motion', without however abandoning their theoretical constraints (e.g. the circularity of planetary motion and the centrality of the Earth). Greek models were embedded in a constraint space, which changed considerably under the impact of the Scientific Revolution (1543–1687).

We can imagine a constraint space as being constituted by a number of different constraints.

- *Empirical constraints*: the availability of reliable empirical data (experimental and observational results); fundamental physical constants (like Planck's constant h and the velocity of light c); the existence of empirical laws—say Kepler's laws of planetary motion or Snell's law of refraction—whose main feature is that they apply only to a limited domain of phenomena and that they can be derived from more fundamental laws (Weinert 1993).
- *Theoretical constraints*: these are physico-mathematical principles (like the principles of Special and General relativity and symmetry principles); methodological norms (like unification, logical consistency within a theory and coherence with the established cannon of knowledge; testability and fruitfulness); mathematical requirements (like numerical predictability, differentiability of functions).
- *Metaphysical constraints*: such principles appear in the form of postulates like the uniformity of nature, determinism or indeterminism, causality and predictability, perfection and harmony in nature. On the philosophical level, they are often associated with demons. Prior to the Scientific Revolution, metaphysical principles did much of the 'explanatory' work, with which the world around us was 'understood'.

> Nature does nothing in vain.
> Nature abhors the vacuum.
> Nature makes no jumps.

The Scientific Revolution did not abandon all metaphysical assumptions, since it did not ban the demons. But it replaced such metaphysical by mathematical explanations. Some models satisfy more of these constraints than others. Thought experiments, as conceptual models, satisfy empirical constraints in the sense of being compatible with a physically possible world; they satisfy theoretical constraints, like logical consistency, conceptual coherence, and fruitfulness. Thought experiments even satisfy metaphysical constraints, like determinism, since Laplace's Demon sets out to demonstrate that the universe is like a gigantic clockwork (see Part II). Other demons, like Maxwell's Demon, were employed to

---

[4] A modern Popperian philosopher would demand that the observations must be able to 'falsify' the theoretical assumptions, which consequently ought to be changed or even abandoned. The Greeks, however, adapted their assumptions to the observations.

show that the world was probabilistic to a certain extent (Part III). Nietzsche's Demon entertained grand cosmological visions of an eternal recurrence of events (Part IV).

Having thus clarified the 'nature of constraints' we can finally say what it means that 'fit' consists in the satisfaction of constraints. As models satisfy constraints to a lesser or greater degree, fit comes in degrees. It is not a one-to-one mapping of theoretical and empirical elements. It does not require 'similarity', 'resemblance' or structural isomorphism. A model 'represents' a section of the empirical world, if it satisfies a certain number of constraints. The more constraints a model satisfies, the better its representational accuracy. 'Fit' means that the model structure successfully accommodates the empirical and theoretical constraints in its constraint space (Weinert 2006). But as the constraint structure can undergo change, the representational accuracy of models will change too. Constraints should enhance the representational force of a given model. Constraints should be testable and justifiable. Representation therefore has a pragmatic function: 'fit' is dependent on the availability of constraints and the inferences, which a model user wishes to draw about the target system (cf. Suárez 2004; Bailer-Jones 2003). However, there must be a limit to representational freedom, if the representational function of models is to succeed. The constraints encapsulate objective information about the target system, in terms of the topological and algebraic structure of the system. What is required of the source system, model A, is that it has an internal structure that, at a minimum, gives informed agents the possibility to correctly draw inferences about the real system, target B (Suárez 2004: §3; Morgan/Morrison 1999: §2.3.2; Morrison 1999: §3.5).

How does this characterization do justice to thought experiments? On the model-based account these experiments explore the consequences of counterfactual and hypothetical scenarios, or possible worlds, for the purpose of probing our knowledge of the actual world. It is of course possible to construct possible worlds, which defy the laws of the physical world. In some cases this disregard for the laws of nature will lead to imaginative fiction—floating mountains in the film *Avatar* or time travellers who change the past—for the purpose of entertainment. In other cases deviations from the laws of nature—a varying speed of light or an inverse-cube relationship for gravity—leads to an investigation of the consequences of such deviations for the physical universe. This is the case, for instance, when the strength of fundamental constants is varied in computer simulations with the result that even a slight change in the value of certain fundamental constants would make carbon-based life impossible.

In a similar way, in order to account for slight deviations of the planets from a strict $1/r^2$ relationship, as postulated in Newton's law of gravitation, alternative dependencies, like $1/r^{2+n}$ were investigated and subsequently found to be wanting (cf. Kragh 2007: 108–110). The fictional scenarios, in which the physical laws are broken, may depict *logically* possible worlds. Scientific thought experiments, by contrast, investigate *physically* possible worlds, and their consequences for 'our' world. For this reason:

> (t)he strength of possibility, physical or logical, depends on whether the thought experimenter has constrained herself to constructing only models were the actual physical laws obtain. (Cooper 2005: 338; cf. Giere 2004: §6)

A world with varying speeds of light and different values for the fundamental physical constants is physically possible but it is not the actual world. However, if a demon is employed to explore such possible worlds the investigation will provide us with important information about the actual world, since it helps us to determine what the necessary and sufficient conditions are for the actual physical world to function.

In which way do conceptual models of physically possible worlds represent? We characterized representation in terms of fit, and fit as the satisfaction of constraints. The constraint space still exists but the constraints take on different values for the sake of investigating the consequences of the imaginary scenarios. The model represents a physically possible world, which deviates from the real world by the degrees to which the constraints are altered for the purpose of modelling. A demon may come in handy at this stage. The modelling does not serve the purpose of creating a separate fictional world—a fantasy world of escape; the modelling serves the purpose of learning something about the actual world, the actual universe, by envisaging a physically possible world.

Chapter 5

The Function of Thought Experiments

> *It is because the mind, the weaver of illusion, is also the only guarantor of reality that reality is always to be sought at the base of illusion.*
>
> Eddington, *The Nature of the Physical World* (1932: 319)

The discussion so far has encountered a number of suggestions as to the function of thought experiments. Ernst Mach stressed the role of instinctive experience in thought experimentation, and placed thought experiments as the middle way between, in Francis Bacon's analogy, the bee—the accumulation of facts—and the spider—the flight of pure thought. Mach stresses that thought experiments in science must stay close to empirical facts (Mach 1883; Kühne 2005: Part II.2). Max Planck, too, emphasized the heuristic value of thought experiments: they help the scientist to formulate hypotheses; to gain new insights into lawful connections in nature, irrespective of whether they are subjected to empirical tests. Thought experiments are not bound by constraints of precision but their premises must not contain contradictions. Thought experiments require idealizations and abstractions, since their main function is the 'continuous variation of facts in thought' (Mach 1905, quoted in Kühne 2005: 198).

The historian of science T.S. Kuhn also held that thought experiments do more than remove contradictions (Kuhn 1964; cf. Cooper 2005; Humphrey 1993). They teach us something about the (in-) appropriateness of lexical terms, with which the natural world is described. Aristotle, for instance, failed to distinguish between 'average' and 'instantaneous' velocity. In particular, thought experiments may indicate how nature fails to correspond to the accepted system of expectations. That is, thought experiments become an important analytic tool during a crisis in science. During a crisis, the old paradigm becomes questionable because it cannot solve a problem, which is within its remit to master. If no solution is found the problem turns into an anomaly. An anomaly is not just a discrepancy between a theory's predictions and the measurement results. Such discrepancies are normal in science. An anomaly is a serious disagreement between the representation of the world, according to the accepted paradigm, and the empirical world, as revealed in observational and experimental results. Kuhn claims that the anomaly must already

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_5

have been present in the scientist's mind, if only dimly and unclearly. The function of thought experiments is to bring such vaguely perceived anomalies to light. During a crisis, when scientists are looking for solutions to an anomaly, a thought experiment may well contribute to a scientific revolution, i.e. a replacement of one paradigm by another.

However, as we have seen, many thought experiments do not render implicit anomalies explicit. Think of Aristotle's tower experiment or Einstein's elevator experiment: rather than making explicit an anomaly, the first confirms normal appearances and the second establishes a new, surprising connection between acceleration and gravitation. In Kuhn's characterization an anomaly is a serious disagreement between a paradigm and the world, but no thought experiment is needed to highlight the discrepancy. A thought experiment deals with hypothetical and counterfactual situations, to which, in Lichtenberg's words, 'nothing objective may correspond.' 'Thought games' may of course discover anomalies, as Galileo's thought experiment against the Aristotelians showed, but they cannot recover anomalies because they are not real experiments.

The function of a thought experiment cannot lie in a resuscitation of dimly perceived anomalies. What differentiates hypothetical or counterfactual 'what-if' questions à la Lichtenberg from thought experiments, conceived as conceptual models, in modern science? Lichtenberg poses 'what-if' questions but does not explore the conceptual consequences of 'what-if' scenarios. What is required to investigate 'what-if' questions?

One proposal is that beyond their hypothetical and counterfactual character—from which inferences about the natural world are drawn—thought experiments satisfy an additional requirement: they 'must stand in a privileged relationship both to past empirical observations and to some reasonably well-developed background theory' (Irvine 1991: 150; cf. Humphreys 1993: 220–221). By this additional criterion many of Lichtenberg's 'what-if' questions do not qualify as thought experiments because they do not relate to a background theory, which consists of empirical and non-empirical principles and values. Admittedly, by general consensus, thought experiments do not test the empirical adequacy of theories, since they do not produce new knowledge about the empirical world. But choice between rival theories is governed by non-empirical virtues as well: 'internal consistency, external coherence with other theories, simplicity, and explanatory power' (Bokulich 2001: §6). It has already been pointed out that one function of thought experiments is to investigate the empirical implications of scientific theories. However, following in the footsteps of Kuhn (1973), it may be said that 'a central function of thought experiments is to test and evaluate the internal consistency, external coherence, simplicity and explanatory power of our theories' (Bokulich 2001: 302).

> The function of a thought experiment is to draw out the physical implications of our theories and to test their non-empirical virtues. (Bokulich 2001: 303)

This thesis can be illustrated in the light of our previous considerations:

- Aristotle's tower argument seeks to draw out the physical consequences of a moving Earth. He finds them 'absurd' but only because he is not familiar with the notion of inertia. His tower experiment seems to underline the *external coherence* of the behaviour of the falling object with the geocentric worldview. It also seems to show the simplicity of the concept of a stationary Earth.
- Archytas examined the *internal consistency* of the Aristotelian-geocentric view of the finitude of the cosmos and found it wanting. He does not anticipate Poincaré's thought experiment, which shows that a surface, like a sphere, can both be bounded and unlimited. The findings of Archytas's 'demon' do not fit in with the Aristotelian cosmology and are therefore not coherent with geocentrism.
- Einstein's elevator experiment establishes the equivalence principle, which demonstrates the possibility of a simpler, more unified view of the equality of inertial and gravitational mass. It also demonstrates the *explanatory* power of the equivalence principle because it leads to the replacement of the notion of gravitation by that of the field, or the curvature of space-time.
- As the subsequent Parts and Chapters will show experiments employing demons, like Laplace's, Maxwell's, Loschmidt's and Landsberg's Demons, have *explanatory power* with respect to the laws of physics.

This observation spells out nicely the additional requirement, demanded by Irvine, and characterizes the thought experiments we have reviewed. For instance, both Archytas's spear experiment and Aristotle's tower experiment are genuine thought experiments, since they satisfy the central function, i.e. the testing of non-empirical values of theories against background knowledge.[1] If the business of thought experiments is essentially one of testing the implications and consequences of theories and theoretical models—albeit in a non-empirical fashion—we can finally ask what thought experiments tell us and do not tell us about the world.

---

[1] Inspired by his Critical Rationalism K. Popper (1959: Appendix XI; cf. Kühne 2005: 328–35) distinguished the critical and heuristic functions of thought experiment—which he welcomes—from the merely apologetic function, which he rejects. Thought experiments, like any other theory tests, must serve in attempts to refute theories. A thought experiment, for instance, may show that the internal consistency of a theory is violated. In the light of this criterion Popper would presumably accept Archytas's arrow experiment because it seems to show the logical inconsistency of the Aristotelian assumption of a finite universe. By the same token he would presumably reject Aristotle's tower experiment because it seeks to support the geocentric worldview and the postulate of a stationary central Earth. However, Aristotle does more than merely seek to examine the internal consistency of his theory of motion and its external coherence with the appearances; his thought experiment is intended to show that the notion of a moving Earth is internally inconsistent. That is, both Archytas and Aristotle use their thought experiments to investigate the non-empirical values of the respective theories. If such an investigation is the central function of thought experiments, then Popper's distinction between critical and apologetic thought experiments becomes redundant: Archytas's 'refutation' is at the same time a 'confirmation' of the logical consistency of the infinitude of the universe; Aristotle's 'confirmation' of geocentrism is at the same time a 'refutation' of early Greek heliocentric speculations (of Aristarchos of Samos). In other words, due to the essential indecisiveness of thought experiments, Popper's distinction is untenable.

Chapter 6
What Thought Experiments Tell Us and Don't Tell Us About the World

Metaphors are an essential part of thought—including scientific thought.

Ruse, *Darwin and Design* (2003: 284)

Do thought experiments produce genuine knowledge about the real world? Similar to Popper's critical function of thought experiments, Brown distinguishes destructive thought experiments, whose function is to highlight conceptual and logical problems within a particular theory, from constructive thought experiments, whose function is to establish a positive result (Brown 1991: Chap. 2; cf. Bunzl 1996). Recall that on the Platonic view of thought experiments, at least the so-called Platonic experiments are constructive in the sense of producing new knowledge, even though they are based on a priori reasoning. This constructive sense, however, invokes the paradox of thought experiments: what positive results could they possibly establish? Thought experiments cannot replace real experiments and thus cannot establish empirical claims. Their indefinite character means that they can be retold and refashioned in different guises. Nevertheless, if a thought experiments convicts an old belief system of, say, logical inconsistency, it may be said to contribute to 'knowledge' in Popper's sense of falsificationism. Galileo's thought experiment about falling objects is said to have 'destroyed' the Aristotelian theory of motion, although the Scholastics at the University of Paris had already questioned the coherence of the Aristotelian theory of motion. Falsification means that an old theory has been shown to be mistaken, on the strength of empirical evidence. In this sense, falsification constitutes progress. But falsification is a rule, according to which an empirical theory must be shown to be false through empirical means. However thought experiments test non-empirical values. Do they thereby contribute to 'knowledge'? To use the notion of 'knowledge' in this context is misleading. Thought experiments are ambivalent and vague and, as conceptual models, qualitative in character. As such they cannot add to the store of empirical or theoretical knowledge about the world.

But if a distinction is drawn between **understanding** and **knowledge** an important function can be attributed to thought experiments (Weinert 2004: §3.1). Thought experiments increase our understanding of the natural and social world,

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_6

and they can do so even if they are mistaken. On this account it is confusing to say that the 'new knowledge' which thought experiments presumably provide 'involves increased understanding of the conditions under which the model holds' (Humphreys 1993: 220).

Knowledge in the sciences is associated with a theory's well-confirmed empirical phenomena, the derivation of precise predictions from the theory, mathematical deductions of less fundamental from more fundamental laws. Knowledge, to use a helpful slogan, is 'justified true belief.' *Understanding*, by contrast, operates under fewer constraints. It is concerned with the interpretation of theories or theoretical models; it revolves around fundamental notions like causality, chance, determinism, energy, indeterminism, mass, motion, space and time, in an attempt to make sense of the world around us. Throughout his career, the German physicist Werner Heisenberg was much concerned with questions of understanding and interpretation in the physical sciences. From his early publications on quantum mechanics to his last essays on philosophy, he returned repeatedly to the concept of understanding in physics. Together with Albert Einstein, Max Planck and Max Born, he was one of the most philosophical physicists of his generation. It was common practice among many of the founding fathers of quantum mechanics and relativity theory to include philosophical discussions in their technical papers. For Heisenberg understanding in physics meant the reduction of the complexity of phenomena to a few basic and quite general concepts (Heisenberg 1973: 46). The possession of such concepts would allow the representation of the underlying unity in a great number of phenomena. But crucially, the discovery of new phenomena would require the revision of concepts, which had served well in the representation of old domains. Heisenberg, like Bohr, stressed that Einstein himself had introduced a revision into the concepts of space and time. The abandonment of concepts like causation, they held, was similarly a consequence of the new discoveries in quantum mechanics (Heisenberg 1928: 21–28; Heisenberg 1931: 40–47). In a more specific sense, then, Heisenberg held that understanding meant the ability to detach oneself from old concepts, when new domains of experience were under consideration.

> We have understood a group of phenomena when we have found the right concepts for describing these phenomena (…) it is always the simplicity of the concepts in comparison with the great wealth of complicated experimental material, which convinces of their correctness. Usually in a new field many very different experiments can be carried out; and if all these experiments allow a description by the same simple new concepts, these concepts will finally be accepted as the correct ones (Heisenberg 1969: 337; cf. Heisenberg 1967: 411–414, Weinert 2004: Part I).

For the Austrian physicist Erwin Schrödinger the task of understanding was intimately connected with human ability to construct conceptual models (Schrödinger 1928; Schrödinger 1947). Such mental constructions assign underlying *structures* to the observable phenomena. The complexity of the phenomena could be coherently ordered by a *Gestalt,* even though not all aspects of it were subject to observation and experimental testing. Although Schrödinger uses the term *Bild* (picture, image), the primary aim of the conceptual models does not seem to be direct visualization. Rather, some underlying order (*Gestalt)* is to be assigned to the observable phenomena, which renders them understandable. This underlying structure could be expressed in purely mathematical terms or as an analogy with some familiar structure. Perhaps some idealised configuration could represent the underlying order. Schrödinger's notion of conceptual model harbours a complexity, which goes beyond simple mechanical models. It includes almost all the model types, distinguished earlier.

Thought experiments, as conceptual models, contribute understanding without adding to the store of empirical knowledge. Thought experiments in science, as E. Mach and M. Planck already insisted, must not be completely detached from the world of empirical facts. Nor must they be limited to simple 'as-if' questions, even though they employ hypothetical and counterfactual reasoning.

> Only instances of hypothetical and counterfactual reasoning which have their parameters in large measure determined by a corpus of relevant observational and theoretical concerns will count as genuine thought experiments (Irvine 1993: 159).

For thought experiment to contribute or increase our understanding of the surrounding world they should satisfy some characteristics (Irvine 1993: 159):

- 'A thought experiment must be relevant to the testing of some hypothesis (…) which has arisen within a particular observational/theoretical context.' The testing, as we have observed, takes the form of testing the non-empirical virtues of theories.
- As should now be clear 'at least some features of the thought experiment must be grounded in the observable world if it is to have any relevancy to general scientific inquiry.' This criterion is even satisfied by Lichtenberg's aphorisms.
- 'Good thought experiments, like most good physical experiments, are repeatable.' Hence thought experiments should be construed as conceptual rather than mental models.
- 'It must be possible to identify a number of independent (or antecedent) variables within the thought experiment in order to determine correlations between variations of these variables and a further set of dependent variables used to characterize the experiment's outcome.' It should be observed, however, that these variables are often of a non-quantitative kind—as in Archytas's argument or Einstein's elevator argument—and where they take on numerical values, these values are not essential for the conclusion—as in the modern version of Aristotle's tower argument.
- 'The outcome of the thought experiment should have some repercussions for the original background theory.' The 'evidence for or against some general conclusion' often concerns the non-empirical values, like logical consistency or conceptual coherence.

Even where thought experiments satisfy these characteristics, their status as conceptual models gives them the freedom to be mistaken. Whether mistaken or not, thought experiments construct models of physically possible worlds. It is in these constructions that demons prove to be particularly useful.

Chapter 7
Enter the Demons

New views through old holes.
Lichtenberg, *Sudelbűcher* (F879)

Like many scientific models, thought experiments use abstraction and idealization to perform their functions in scientific reasoning. Furthermore, thought experiments use hypothetical reasoning and in this respect resemble ‘as-if’ models. With the exception of analogue models, scientific models attempt to represent systems in the natural and social world. In this respect, too, thought experiments resemble scientific models and therefore they must stay close to the coastline of empirical facts. If thought experiments are conceptual models, they are an important addition to the class of recognized models in the sciences. As observed, thought experiments introduce counterfactual scenarios, which, by definition, refer to (logically or physically) possible worlds. Demons enter the story because they help to show, which possible worlds can be inferred from the existing store of knowledge. The demons of science do not usually defy the laws of nature but they test the laws to their very limit. Although none of the thought experiments reviewed so far required the services of a demon, demons would often have been the right agents to perform the counterfactual feats, which thought experiments typically envisage.

- Although humans lack the physical possibility to explore the edge of space, Archytas’s space traveller—as a demon—suffers from no such limitations. What remains a mere logical possibility for humans, without violating the laws of nature—walking on water, flying unaided through the air—becomes a physical possibility for a demon. Archytas’s space traveller must be a demon.
- No human could ride on a beam of light or pull a rope attached to a lab suspended in empty space, but we have no difficulty in imagining that a demon could perform these tasks. For a demon they are physically achievable.

Demons are therefore useful agents to perform the superhuman deeds, which occur in imaginary experiments. ‘What-if’ questions, which invite the depiction of possible worlds, can therefore conveniently be associated with the work of demons.

- What if a Demon could travel to the edge of space?

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_7

- What if a Demon could ride on a beam of light?
- What if a Demon could be suspended in gravitation-free space?
- What if a Demon could travel through the interior of the Earth?

As the subsequent Parts and Chapters will discuss, some of the famous demons of science have been employed to explore the conceptual, logical and philosophical consequences of scientific theories. Demons are particularly useful in this respect because they can expand the space of possibilities. They help to explore what science can and cannot tell us about the world.

Before we proceed to discuss some of the famous demons—Laplace's Demon (Part II), Maxwell's Demon (Part III), Nietzsche's Demon (Part IV)—it will be useful to bring to mind some other demons, and their deeds, who have been explicitly invoked in the history of ideas. For the purpose of coherence only demons who stand at the crossroads of the physical sciences and philosophy will retain our attention in later Parts.

7.1 Freud's Demon

Consider, for instance, the question of our mental life. If a demon could investigate the firings of the neurons—if a demon could be sent, like a probe, into a person's brain—would this exploration help to discover whether humans are driven by unconscious motives? Sigmund Freud seems to have been inspired by Maxwell's Demon when he described a guardian, equipped with supernatural powers, who sits at the threshold between the conscious and unconscious compartments of the mind. This Demon sorts out mental processes as to whether they are fit for admission from the unconscious to the conscious part of the mind.

> Let us therefore compare the system of the unconscious to a large entrance hall, in which the mental impulses jostle one another like separate individuals. Adjoining this entrance hall there is a second, narrower room—a kind of drawing-room—in which consciousness, too, resides. But on the threshold between these two rooms a watchman performs his function: he examines the different mental impulses, acts as a censor, and will not admit them into the drawing-room if they displease him. (Freud, *Lectures* Vol. XVI, 1916–1917, Part III: 295)

Freud is aware that the spatial arrangement of the mental apparatus in this thought experiment may be misleading or even incorrect but he insists that it nevertheless serves the heuristic purpose of helping to make sense of the behavioural 'observations'.

> They are preliminary working hypotheses, like Ampère's manikin swimming in the electric current, and they are not to be despised in so far as they are of service in making our observations intelligible.

Clearly, then, Freud saw the function of thought experiments as increasing our understanding, in his case of mental life. Trained as a scientist, Freud was very well aware of methodological issues. He echoes Mach and Planck in insisting that thought experiments must stay close to the empirical world.

> I should like to ensure you that these crude hypotheses of the two rooms, the watchman at the threshold between them and consciousness as a spectator at the end of the second room, must nevertheless be very far-reaching approximations to the real facts. (Freud, Vol. XVI, 1916–1917, Part III: 296)

7.2 Descartes's Demon

The French philosopher and physicist René Descartes also employed a demon to aid understanding, not of human psychology, but of our mental abilities. Descartes's Demon is employed to drive his methodological doubt about knowledge of the external world to the ultimate limit. The Demon may deceive Descartes about the veracity of his thoughts of the external world. But the Cartesian knows that it is the Demon who deceives his mind. Hence, Descartes concludes: 'I think therefore I am' (*cogito ergo sum*).

It is important to emphasize that Descartes's Demon is part of the Cartesian solution to the then vexing problem of skepticism. The context of the Cartesian doubt is the widespread skepticism in intellectual circles, especially in France, in the 16th century. For skeptical writers like Michel de Montaigne neither our senses nor our minds guarantee the certainty of knowledge. The 'beginning and the end' of human knowledge lie in sensual experience but the five senses are not sufficient to perceive the 'essence' of things with certainty. (Montaigne, *Essais* II, XII: 575, 572)

> The uncertainty of our senses renders everything they produce uncertain. (Montaigne, *Essais* II, XII: 584; translated by the author)

The mind, by contrast, is a 'vagabond', which cannot be constrained by 'order' and 'measure'.

> It is an instrument made of lead and wax, stretchable and pliable, which fits every bias and all measures. (Montaigne, *Essais* II, XII: 548, 541; translated by the author)

Prior to the 17th century, when both empiricist and rationalist philosophers went in search of 'certain knowledge', the French essayist Michel Montaigne expressed the view that genuine knowledge of the natural world was unattainable. In the wake of the Cartesian solution Blaise Pascal, in his *Pensées* claimed that 'we have an idea of truth invincible to all scepticism.' (*Pensées* 395) But the skeptical Montaigne, in his *Essays* (1580), composed in the 16th century, held that people had to content themselves with mere opinions. Opinions, however, diverge. Adopting skepticism has the consequence of believing that there is no evidence, which could lower the credibility of one opinion and increase the credibility of another, on the strength of mutually agreed evidence. Skepticism towards knowledge, however, was not as unreasonable as it sounds today, especially since 17th century philosophy attempted to secure certain knowledge.

One of the most advanced natural sciences in the 16th century was astronomy. But even in astronomy different models vied for the attention of contemporaries well into the 17th century. Even when it looked as if the Copernican model was gaining predominance in the 17th century, G. Riccioli's textbook of astronomy (1651) still lists 5 competing models. To the skeptical philosophers at that time this plethora of planetary models must have re-emphasized their skeptical beliefs that different incompatible models seemed to make sense of the appearances. But no single planetary model could claim to be better than the others, at least not until Newton demonstrated how and why planets orbit the central sun in elliptical orbits. In the middle of the 16th century, Tycho Brahe proposed a compromise model, according to which the Earth was the centre of the universe, with the moon and the sun orbiting around it; but all other planets moved around the sun (Figure 4.1c). But there was, as of 1600, no decisive empirical evidence, which could rule in favour of one and against the other models.

Nicolaus Copernicus's great book *De Revolutionibus* (1543) itself bears witness to the skepticism of the age. Copernicus died on May 24, 1543 before his book could be published. It fell to the Lutheran theologian and preacher, Andreas Osiander, who was based in Nuremberg (Germany), to oversee the publication of *De Revolutionibus*. Not only did Osiander change the title of the Copernican treatise, he also added an anonymous, unauthorized Preface, in which he defended the book along skeptical lines. Kepler later identified Osiander as the author of the anonymous Preface. It is philosophically significant because Osiander tries to interpret *De Revolutionibus* as a treatise, which, contrary to first impressions, does not challenge the accepted geocentric worldview. In order to soften the conflict between the Church and heliocentrism, Osiander inserts his Preface in an attempt to present the Copernican hypotheses as mere calculating devices. They have the license to be false or replaceable as long as 'they reproduce exactly the phenomena of the motions.' (Osiander, Letter to Copernicus 20 April, 1541, quoted in Rosen 1984: 193–194) Reminding the reader of the newness of the heliocentric hypothesis, Osiander spells out the astronomer's dilemma. On the one hand the astronomer cannot know the 'true causes' of the celestial motions. On the other hand, the astronomer can establish fairly accurate descriptions of 'the history of the celestial movements.' How is this dilemma to be resolved? Osiander's recipe is the *locus classicus* of instrumentalist, skeptical philosophy. The astronomer can establish *how* the planets move but not *why*. Yet the human mind is exercised by theoretical curiosity. Even though no true explanation can be given, any explanation is better than no explanation. It is then the job of the astronomer, in the case of doubt,

> to think up or construct whatever causes or hypotheses he pleases such that, by the assumption of these causes, those same movements can be calculated from the principles of geometry for the past and for the future too.

It is therefore not necessary for the hypotheses to be true or even probable. Osiander holds that:

(…) it is enough that they [the hypotheses] provide a calculus, which fits the observations.

Why should the reader then even read the Copernican tract? Osiander makes an appeal to simplicity. Some hypotheses render the calculations simpler; make the observations easier to understand. They may even give rise to more reliable predictions.

> Therefore let us permit these new hypotheses to make a public appearance among old ones which are themselves no more probable, especially since they are wonderful and easy and bring with them a vast storehouse of learned observations. As far as hypotheses go, let no one expect anything in the way of certainty from astronomy, since astronomy can offer us nothing certain, lest, if anyone take as true that which has been constructed for another use, he go away from this discipline a bigger fool than when he came to it. Farewell. (Osiander 1543: 3–4)

In the face of doubt, he permitted heliocentrism as a mathematical hypothesis, but not as a realist claim about physical reality. By deflecting the Copernican hypothesis along instrumentalist lines, Osiander sought to remove its sting. It was another mathematical device, with no firmer grip on reality. It had as little probability as the established Greek hypotheses. The true causes of planetary motion cannot be known, because the human mind is too weak to apprehend the celestial sphere. In the absence of physical understanding, revelation takes its place. Osiander captures the essence of Montaigne's skepticism. Copernican heliocentrism did not romp to an easy victory. It faced hostility and doubt.

Descartes wanted to cut through the Gordian knot of skepticism. He employed a Demon to defeat it. The Demon is used as a methodological device. Its purpose is to take skepticism very seriously and turn it against itself. The Skeptics doubted in order to reduce all apparent knowledge to mere opinion. Descartes doubts in order to achieve certainty of knowledge. In his thought experiment Descartes allows the Demon to deceive him into believing that all his thoughts about the external world are mere illusions. The Demon's deception only underscores the skeptic's belief that our senses are untrustworthy. But even the Demon cannot avoid deceiving someone—in this case Descartes. Hence Descartes concludes that 'I am a thing that thinks, feels, doubts, desires and can be deceived.' One truth is immune to all doubt: I think, therefore I am. Perhaps my thoughts are mistaken; perhaps they are illusions, instilled in my mind by an evil Demon. But they are *my* mistaken thoughts and *my* illusions. I *am* the one the Demon is deceiving. The Demon cannot change the fact that they are *my* mental processes. Hence, Descartes concludes, he had a basis in thought, which lay the foundation to a more reliable form of knowledge than the skeptics would accept.

7.3 Mendel's Demon and Evolution

Maxwell's Demon, who will occupy us in later Chapters, was very influential in discussions in physics about the arrow of time. But Maxwell's Demon made his presence felt both in psychology and biology. He influenced Freud. He also inspired the evolutionary biologist Mark Ridley to dream up a biological demon—Mendel's

Demon—who, instead of working his magic on gas molecules, sets to work on genes. Whilst Maxwell's Demon has the superhuman power to control the paths of individual gas molecules between two gas chambers, Mendel's Demon 'sits in the parent and tosses a coin over each gene to decide whether it will be allowed into each offspring' (Ridley 2000: 271). Mendel's Demon has the ability to 'control the inheritance of genes' and thereby becomes the 'executive of gene justice in all complex life' (Ridley 2000: 230, 199). Mendel's Demon is a metaphor for a *biological* mechanism, which operates in genetic inheritance. Maxwell's Demon is a metaphor for a *physical* mechanism, which operates on the flow of gas molecules. But whilst many physicists consider the workings of Maxwell's Demon to be physically impossible, Mendel's Demon does important work in the evolution of complex life.

> Mendelian inheritance controls how genes are inherited in complex life. It combines sex, reproduction, and the probabilistic rather than certain inheritance of genes. Mendel himself was an Augustinian friar, and I like to imagine the chance mechanism as a rather monkish figure—Mendel's demon—who stands over each gene in a parent and decides whether it will be inherited in the next generation, and which other genes it will be passed on with.

Gregor Mendel and James Clerk Maxwell were contemporaries. Whist Mendel was an obscure Augustian monk labouring in Brünn (Brno, now in the Czech Republic), James Maxwell was a famous physicist, working in Aberdeen, London and Cambridge.

> Mendel published his ideas in 1866; Maxwell described his demon five years later. Maxwell's demon is a hypothetical demon. It stands by a hole between two parts of a vessel and, by allowing only the fast-moving molecules through in one direction, can make (without expenditure of work) one part of the vessel hot and the other part cold. Maxwell's demon is an anti-randomizing demon, who opposes the random movement of molecules and produces a more ordered state of the vessel—that is, it comes to have a hot half and a cold half rather than a uniform temperature throughout. Mendel's demon, by contrast, is a more realistic demon. It is a randomizing demon, who creates an ordered state (that is, complex life) by opposing the disruptive force of natural selection. (Ridley 2000: x)

More than an analogy exists between Maxwell's Demon and the evolution of life and the universe. If Maxwell's Demon were successful, a limitless supply of energy could be generated. According to Maxwell's set-up, his Demon expands no energy whilst sorting the fast from the slow molecules. Maxwell's Demon, if his work could be achieved, would create heat in one chamber simply by allowing only the fast molecules to pass through the opening. According to modern physics heat is not, as used to be assumed, some substance but average kinetic molecular energy. A Maxwellian Demon could be installed in every house, by every window and door, refusing entry to slow air molecules from outside and allowing entry to the fast molecules, which have been accelerated by the sun's rays or intermolecular collisions. The Demon's actions run counter to the universal increase of entropy or, metaphorically speaking, the increase of disorder, which accompanies all physical activities. However if there is a universal tendency for disorder to increase and order to decrease, why do we observe order all around, from the formation of galaxies to the evolution of complex life?

> In the sense of thermodynamics, the development of life is like swimming against the current of entropy. The spontaneous origin of order without external influence and the expenditure of energy would contradict the second law [of thermodynamics, i.e. the increase of disorder] and require 'demonic' forces. (Mainzer 1996: §3.4.2; cf. Penrose 2010: §2.2)

For life to be possible on Earth and to evolve towards complexity, and hence increase of order, energy must be drawn from the sun. The sun is a low-entropy source of energy, i.e. the energy, which arrives from the sun, as carried by photons, is much higher than the energy, which the Earth reflects back into space. This surplus of energy makes life possible on Earth. Overall, the total energy degrades or the total entropy of the universe increases but life on Earth benefits from the arrival of high-energy photons.

> Living systems are open systems, which compensate for their entropy production by constant exchange of energy with their environment. This metabolism of open systems solves the thermodynamic pseudo-problem of Maxwell's demon and the theory of evolution. (Mainzer 1996: §4.44)

Human beings constitute one of the highest forms of complexity and Darwin's theory of evolution was designed to explain, through the principle of natural selection, the evolution of life on Earth from its simplest to its most complex forms. For Darwin the principle of natural selection was required to explain not just the anatomy of human bodies but also the emergence of human minds. But the universe itself undergoes evolution, from a smooth beginning in the Big Bang, to its ultimate demise in a 'heat death' or a Big Chill.

Evolution, then, points to the topics, which later Chapters will explore. The emergence of complexity on Earth during the expansion of the universe seems to indicate a cosmic arrow of time but so does the Second law of thermodynamics. It states the principle of the universal increase of entropy (or disorder). Humans and living organisms in general, have temporal experience, as revealed in the distinction between past and future and the awareness of the 'flow' of time. Hence there are many arrows. But how are they related? Is there a master arrow of time? The complexity of the human mind includes what many believe to be free will. But if the universe is a deterministic machine, as Laplace's Demon envisaged it, does this Demon imply a static arrow-less world, in which human beings are deprived of free will? Maxwell's anti-randomizing Demon seems to show, by contrast, that the law of increasing entropy (disorder) may be probabilistic, not deterministic. Fluctuations in entropy could decrease the disorder. If the world is indeterministic can an arrow of time be derived from the increase in disorder? And how does the mind fit into an indeterministic world? Yet the increase in entropy seems to be such a universal phenomenon that it suggests itself as a criterion for the arrows of time. But where? Even if a local arrow of time is discovered on Earth, does the universe itself exhibit an arrow of time? Nietzsche's Demon claims the eternal recurrence of events. The history of the universe will repeat itself *ad infinitum*. Yet, this thesis is incoherent. Nietzsche's Demon is redundant. It requires Landsberg's Demon to gain a vista of the evolution of the universe or, more dramatically, of the multiverse, in which it is embedded.

A study of the demons of science will enable us to provide some answers to these questions. It will allow us to acquire an understanding of a cluster of fundamental notions—determinism, causality and free will; entropy, order and disorder; indeterminism, evolution and the nature of the mind; the anisotropy of time and the temporality of the world—which form a common toolbox, shared by science and philosophy. It will also allow us to determine what can and what cannot be said in the name of demons. Let the demons enter!

Part II
Laplace's Demon

> *The determinism of the physical laws simply reflects the determinism of the method of inference.*
>
> Eddington, *The Nature of the Physical World* (1932: 271)

Laplace's famous Demon evokes an image of the universe as a gigantic clockwork. A clockwork works with predictive accuracy. But contrary to appearances, it does not distinguish between past and future. If the clock ticks regularly and reliably it is possible to predict the future and past positions of the hour hand: it requires 3600 ticks—one tick per second—for the hour hand to advance by one unit (one hour). If the clock's mechanism were reversed it would require 3600 ticks to retrograde it by one hour. Barring minor irregularities a clock is therefore a deterministic system. The clock does not distinguish past and future.

Laplace's thought experiment attempts to show that the universe, too, is a deterministic system. It does not distinguish between past and future; it recognizes no arrow of time. To a superhuman mind, the future and past are equally present. From the present point of view, the future is as predictable as the past is retrodictable. The image of a clockwork universe translates into the view of a deterministic scientific theory. The French mathematician and physicist Pierre-Simon Laplace invoked the services of a superhuman being—a Demon—to investigate the properties of a deterministic scientific theory, like classical mechanics. To the Laplacean Demon, using the laws of classical physics, the whole universe appears like a long filmstrip, in which every frame, in the past, present and future, is already present. But if the Demon already knows what will happen in the future, that what we do today will determine what we shall do tomorrow, it seems that determinism implies a world without an arrow of time and free will. The past determines the present and the present determines the future. Past events cause future events. This part of the book will consider some of the philosophical aspects of determinism: What does it mean? Does a deterministic theory correspond to a deterministic world? Does determinism mean the same as causality? And does it deny an arrow of time and deprive humans of free will?

Chapter 8
Laplace's Demon: Causal and Predictive Determinism

> *If the whole prior state of the universe could occur again, it would again be followed by the present state.*
>
> Mill, *A System of Logic* (1843: Bk. III, Chap. VII, §1)

The clockwork universe is an image developed by the mechanistic worldview, which arose as a result of the Scientific Revolution (1543–1687). The mechanical worldview is a philosophical view, which claims that all events in the universe can ultimately be reduced to mechanical principles: all events can be explained in terms of matter and motion, under the rule of physical laws. Similar to a clockwork, all events, whether past, present or future must be completely determined. What does it mean that all events are completely determined? There are several versions of determinism.

The ability to determine from the present state of a physical system, in terms of specific parameters, its state at an earlier or later point in time, in conjunction with the knowledge of lawful regularities under which the system evolves, is one way of defining *determinism*. It is *predictive* determinism, since the emphasis is on an agent's ability to make unique predictions of future events from the present state of affairs. The predictor must have precise knowledge of the specific state of the system, at the current time, and the fundamental laws, according to which it will evolve to its future state.

But at times the Laplacean Demon stipulates further that the physical world is locked in a causal chain of events. Not only does appropriate knowledge lead to accurate predictions; the whole universe is bound in a unique concatenation of events. The Laplacean Demon, in this guise, embraces *causal* determinism.

Determinism—and its identification with causation—is a central feature of classical physics and its philosophy. It origins lie with the founding fathers of modern science and are epitomized by Laplace's Demon. In an often-quoted passage, Laplace derives determinism—the ability of a superhuman intelligence to predict future events—from the *axiom of a universal causal chain* of all events. This axiom, which Laplace adopts from Leibniz, is the Principle of Sufficient Reason. It states, in Laplace's words:

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_8

> The present events have a profound link with the preceding events, which is based on the obvious principle that a thing cannot begin its existence without a cause, which precedes it. (Laplace 1820: Introduction; Laplace 1774; cf. van Strien 2014; Koźnjak 2015)

This statement of the universal law of causation has had a respectable tradition in philosophy. Leibniz maintains:

> (t)hat everything is caused by a determined destiny is as certain as 3 × 3 = 9. For destiny consists of the interdependence of everything as in a chain, and will take place infallibly as much so before it has occurred, as when it has occurred. (Quoted in Mittelstaedt 1976: 134–135)

From this concatenation of events, Leibniz adds, with a demon-like figure in mind:

> (…) that everything proceeds mathematically – that is, infallibly – in the whole wide world, so that if someone could have a sufficient insight into the inner parts of things, and in addition had remembrance and intelligence enough to consider all the circumstances and to take them into account, he would be a prophet and would see the future in the present as in a mirror. (Quoted in Cassirer 1956: 12)

What do Leibniz and Laplace mean by the phrase 'the universal chain of events'? Laplace speaks of a 'profound link' between past, present and future events. Such a link could not be one of the invariable successions of events, like day and night or summer and winter. Although they follow each other invariably, the one is not the cause of the other. Rather, the existence of the effect has to be *conditionally* dependent on the prior existence of a cause. Under this specification, the invariable succession of day and night, of winter and summer, is not a case of causation. The appearance of the day, or the summer, is not conditionally dependent on the disappearance of the night, or the winter. Rather the succession of day and night, of winter and summer, are mere correlations, due to an underlying cause: the daily and annual rotation of the Earth. From this characterization of causation, the Laplacean principle of causal determinism in the whole universe does not yet follow. For it may well be that some cause always produces the same effect (under the same circumstances, the warmth of the sun melts ice), yet this local cause reveals little about the future state of the *whole* universe. To derive the predictability and retrodictability of future and past states of the whole universe from knowledge of the present state, Laplace must make a further assumption:

> We ought to regard the present state of the universe as the effect of its antecedent state and as the cause of the state that is to follow.

Laplace stipulates a Leibnizian chain of events between cause and effect. It is the association of determinism with a unique cause, preceding each subsequent event, which gives birth to Laplace's vision of a Demon. It is a causal, dynamic form of determinism. The Demon, in order to function, must know the present state of the universe and its laws.

> An intelligence knowing all the forces acting in nature at a given instant, as well as the momentary positions of all things in the universe, would be able to comprehend in one single formula the motions of the largest bodies as well as of the lightest atoms in the world, provided that its intellect were sufficiently powerful to subject all data to analysis; to it nothing would be uncertain, the future as well as the past would be present to its eyes. (Laplace 1820: Introduction, VI-VII; the translation is quoted from Nagel 1961: 281–282)

The Laplacean Demon delivered an extreme version of the mechanistic worldview, which itself had been developed by Descartes and Newton, by Galileo and Boyle. It gave rise to the *clockwork* image of the cosmos. Laplace was not alone in his adoption of the clockwork universe. Paul Henri Thiry d’Holbach, one of the leading figures of the French intelligentsia, presented the cosmos precisely as a network of interlocking causes and effects. ‘The universe’, he wrote, ‘reveals to us an immeasurable and uninterrupted chain of causes and effects’ (D’Holbach 1770: Chap. 1). Roger Joseph Boscovich—the author of *A Theory of Natural Philosophy* (1768) and Continental proponent of Newtonian mechanics—also invokes the services of a demon to depict a deterministic worldview.

> Now, if the law of forces were known, and the position, velocity and direction of all the points at any given instant, it would be possible for a mind of this type to foresee all the necessary subsequent motions and states, and to predict all the phenomena that necessarily followed from them. (Quoted in Barrow 2007: 63; cf. Koźnjak 2015)

A century later the French physiologist Claude Bernard (1865: 69, 87–89) still regards determinism as an absolute principle of science. By it he understands ‘the absolute and necessary relation between things’ in animate and inanimate matter. However, the British astronomer Arthur Eddington, who—as Parts III and IV will show—became much preoccupied with the notion of the arrow of time, applied greater caution. He converted from the philosophy of determinism to the philosophy of indeterminism, under the impact of the new scientific discoveries after 1850. Eddington’s Demon is closer to the abilities of human scientists than the omniscience of the Laplacean Demon. He is able to deduce general laws of nature, without being able to predict particular events or properties:

> An intelligence, unacquainted with our universe, but acquainted with the system of thought by which the human mind interprets to itself the content of its sensory experience, should be able to attain all the knowledge of physics that we have attained by experiment. He would not deduce the particular events and objects of our experience, but he would deduce the generalisations we have based on them. For example, he would infer the existence and properties of radium, but not the dimensions of the earth. (Eddington 1936: 327)

The *chain* metaphor, which served d’Holbach to emphasize the interrelatedness of Nature, reinforces the uniqueness of the chain of events. The identification of causation with determinism is only possible under this extra assumption that ‘the same cause always leads to the same effect’. If the universe is indeed a vast interlocking system of an ‘immeasurable and uninterrupted chain of causes and effects’, and if posterior events are uniquely determined by prior events, then Laplace’s Demon embraces causal determinism. Laplace, d’Holbach and Bernard took the interpretation of classical physics a step beyond the philosophical understanding of its founding fathers. Their demons see the world as an interlocking chain of events. The demons, which Leibniz and Eddington have in mind, are 'prophets' who can 'foresee' the future.

We can therefore distinguish several kinds of determinism: metaphysical (or ontological), predictive (or scientific) and causal determinism. Briefly, metaphysical determinism is the view that the universe (or a physical system) follows one unique trajectory from the past to the future, in which no branching occurs. If two worlds share the same laws and their conditions coincide at one point in time, then they will agree at all points in time. Predictive determinism is the view that the events in such deterministic worlds are predictable (and retrodictable). Causal determinism is the view that an ontologically determined world has a causal structure such that it forms a unique chain of cause and effect events.

In more detail, *metaphysical* and *predictive* determinism can be presented as a joint argument scheme, with two premises and a conclusion.

- **P_1**: The initial conditions of a physical system—like the solar system—are given in terms of a number of specific parameters characterising its motion (momentum and spatio-temporal location)
- **P_2**: The system will be governed by lawful regularities, especially as specified in Newton's mechanics and its extension to other domains of classical physics. Differential equations play a particularly important role in these computations.

∴ **C**: Under these two premises, the past and future trajectories of the system are uniquely ontologically specified, in the sense that their probability of occurrence is one, if other disturbances on the system can be excluded. If observers have knowledge of premises P_1 and P_2, they will be able to make fairly precise predictions about the past and future trajectory of a system, like the solar system.

Note that this characterisation makes no reference to any causal influences on the system's dynamic behaviour. The mathematical equations lay down the trajectory of the system (for instance a planet orbiting the sun), without specifying any particular cause for this determinate behaviour. The ancient Greeks demonstrated how this can be done. They made fairly precise predictions of the motion of the planets, without specifying any particular physical cause of the planetary circles. They assumed that what they believed to be the circular orbits of planets was their natural motion. Even when Copernicus changed the Greek geocentric vision to the modern heliocentric worldview, he did so without stipulating a physical cause for their motion. Newton's laws provided the classic mathematical explanation of why planets move in the elliptical orbits, which Kepler had found. But Newton's laws do not specify a physical cause—or mechanism—for planetary motion. Newton wrote down his famous inverse-square relationship between two gravitational bodies but declined to speculate how a gravitational force can act on a planet through empty space to keep it in orbit (cf. Weinert 2009: Part I).

Causal determinism is therefore the stronger view, adopted by Boscovich, Laplace and d'Holbach. It adds a further premise to the argument (see Bunge 1979: 4, Earman 1986: 4–6).

P$_3$: The unique trajectories of the systems are produced by a chain of interlocking causes and effects such that *successive* stages of the universe can always be construed as the unique effects of anterior states; in turn the current state becomes the unique cause of further posterior states, lying in the future. Future states are conditionally dependent on past states.

Both Boscovich's and Laplace's version of determinism, in fact, vacillate between *predictive* and *causal* determinism. From the causal concatenation of cosmic events, Laplace shifts the focus at times to their predictability. The Demon can represent both versions of determinism. It is the Laplacean Demon's *predictive* rather than his *causal* determinism, which has become the focus of many criticisms in the 20th century.

For instance Nernst (1922), von Mises (1930, 1931) and Frank (1932), objected to the extreme idealisation of a superhuman being, needed for the formulation of the Laplacean view. This superhuman mind needs to be able to determine the precise initial conditions and the exact form of the equations for a system of particles. This ability lies beyond human capacity. The force laws, applying to most phenomena in the material world, are far more complicated than those of celestial mechanics. Laplace admits that a human mind remains far removed from the attainment of such an ideal, although it could achieve certain approximations, as in a field like celestial mechanics. Popper (1982), however, interprets the Laplacean Demon as a super-gifted scientist, surpassing a human scientist only by degrees of predictive ability. Popper then tries to show that the calculating precision, required to exercise Laplacean predictability, cannot be achieved in principle. To characterize predictive determinism, Popper introduces the idea of predictors. Predictors are predicting devices, which have the ability to calculate the future and past states of the world from knowledge of present conditions and appropriate laws. Predictive determinism makes the calculating abilities of predicting machines (predictors) to acquire knowledge about the system's future evolution an essential feature of determinism. Insofar as the cognitive abilities of competent observers can be treated as approximations to the predictors, they also, like the predictors, should have the ability to predict all their future states. This ability requires the knowledge of boundary conditions and the differential equations, which govern the system. But predictors cannot calculate all their future states, because the information a predictor receives about its own state 'is liable to interfere strongly with that state and thereby to destroy the predictive value of the information' (Popper 1950: 189; Popper 1982).[1]

De Broglie (1941: 59–61) raised three difficulties for Laplace's predictive determinism: (**a**) the assumption of precise predictability is unrealistic, due to the

---

[1] Popper's argument is similar to his 'unplanned planning' argument against directed economies or societies: even if all future states are perfectly planned, contingencies will arise, which require new unexpected plans. Of course the new, modified plan, will also run into problems of unplanned planning. These new 'unplanned plans' will interfere strongly with the previous plans and destroy their predictive value (Popper 1957: Chap. 21).

universal interaction between all bodies in the universe; (**b**) all our observations and measurements are subject to error and (**c**) the atomic realm imposes inherent limits on the precision, with which the parameters of atoms could be known simultaneously Bohm (1957: 158–60). emphasised the inadequacy of Laplacean determinism on the ground of the effect of outside contingencies and chance fluctuations, as well as the emergence of qualitatively new causes, new laws and new contingencies in the infinity of time. As we shall see, the long-term evolution of the solar system is indeed subject to such fluctuations.

What these objections indicate is that humans fall short of the calculating abilities of the Laplacean Demon. Causal determinism is a far stronger assumption than predictive determinism. Mathematical prediction is possible even if, as in Greek astronomy and Newtonian mechanics, the predictor is ignorant about a causal chain. The world could even be highly predictable and computable, and yet it may be ontologically indeterministic. Consider, for instance, the case of radioactive decay, in which subatomic particles are transmuted into other particles through the emission of charged particles. No particular cause within the atom has ever been identified, which could be said to be responsible for the emission of, say, an alpha particle at a particular time. It seems that the emission is not causally determined. The event is *causa sui*, that is, it is not the effect of a (known) prior cause. Yet it is possible to calculate the probability of an emission event for an ensemble of particles. If radioactive decay is truly indeterministic, a demon's knowledge would be confined to statistical predictions. Prediction, then, can be of a deterministic kind (according to the Laplacean Demon) or of a probabilistic kind (according to a Maxwellian Demon, introduced in Part III).

The world could also be highly deterministic yet not be computationally predictable (Penrose 1989: 220, 278; Penrose 1994: 33). According to the theory of statistical mechanics, the paths of gas molecules in a confined container follow deterministic trajectories. At least this is what human scientists assume, even though they are incapable of predicting the individual trajectories. But to a demon —like Maxwell's Demon—the individual paths of each molecule in a gas cloud are both determined and predictable.

So what is the point of Laplace's Demon? The Demon presents a thought experiment of a physically possible world. It is a model of a completely mechanistic and deterministic world. The challenge is to find out whether this model depicts the actual world. (Recall that thought experiments probe the extent and limits of our knowledge of the universe.)

While these objections show in various ways the limits of Laplace's conception of what came to be seen as *predictive* determinism, there is a further objection, which goes to the heart of Laplace's causal determinism. As we have seen above, Laplace accepted Leibniz's Principle of Sufficient Reason: 'Every event must have a cause.' The whole anterior state of the universe causes the subsequent state of the universe to move along *one* unique trajectory, at the exclusion of all others. This makes sense in a clockwork universe: the hands of the clock move steadily forward from one position to the next in either direction, depending on the accuracy of the mechanism. If this premise is denied, then Laplace's assumption of the unique causal concatenation of all events breaks down. There may be limits to determinism even in the classical realm. Randomness or at least indeterminism may lie at the root of things. If the evidence does not warrant the statement that every event must have a unique cause or at least that every cause leads to a unique effect, then one of the premises of Laplace's argument is denied. The derivation of Laplacean causal determinism must fail. The failure of causal determinism would have an impact on predictive determinism.

Nevertheless the association of determinism with predictability has proved useful to the mechanistic worldview. Predictive determinism, even in classical physics, need not require the predictability and retrodictability of the future and past states of the *whole* universe. Such foresight may remain the prerogative of a Laplacean Demon. Natural systems, like the solar system, can be modelled in isolation from the rest of the universe to derive a limited knowledge of their future or past spatio-temporal states. Even then it only describes an idealisation. The boundary conditions cannot be known to an infinite degree of perfection, and the universal laws themselves relate parameters in degrees of idealisation and abstraction. The idealised picture for a human, if not the Laplacean, mind is then: from the knowledge of boundary conditions and the governance of universal laws it should be possible to make limited predictions about the future spatio-temporal location of the system under consideration. Equally, it should be possible to retrodict past spatio-temporal states. Heisenberg (1973: 52; cf. Earman 2004) describes this predictive determinism as the principle of Newtonian physics (Fig. 8.1).

Astronomy provides paradigm cases of both prediction and retrodiction. In 1928, Arthur Eddington predicted 'a total eclipse of the sun visible in Cornwall (…) for 11 August 1999' (Eddington 1928: 299). On August 2, 1977 Voyager 2 was launched and flew past its target—the planet Neptune—on August 25, 1989 at a distance of less than 5000 km. Similar to Eddington's solar eclipse, this mission required precise predictions of the orbits of Neptune and Voyager 2.

On the night of September 23, 1846 the German astronomer Johann Galle in Berlin discovered the planet Neptune (cf. Peterson 1993: 110). But calculations of Neptune's orbit show that Galileo Galilei must have perceived it 234 years earlier,

Initial Conditions & Universal Laws →

Predicting the State of the System in the near Future

Initial Conditions & Universal Laws →

Retrodicting the State of the System in the near Past

Fig. 8.1 Heisenberg's description of predictive determinism since Galileo made entries in his diaries regarding an unidentified object. There is reason to believe that Galileo actually saw Neptune because the planet would have been at these locations in the sky on December 27, 1612 and again on January 28, 1613 (Zeilik 1988: 219). The same is true of the planet Uranus, which was discovered by Wilhelm Herschel on March 13, 1781. Retrodictions of the planet's orbit allowed astronomers to determine that there had been 22 sightings of the planet prior to Herschel's discovery (Peterson 1993: 103).

Chapter 9
Causality, Determinism and the Block Universe

> *As far as mechanics is concerned, we could also remember events in the future.*
>
> Hemmo/Shenker, *The Road to Maxwell's Demon* (2012: Chap. 10.6)

One lesson we can draw from Laplace's Demon is that notions like determinism and causality should be kept apart. The Greeks made fairly accurate predictions of planetary motions without having any knowledge of their dynamical cause. Newton's laws equally allow the prediction and retrodiction of planetary orbits, even though his notion of gravitation implies a dubious action-at-a-distance. By contrast, one may have causal knowledge of events, without the benefit of precise predictions. Evolutionary biologists, for instance, are able to causally explain the splitting of lineages in the past, without being able to predict when lineages will split in the future. These distinctions should be reflected in a model of causality to be developed later (see Part III, Chap. 5.2.1).

In some cases, determinism and causality merge into the notion of causal determinism. But causal determinism is a metaphysical assumption. A Laplacean Demon could endorse predictive determinism, whilst dispensing with its causal, dynamic component. The Demon could focus on the succession of states, even though there either is no causal link between them or the Demon wishes to ignore it. Metaphysical determinism—as a unique chain of events—could be compared to the pre-existing frames in a long filmstrip. The Demon's gaze may follow this chain in either direction, as if it were a temporal highway. In each direction the traveller encounters a unique set of events. From any moment in this chain it is possible to proceed to future events or to retrodict past happenings even though their causal connection, if any, may be unkown. At any moment in time the laws fix the future and past trajectories of events, captured by one of the frames of the filmstrip.

As mentioned above, Laplace's version of determinism can be construed, alternatively, as predictive *or* causal determinism. Laplace's famous Demon may serve as an example of *predictive determinism*, since he appeals to a superhuman intelligence to which neither past nor future states are uncertain. As Boscovich's Demon also makes clear, *predictive* determinism refers to the cognitive abilities of

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_9

competent observers to predict the past and future states of the world, from the present state. Such competent observers must possess sufficient knowledge of the initial conditions and the universal laws, which govern them. The Demon can predict successive states of the universe—or the world-lines of particles in what will soon be characterized as space-time—because the Demon, and his human apprentices, know the dynamic laws and the initial conditions of the system, whose events are to be predicted. The Demon knows the data to much more perfection than human scientists. But the Demon departs from his human counterparts if he not only predicts events but is able to survey the *whole* universe. The whole universe looks to him *as if* the events already had left their traces on the canvass of space-time, like frames in a long film strip. But if for the Laplacean Demon all events have already happened, irrespective of their location on the axis of time, then the Demon endorses a deterministic view of the universe. For the Laplacean Demon the whole world history is engraved in the fabric of space-time. He can see all the individual frames in one fell swoop. But humans, due to their cognitive limitations, must experience them sequentially. This sequential flow of the images creates the illusion of the passage of time. But since Laplace bases his superhuman intelligence on the assumption that 'we ought to regard the present state of the universe as the effect of its antecedent state and as the cause of the state that is to follow', he goes beyond predictive determinism and assumes a stronger *causal version of determinism*. The past, present and future states of world history are fixed but they are also locked in a causal chain, since any present state will always be the causal antecedent of a subsequent state, just as any past state was the antecedent to the present state. This assumption allows Laplace to identify causation and determinism to turn the Demon's position into one of causal determinism. The Demon's vision poses some puzzling questions for his human apprentices. Does he see the whole universe as if frozen in time, like a timeless filmstrip? Or does he witness an evolution of the universe from its beginning to its end? Does he observe a static or dynamic universe? Does he detect an arrow of time?

Such questions suggest that metaphysical determinism comes in two flavours, a static and a dynamic form (see also Ĉapek1951; Zeh 1992: Chap. 1):

(a) A *static* form of determinism. For the Laplacean Demon the world is like a map on which the coordinates of all events are already entered. There is no distinction between past, present and future. Therefore, past, present and future are equally real. This is an extreme form of determinism, since both the emergence of novelty and the passage of time are denied. It is a forerunner of the conception of the **Static Block Universe**, which is surprisingly popular with many scientists. For the omniscient Laplacean Demon time does not flow since all the frames are already in place and visible to the Demon's inquisitive eyes.

(b) A *dynamic* form of determinism. The evolution of the universe is 'predestined' forever from past to future, because deterministic laws will allow only one unique trajectory, given the initial conditions of any anterior state. Nevertheless, the world dynamically evolves from state to state. Hence this

> version of determinism allows for a dynamic evolution of the universe. It is equivalent to the view that world-lines in space-time possess evolving histories. This view is sometimes known as the **Evolving Block Universe**. In such a dynamic universe time would flow even for the Demon because new frames are added to the film, as the universe unfolds from its beginning in the Big Bang to its ultimate end in a Big Chill. This dynamic form of determinism raises questions of becoming.

Both versions—the static and the evolving block universe—have been associated with Einstein's Special theory of relativity (Weinert 2004). From the static Block Universe it seemingly is but a small step to the assumption of fatalism and the denial of free will. The Special theory has also been associated with such claims. The question therefore arises whether classical mechanics and the Special theory do support such wide-ranging inferences. (Part II, Chap. 11)

The various demons, who have crossed our path, are right that notions like causality and determinism, predictive and causal determinism are related but they should nevertheless be kept apart.

The notion of (predictive) determinism requires two essential ingredients: knowledge of initial data (or boundary conditions) and the fundamental laws from which the trajectories are computed. The initial data may change from time to time but the fundamental laws remain the same. They remain invariant over time. What is more, the fundamental laws of motion make no distinction between past and future; they indicate no arrow of time. They are said to be time-reversal invariant. Such laws seem to strengthen the case for determinism, of the static flavour, and the Block Universe.

Chapter 10
The Time-Reversal Invariance of Fundamental Laws

> *So if the laws of Nature are indifferent as to the doing and undoing of an event, they must be indifferent as to a direction of time from past to future.*
>
> Eddington, *The Nature of the Physical World* (1932: 66)

The time-reversal invariance of fundamental laws is one of the basic arguments not only for determinism, but also in favour of the afore-mentioned Block universe. The fundamental laws make no distinction between past, present and future and hence do not give rise to an arrow of time. It is an essential feature of many fundamental laws that they are time-reversal invariant. Technically, this means that a physical law, say *L*, which includes a parameter *t*, allows physically possible models with either $+t$ and $-t$ as temporal parameters. Consider a ball, which moves inertially (i.e. without acceleration) in the positive x-direction along a trajectory, *h* (cf. Fig. 11.7). Let the ball obey the equation $x = vt$. The state of the ball is described by its initial momentum, m_i, and its position, p_i and evolves, under the dynamic equation, to its final state (m_f, p_f). Reversal means that the sign of the momentum, *m*, and the direction of time, *t*, are reversed. The instantaneous state is reversed by a reversal operator, *R*: $R(m, p) = (-m, p)$. Time reversal is achieved by the time operator, *T*, which act on *R* and the trajectory, *h*: $Th\{Rm(-t), p(-t)\}$, where *t* lies between t_f and t_i. The equation has the same form in both temporal directions (see Frigg 2008: 181; Uffink 2001: 314).

This characterization marks the time-reversal invariance of *laws*. The time-reversal invariance of laws means that the equations of physics admit—as physically possible—worlds, which are either generated by a positive time parameter (+*t*) or by a negative time parameter ($-t$). For an equation to be time-reversal invariant the replacement of the parameter +*t* by the parameter $-t$ must yield physically admissible models. That is, every process that is allowed to happen in one temporal direction (+*t*) is also allowed to happen in the opposite temporal direction ($-t$). The initial conditions of the time-reversed process become the time-reversed final conditions of the initial process. This process is satisfied in cosmological models, which allow the universe to expand to a maximum point and then to re-collapse, under specified conditions, to its original state. In such so-called

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_10

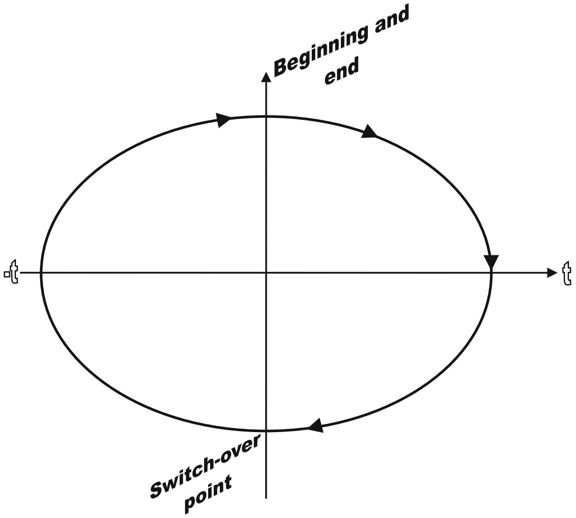


Fig. 10.1 A closed Gold universe, according to which the universe, after the beginning, expands to a maximum point and switches over to a retraction phase. According to the Gold universe model at this point the arrow of time would reverse so that the final condition becomes the time-reversed initial condition

Gold universe models the arrow of time is supposed to reverse at the moment of retraction—the switch-over point (Fig. 10.1).

But the dynamical behaviour of physical systems is dependent on both their initial conditions and their dynamic laws. It is possible for the fundamental equations to be time-reversal invariant, yet for systems they govern to be either temporally symmetric or asymmetric. This second sense of time-reversal invariance refers to the *solutions* of the equations. The equations may allow natural processes, which as a matter of fact do not occur in nature.

A good illustration of time reversal invariance or temporal symmetry in this sense is to picture a film of a pendulum—an ideal pendulum, which does not suffer damping—that oscillates with a certain frequency, v. A viewer would not be able to tell whether the film was running backward or forward in time. The same is true of an animation of planetary motion. As a matter of fact, it is the case in our solar system that all planets orbit the sun from west to east. But an observer who was ignorant of this fact would not be able to tell which of two films—one showing the planets moving in the familiar direction and the other in the opposite direction—was a correct representation of reality. Even if observers did know the true motion of the planets around the sun, on seeing the film running in reverse, they would only be able to conclude that the planetary system shown was not the familiar solar system but that it was a physically possible planetary system elsewhere in the universe. The particular direction of orbits depends on special initial conditions but the planetary laws have no preference for either direction. A time-asymmetric picture would result if the viewer was shown a slow-motion film of a bullet leaving a pistol shaft or a vase falling from the top of a table and breaking on impact. In these cases, if the film were shown running in reverse, the viewer would be in no doubt as to the true sequence of events. S/he would judge the reverse scenario to be physically impossible (especially if s/he was asked to judge intuitively without employing any previous knowledge of the laws of physics). The reason for this verdict is that these cases display a familiar temporal asymmetry from ordered smooth conditions to disordered final conditions. If a film was shown of the pieces of a vase, gathering together on the floor and returning to the top of the table, it would be very natural to reject the suggestion that the backward-running film was showing a physically possible process. Yet the laws of physics do not forbid a vase picking itself up from the floor and reconstituting itself on the top of the table. It is simply very unlikely to happen in the lifetime of the universe. It is an atypical case; a case in which the law is time-reversal invariant but the solution is temporally asymmetric.

A process is time-reversal invariant if a viewer cannot decide—as in the case of orbiting planets—whether a film of the events was running in the forward or reverse direction of time. Yet the example of the film of a falling vase shows the importance of de facto irreversible processes in the past-future asymmetry. De facto irreversible processes are to be understood as complex processes whose time-reverse is highly improbable in the history of the entire universe, although they remain theoretically possible. If they occurred, they would not violate the fundamental laws of the micro-processes in terms in which the macro-processes are to be understood. The fact that the reversibility of physical processes, if it occurred, is highly atypical but does not violate the fundamental laws of physics constitutes the de facto irreversibility of many physical processes.

So a contrast exists between the familiar world of asymmetric processes and the Laplacean world. The familiar world seems to exhibit arrows of time. The Laplacean world is governed by *t*-symmetric deterministic laws, which make no distinction between past and future. For the Demon the whole history of the universe exists all at once. Does the Demon's global vision mean that determinism is incompatible with the arrow of time?

Chapter 11
Determinism and Its Implications

> *It is impossible to trap modern physics into predicting anything with perfect determinism because it deals with probabilities from the outset.*
>
> Eddington, *New Pathways in Science* (1935: 105)

Consider again Laplace's Demon and his belief in metaphyscial determinism. In one fell swoop the Demon is able to survey the history of all particles—small and large—as trajectories in space-time. The Demon therefore 'sees' the whole vista of the universe, in its minutest detail, laid out like a map of events. For the Demon the future would not be as uncertain as it is for his human counterparts. If you were destined to win the lottery in the future, it would be certain for the Demon but still uncertain for you. Hence for the Demon past, present and future would be equally real. From the point of view of the Demon it looks as though time did not pass, as if there were no passage of future events into past events through the agency of the present. For a human observer the past can no longer be changed but the future seems to be still open. Humans feel that the present is the stage on which, thanks to the agency of free will, the branching alternatives of the future can be considered before a decision is taken in favour of one particular option. But for the Demon no choice presents itself: free will and free agency, the rational choice of one option from a pool of equally available alternatives is a human illusion. What will be, will be: for the Demon the future histories are already etched onto the canvass of space-time. In terms of our earlier distinction the Demon inhabits a static Block Universe.

11.1 Determinism and the Arrow of Time

Could the Demon nevertheless experience an arrow of time? In order to approach this question, let us briefly reconsider the Greek view of the universe. It was a geocentric universe: the Earth rested motionless at the 'centre' of the cosmos, the limits of which extended no further than 'our' solar system. The Earth possessed neither a daily nor an annual rotation. The six known planets of antiquity, including

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_11

the sun, and the ‘fixed’ stars rotated in a circular motion around the central Earth. For Plato the planets were the ‘instruments of time’. In the geocentric worldview the regular and circular motion of the planets around the ‘centre’ presented perfect models of the passage of time on the basis of which humans could construct clocks and calendars. Since these ancient times sun dials and other measuring devices—like water clocks—have allowed humans to ‘read’ the local *flow* of time. But the Greeks would not have acknowledged a cosmic *arrow* of time because according to their geocentric view, the universe (beyond the moon) was eternal, symmetric and unchanging. It was a static, not an expanding universe.

What about the Demon? Is a deterministic world a world without time? According to modern cosmology the universe is undergoing an accelerated expansion from the Big Bang 13.7 billion years ago. Its ultimate fate is likely to be a Big Chill—eons of time from now. The Big Chill is the dissipation of all energy differences, making life impossible. The Demon, of course, would see the whole trajectory of the universe from its birth in the Big Bang, to its ultimate demise in the Big Chill (or the Heat Death, as the 19th century called the event). The omniscient Demon would know how long it would take the universe to wind down to a lifeless wasteland of aimlessly swirling particles. The Demon would ‘see’ the whole history of the universe from its fiery beginning through the formation of galaxies and their accelerated expansion, the creation of order, from the emergence of life to its eventual extinction and the slow decline into thermodynamic equilibrium, as a continuous series of changes. But if the anisotropy of time—either on a local or global scale—is to be inferred from the observable changes that take place in the universe, the difference between the Demon and ordinary humans is merely a difference of perspectives. For the Demon the beginning and the end of the universe are already visible as a series of frames on a long filmstrip. For humans the arrow and the passage of time still unfold since the film must run its course. The end of time or of the universe remains a conjecture. For the Demon the succession of events already lies before him, like an evolving landscape: either the events in a local environment or on a cosmic scale. In either case, since time is fundamentally related to material change, time unfolds. Hence it is not necessarily the case that a deterministic world is a timeless world, especially if Laplace’s Demon believes in causal determinism. Even a deterministic world undergoes dynamic evolution, although the evolution happens along a unique trajectory since, in Laplace’s words, the current state of the world becomes the cause of the future state. Although the Demon presumably ‘sees’ every frame, this does not imply that there is no arrow of time. Hence causal determinism constitutes much more of a threat to the notion of free will than it does to the arrow of time.

The Laplacean Demon could be a proponent of the afore-mentioned Evolving Block Universe, according to which the universe evolves from a beginning in the Big Bang and then branches out into an uncertain future. If the Demon can peer into the distant future, to the end of time, he is bound to register a series of material changes, hence a cosmic arrow of time. But is the recognition of a cosmic arrow not undermined by the afore-mentioned time-invariance of the fundamental laws? In order to answer this question it is useful to recall the difference between the *t*-invariance of laws and the temporal symmetry or asymmetry of their solutions. For the moment, it may be observed that even though the fundamental laws of science are time-reversal invariant (unchanged with respect to temporal direction) the initial conditions, on which they operate, undergo dynamic change and this material change is essential for our conception of time. Determinism, as will be discussed shortly, has its limits. We shall also have to consider the work of Loschmidt's Demon (Part III, Chap. 16). It concerns questions of reversibility and irreversibility of the trajectory of systems, which operate under the *t*-invariance of laws.

11.2 Determinism and Fatalism

Determinism should not be confused with fatalism. According to Laplacean determinism, the future trajectory of a system is fixed, *if* initial conditions and deterministic laws are given. As Laplace's Demon stipulates, the current state of the universe becomes the cause of the subsequent state, just as the present state was caused by an anterior state. Since the choice of future states is conditionally dependent on a previous state, the future state is curtailed to one trajectory. Despite first impressions, this conception of causal determinism allows that a change in initial conditions, on the basis of the same deterministic laws, will give rise to an alternative future. Will Aristotle's famous sea battle take place tomorrow? It is bound to take place if none of the adversaries back down and the logic of warfare is allowed to unfold. There is no change in initial conditions, no change in the rules of warfare. What, however, if an emissary arrives to threaten the belligerent generals with crippling sanctions if the sea battle goes ahead? The generals may back down since the effect of the sanctions may be much worse than a victory in sea battle. A change in initial conditions will lead to a different history. Even the assumption of deterministic laws does not exclude the occurrence of alternative events, as long as the initial conditions are allowed to vary. Seen from the perspective of changing initial conditions, the future is still open.

But where would such a change in initial conditions come from, if the system is the entire universe? Recall that the Laplacean Demon has foresight of what is going to happen but he cannot interfere. If two galaxies are on a collision course—as apparently the Milky Way and the Andromeda galaxy are—such a deterministic trajectory would lead to a change in initial conditions compared to the trajectories of non-colliding galaxies. In this way the Demon could distinguish actual from possible histories.

Fatalism embodies a much stronger claim than determinism. It leads to both the denial of an open future and free will. In the common-sense understanding, fatalism is the view that all events are pre-determined and cannot be changed. They cannot be changed because they are governed by some force, be it logical laws, metaphysical necessity or supernatural agency. Inevitability governs the occurrence of events. Let us say that fate is kind to you and will make you the sole winner of next Saturday's lottery jackpot. Would failure on your part to buy a ticket not thwart fate's kind intentions? To think so would be to confuse determinism and fatalism. According to fatalism all events are already fixed in advance of any action on the part of an agent. No amount of human intervention can change them. Fatalism extends the unchangeability of past events to the unchangeability of all events. So even if you fail to buy a ticket your lottery win is guaranteed. Perhaps you will find a winning ticket in the street or it is given to you as a gift. What will be, will be!

In philosophy fatalism is cast as logical fatalism, following Aristotle's discussion of tomorrow's sea battle (see Bardon 2013: 138–152; O'Connor 1971: Chap. 12; Bunge 1959: §4.3). There is a set of propositions about the future, which are either true or false, and predict what will happen. (Either the sea battle will take place or it will not take place.) Let us say that the sea battle will take place, then the statement about this event is true, not only tomorrow but at all times. There is no open branching future, which could depend on human decisions. Logically speaking, fatalism amounts to the denial of indeterminate truth values of future events: they may be true or false.

> The case for fatalism goes something like this. What was true in the past logically determines what will be true in the future, therefore since the past is over and done with and beyond our control, the future must also be beyond our control, consequently there is no point in worrying, planning, and taking pains to influence what will happen. (Horwich 1987: Chap. II.4, 28)

The Laplacean Demon is a determinist, not a fatalist. For the Laplacean Demon the current state of affairs is the cause of the subsequent state of affairs, given the laws of nature. The commitment does not imply that a change in the current configuration of conditions, through some kind of intervention, will not lead to an alternative state of affairs in the future. What Laplace's determinism excludes is that a subsequent state of affairs remains undetermined—or leaves open several possibilities—given the current cluster of conditions and the laws of nature. Laplace's Demon is neither an indeterminist nor a fatalist. But Laplace and his determinism belong to the realm of classical physics. Has modern physics not besmirched the good name of determinism? Or has determinism found refuge in Einstein's Special theory of relativity, as has been claimed? The question is pertinent because Einstein's theory is an extension of the realm of classical physics.

11.2.1 The Special Theory of Relativity (1905)

Two claims are found in the literature: (1) It is frequently asserted that the Special theory of relativity is a deterministic theory (Sect. 11.2.2); sometimes the claim is added (2) that the Special theory implies fatalism (Sect. 11.2.3).

It will be convenient, in order to investigate these claims, to provide a preliminary characterization of the Special theory of relativity. Although it is often characterized as a revolutionary theory, it is really an extension of classical mechanics. Seeing it as a further development of classical mechanics does not belittle Einstein's achievement, since he himself did not claim the crown of a revolutionary theory for the Special theory. As we shall see in a later chapter, the Special theory had radical and counterintuitive consequences for the notion of time. As a first approach, note that the Special theory is based on two postulates:

1. According to the **Principle of Relativity**, all coordinate systems are to be regarded as physically equivalent. This principle already existed in classical mechanics, under the name of Galilean relativity principle, but it was restricted to mechanical phenomena. The reader will be familiar with this principle since all our normal activities—chatting, drinking, eating, reading, writing—happen under the watchful eye of the relativity principle. These activities are carried out in the same way irrespective of whether they happen on a park bench or a moving train. Of course a moving train makes noise and creates wind effects. Such side effects have to be ignored. From the point of view of physics and the physical laws, which underlie these activities, there is no difference between a system being in a state of 'rest' or in 'uniform' motion. (Non-uniform motions, like accelerations or rotations, do reveal the difference between being at rest and being accelerated.) In other words the laws of motion are the same in all inertial reference systems. In a famous thought experiment Galileo illustrated the equivalence of inertial systems. He imagined that in a cabin below the deck of a large boat a sailor observes the flight of insects and the behaviour of 'fish in a bowl'. From the behaviour of these creatures the sailor will not be able to determine whether the boat is at rest or in inertial motion (Galileo 1954: 199–201). Einstein extended the Galilean relativity principle from mechanical to all phenomena, including electromagnetic phenomena.
2. According to the **Principle of the Invariance of the Speed of Light** (in vacuum), light travels at a constant velocity of approximately $3 \times 10^8\,\mathrm{m/s}$, irrespective of either the motion of the source or the direction of the light beam. This latter condition is important as can be seen from a thought experiment. Imagine an exploding fire cracker on top of a moving train, watched by an observer on the embankment (Fig. 11.1). The explosion on top of the train will scatter light

Fig. 11.1 Invariance of the speed of light in the Special theory of relativity

> beams in all directions. How fast does the light signal travel (a) for an observer on top of the train who measures the direction of the light beam in the direction of the moving train and then in the opposite direction and (b) for the observer on the embankment? According to the Special theory, the answer is that they measure the same velocity (≈300.000 km/s), irrespective of the velocity of the train and the direction of the light beam. This situation is strikingly different from a 'classical' situation as can be seen from a slight modification of the thought experiment. Instead of an exploding fire cracker, an observer stands on top of the train and fires a bullet first in the direction of the moving train and then in the opposite direction. What is the velocity of the bullet? The train-bound 'shooter' and the observer on the embankment will disagree on the velocity of the bullet. The embankment observer will need to add the velocity of the train—say 30 m/s—and the velocity of the bullet—say 800 m/s—if the bullet is fired in the direction of the moving train (800 m/s + 30 m/s = 830 m/s). If the bullet is fired in the opposite direction the observer will calculate that the velocity of the bullet was 800 m/s − 30 m/s = 770 m/s. For the train-bound 'shooter' however, the velocity of the bullet is the same in both directions, i.e. 800 m/s, since he is already part of the moving train system.

The velocity of light, *c*, has been measured with increasing accuracy since 1676 when the Danish physicist Olaf Rømer determined its value by observations of the Jupiter moons; he found a value of $c \approx 225.000$ km/s (or 2.25×10^8 m/s), which is relatively close to the modern value of 2.99×10^8 m/s in vacuum. Einstein turned this empirical value into a postulate of his theory. To make *c* a postulate means that it becomes a limit velocity. According to Einstein's theory no particle can travel faster than the speed of light. Light particles, like photons, do travel at the speed of *c*, but material particles with mass, *m*, must travel below this limit velocity. The light postulate has a significant impact on the way events are visualized in coordinate systems, which move relative to each other.

What are coordinate (or reference) systems? As physical systems are either at rest or in motion, the first task is to adopt coordinate axes, which provide the units of spatial and temporal lengths. If a system is at rest it stays in the same location over a certain amount of time (Fig. 11.2a). But if a system is in motion, with respect to the first, then it experiences displacement in space over a period of time, which is either linear (indicating a constant speed, as in Fig. 11.2b), or exponential (indicating an acceleration, as in Fig. 11.2c). The time slices are indicated by simultaneity planes, which run perpendicular to the time axis. On these simultaneity planes, events happen at the same moment in time. The problem is that in Newtonian mechanics there exists no upper limit to the velocity of a particle. Particles are allowed to go infinitely fast, which means that the inclined world-line in Fig. 11.2b leans more and more towards the spatial (x-) axis, as the particle approaches its infinite speed. In a split fraction of an instant the particle would go infinitely far. But Einstein's theory introduces *c* as a limit velocity, which means that all material particles must remain below this limit. In fact only massless

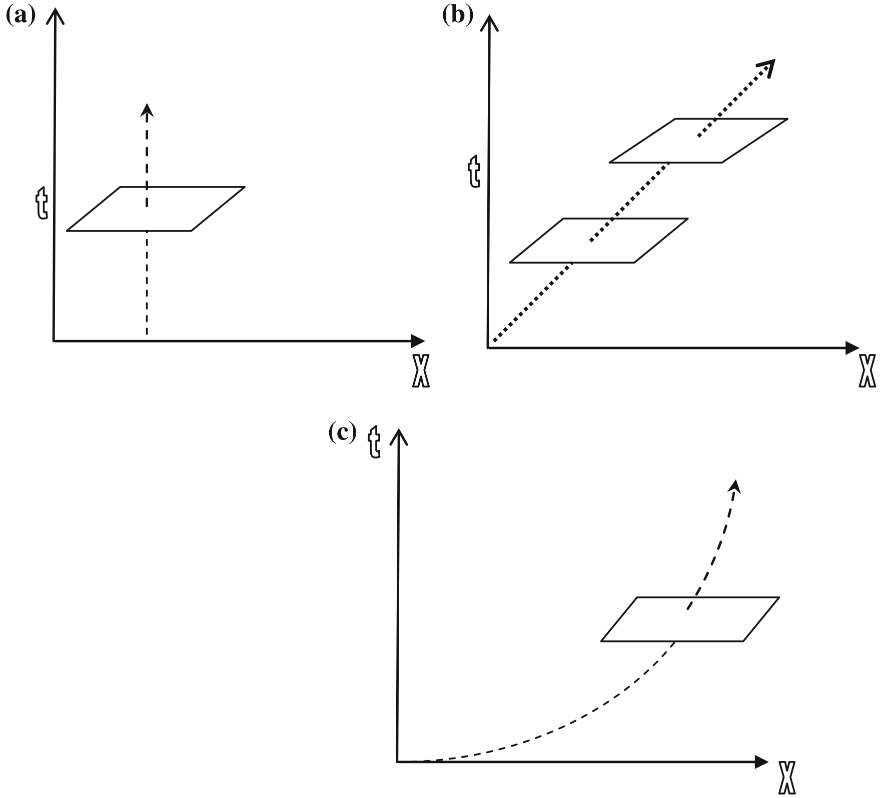


Fig. 11.2 Objects at rest (**a**), at constant speed (**b**) and in acceleration (**c**)

photons travel at the speed of light. Their world-lines mark the boundaries of light cones in space-time diagrams (Fig. 11.3).

Let us now concentrate on a particular event, E_1, at the space-time point near the **Here-Now**. Two observers, moving into their forward light cone, pass the event, E_1. One observer is at rest, and the other observer is moving with constant velocity relative to the first observer. Each observer must specify four data to describe the event. The stationary observer will use the coordinates (x, y, z, t), the moving observer will use the coordinates (x', y', z', t'). What events these observers judge as happening simultaneously is relative to their state of motion (Fig. 11.3a, b). As the simultaneity planes of the two observers do not coincide, the two observers do not judge the same events as being co-present with E. There is thus no absolute sense of *Now*. Hence the separation of the four-dimensional order into space and time coordinates depends on the state of motion of each observer. The lamination of space-time yields different *Nows*. Space and time seem to be relegated to the observer. The Special theory of relativity seems to vindicate the Kantian view that temporal and spatial judgements have their source in human minds.

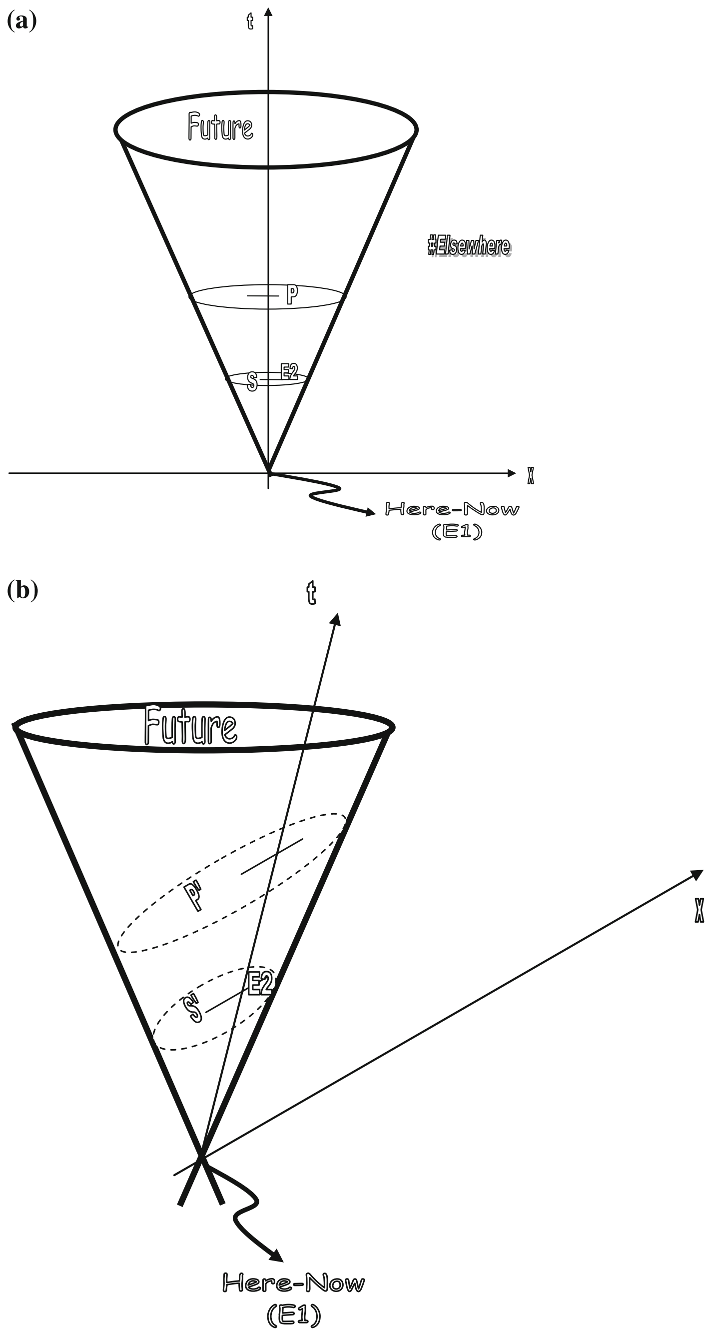


Fig. 11.3 **a** Stationary observer and simultaneity planes *P, S* and events. **b** Moving observer and simultaneity planes to *P*′, *S*′. Two observers, one at rest (as in **a**), the other in relative motion with respect to the first (as in **b**) will not agree on the simultaneity of events

11.2.2 The Special Theory and Determinism

The Special theory and the relativity of simultaneity had significant philosophical consequences, especially regarding our understanding of time (Part IV). But the Special theory has also been associated with determinism and even fatalism (Sect. 11.2.3). If the Special theory is deterministic the consequences seems to be threatening commonplace assumptions: it seems that it makes no room for the arrows of time and offers no hope for a belief in free will. It seems that some form of probabilism—a choice of alternatives towards the future—will be needed both for the existence of temporal arrows and the exercise of free will. As we have seen, classical mechanics is often presented as a deterministic theory. Since the Special theory is 'only' an extension of classical mechanics, it is tempting to conclude that it too is a deterministic theory. In the Laplacean-Newtonian scheme of things the laws of motion and the initial conditions, which obtain at particular moments of time, suffice to completely constrain the behaviour of a system for all future and past times. Similarly for the Special theory:

> Determinism, in special relativity, can be formulated as the fact that initial data on any given simultaneous space **S** fixes the behaviour in the whole of the space-time. (…) There is a stronger statement that one can make, however. If we want to know what is going to happen at some event *P* lying somewhere to the future of *S*, then we only need the initial data in some bounded (finite) region of *S*, and not on the whole of *S*. This is because information cannot travel faster than light (…). (Penrose 1989: 277; bold and italics in original; see Fig. 11.3a, b)

This statement seems to make sense since the Special theory only imposes a limit velocity, in the form of c, but does not change the laws of motion in an essential way. But if the Special theory is deterministic and reflects a deterministic world, it is tempting to conclude 'that there is no becoming, no flow of time, and no free agent in a four-dimensional world' (Petkov 2005: 122). Becoming implies some future branching of events. Free will implies that there exists some choice for an agent, some alternative courses of action. The unspoken assumption is that the Special theory does apply to human beings. Then two alternatives offer themselves: (a) the Special theory implies determinism and has Laplacean consequences; or (b) it will have to be compatible with some form of probabilism or indeterminism.

To inject some form of probabilism into the four-dimensional world of the Special theory it needs to be possible that the initial data, contained within the particular past light cone of some point P, do not constrain events (E_1) at *Here-Now* to one unique trajectory. What happens in the past light cone would have to be compatible with an alternative event E_2 (where $E_1 \neq E_2$). If the Special theory is truly deterministic, in a Laplacean sense, such a possibility must be excluded. But we have already seen that determinism can make room for alternative trajectories, leading to E_2, if it is allowed that an interference at Here-Now (E_1) changes the initial data. An event E_2, which lies in the future light cone of event E_1 may remain ‘indefinite’ in the sense that there are ‘many real alternative possibilities associated with (or corresponding to) the space-time location E_2’ (Maxwell 1985: 25; cf. Dieks 1988; Maxwell 1988; Penrose 1989). Events within a particular light cone of an observer have *time-like* connections, because event E_2 can be reached from event E_1 by familiar signals. As we have seen, a change in initial conditions, whilst keeping the laws fixed, leads to alternative histories. Determinism need not have a strictly Laplacean flavour.

Nevertheless an oddity arises if two *space-like* separated events are considered. These events, which lie outside an observer’s light cone in **Elsewhere**, could only be reached by superluminal signals (Fig. 11.3a). The oddity seems to show that probabilism and the Special theory are incompatible. Probabilism expresses the view that the future is open to alternative histories, whilst the past has frozen to one fixed path. Therefore the past cannot be changed but the future is still open. Unlike Laplacean determinism, however, probabilism assumes that the fundamental laws are probabilistic, not deterministic. This seems to imply that there exists an absolute distinction between the fixity of the past and the openness of the future. But such an absolute distinction between past and future is in contradiction with the Special theory, according to which there exists no universal Now, which constitutes the cut-off point between the fixity of the past and the openness of the future. In particular, observers in relativistic motion with respect to each other do not agree on the simultaneity of events. According to the Special theory time is no longer a universal feature, on which all observers agree. The problem arises for *space-like* separated events: what lies in the future for one observer may already be the past for another observer. Imagine two observers between whom a space-like separation exists, which means that they lie outside each other’s light cones. *A* moves relative to *B* who is seen as ‘at rest’. These two observers do not agree on the temporal order of events so that a question arises for them as to whether a future event is already definite or still indefinite (Fig. 11.4).

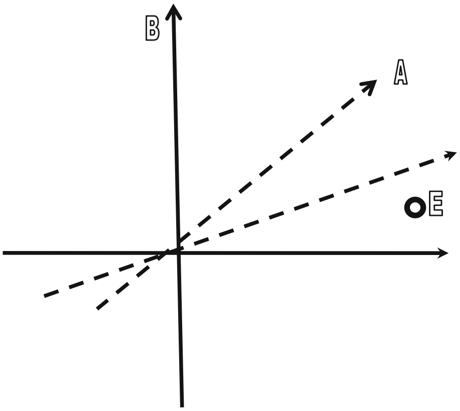


Fig. 11.4 Conventional temporal order for space-like separated observers. At t_1 event *E* lies in the future for observer *B* but it is already past for observer *A*

The two observers who disagree about the simultaneity of events will, for instance, come to different conclusions regarding *gedanken*-experimental rumours, according to which an Andromeda fleet is intent on invading planet Earth. For one of the observers the fleet—event *E*—is already on its way, while for the other the invasion has not yet been launched (and could possibly be prevented) (Penrose 1989: 260, 393; Fig. 11.4). How can the same event already have been decided for one observer but still be undecided for the other? If probabilism means that there must be a universal distinction between a fixed past and an open future, this assumption is in contradiction with Special relativity. Without such an assumption it would not be possible to speak of an open future and a closed past. Hence probabilism, so the argument runs, cannot be compatible with the Special theory. The whole of space-time must already be fixed, in the Laplacean image of the Block Universe. There is no room for uncertainty, no room for an open future, no room for free will or arrows of time. This looks indeed like an odd consequence of the Special theory. But its effect is alleviated by the fact that *space-like* separated events, unlike *time-like* separated ones, have a conventional time-order. *Space-like* connected events are not causally connected and can either see their order reversed or can be made simultaneous with the origin of another observer's coordinate system by an appropriate choice of coordinates. In fact for such *space-like* separated events it is meaningless to ask which event comes first (Penrose 1989: 371). By contrast, *time-like* connected observers may disagree about the simultaneity of events but they will agree on their temporal order. For time-like connected events, space-time allows for an arrow of time, as will emerge presently.

If the Special theory is indeed deterministic it seems to imply a static interpretation of space-time, namely that space-time already lies in wait, like the frames of a film or contour lines on a map, containing all events, which will ever happen. Such a static Block Universe, which Laplace's Demon is free to presuppose, is however only *one* interpretation of Minkowski space-time and not a necessary consequence of it. An alternative interpretation—the Evolving Block Universe (Ellis 2013)—suggests that the future is still open and that the four-dimensional world unfolds with the 'creation' of new events. The universe originates in the Big Bang and expands to its current constellation towards what looks now like a Big Chill. Such a view makes room for an asymmetry between past and future. In an Evolving Block Universe, Laplace's Demon would see an arrow of time (Part II, Sect. 11.1).

It is a dynamic view of space-time. But how does this interpretation avoid the argument, just discussed, that probabilism is incompatible with determinism in the Special theory? The argument assumed that probabilism implies an open future, since the laws themselves are probabilistic. Events occur with greater or lesser probability. But the laws may remain deterministic, if the boundary conditions are allowed to vary. Time-like connected events permit alternative trajectories, since it is sufficient for chance events to change the initial conditions whilst keeping the laws invariant. But *space-like* connected events are too far apart in space and too close in time for any finite signals to link them. How they are temporally ordered is a matter of convention; their temporal order therefore has no physical significance and does not support probabilism in the above-mentioned sense.

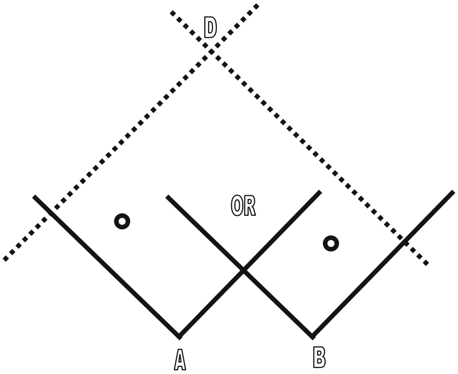


Fig. 11.5 Mutually inaccessible events for space-like separated observers *A* and *B*. However, these events become accessible to Popper's Demon, who resides at *D* outside these light cones and retrodicts the events since they lie in his past light cone

But if the Laplacean Demon perceives an arrow of time, even in a deterministic world, does he retain complete predictability? Imagine two observers whose light cones overlap (Fig. 11.5). There will be events in *A*'s and *B*'s future light cones, which are not yet accessible to *A* or *B* respectively since they lie outside the overlapping (OR) regions, which *A* and *B* share. Let us introduce another demon—dubbed **Popper's Demon**—who resides at a point, *D*, from where he can see both light cones and their overlap. He would be able to calculate what happens. However for such a Demon the calculation would be a retrodiction! Karl Popper concludes from this thought experiment that the Special theory is not a fully deterministic theory, since the Demon can only retrodict.

> If we try to introduce the (…) demon into special relativity, we find that we can calculate, from the demon's region of information, a lower bound for the demon's spatio-temporal position D; and we further find that the demon calculated only an event within his own past. (Popper 1982: 61; italics removed)

Popper's Demon can inform *A* and *B* about events, which are still inaccessible to them. Popper's argument is inspired by his reservations about predictive determinism. Predictors cannot predict all of their future states because of the threat of unexpected or unplanned interferences (Part II, Chap. 8). It remains the case, however, that *time-like* connected events are predictable but with the possibility of alternative events in the future light cone of *P*, *S*. So if the Special theory is not fully deterministic, in the Laplacean sense, what are we to make of the claim that it implies fatalism?

11.2.3 Fatalism and the Special Theory

If the Special theory is not fully deterministic in the Laplacean sense, we should not expect it to be fatalistic. Yet the claim is sometimes made in the literature that the Special theory leads to fatalism and the denial of free agency. (The question of free will has implications for the nature of mind debate, which will be taken up in

Part III). Space-time and the world-lines, which furrow through it, lie before the eyes of a Laplacian Demon like a landscape in four dimensions. (Popper's Demon enjoys the panorama of many overlapping light cones.) This view seems to have severe implications:

> The implications of a four-dimensional world for a number of fundamental issues such as temporal becoming, flow of time, free will, and even consciousness are profound – in such a world (often called the block universe) the whole histories in time of all physical objects are given as completed four-dimensional entities (the objects' worldtubes) since all moments of time are not 'getting actualized' one by one to become the moment 'now', but form the fourth dimension of the world and hence are all given at once. And if temporal becoming and flow of time are understood in the traditional way – as involving three-dimensional objects and a three-dimensional world that endure through time – there is no becoming, no flow of time, and no free will in a four-dimensional world. (Petkov 2005: 122)

Becoming and the flow of events become relative to the perceptual mode of living beings. Sentient creatures are compelled to explore little by little the content of the four-dimensional world, as each travels on a time-like trajectory in space-time. But for a Laplacean Demon everything is already written down, from past to future. (Recall, however, that such a view presupposes a static rather than a dynamic view of space-time.)

> Nature will take one of the alternatives open to her, and it is this that we must imagine inscribed, even though we do not know what 'it will be'. (Costa de Beauregard 1966: 430)

If past and future are already determined and events exist eternally, their grooves etched into space-time, it seems that the Special theory can also be mustered to embrace a fatalistic view of human existence.

> In the Minkowski four-dimensional world, there is no free will since the entire history of every object is realized and given once and for all as the object's worldtube. Therefore, free will may exist only in a three-dimensional world. (Petkov 2005: 152; cf. Lockwood 2005: 162–164, 254–256; Deutsch 1998: Chap. 11)

It seems that the Block universe commits us to fatalism:

> Should we all, then, if we find ourselves persuaded by the tenseless view [i.e. Block universe], become fatalists? (…) If being a fatalist means believing that, as of your current here-now, every meaningful question about the future already has a determinate answer, then yes: accepting the tenseless view of time requires you to be a fatalist.

But fatalism of this kind, continues Lockwood, is not to be confused with a passive attitude towards one's own life: 'stoically waiting for the inevitable to happen.'

> Specifically, it would be thoroughly *ir*rational, at least in ordinary circumstances, to take a fatalist attitude, thus understood, of outcomes that you believe to be causally dependent, in part, on decisions that you yourself have yet to make, or still have time to rescind. (Lockwood 2005: 162; italics in original)

Such views attribute to scientific theories wide-ranging claims. A scientific theory, like the Special theory of relativity, has the power to inform us that the material world is a four-dimensional Laplacean Block Universe, the laws of which determine both the motions of all material particles *and* the life events of its conscious human inhabitants. It is worth recalling, however, that Laplacean determinism and fatalism do not make equivalent claims. The world-lines of a particle are only determined in terms of the laws of motion on the basis of given initial data, which are situated on a simultaneity slice. If these initial data change, the trajectory of the particle will be diverted, under the governance of the existing laws. Fatalism, however, regards all events as pre-determined, irrespective of human agency and any interference in initial conditions (cf. Earman 1986: Chap. II, §10). The claim that a Laplacean Demon is able to see the trajectories of all particles, including human beings, into the whole of future—the Demon is able to preview the shape of an object's worldtube—in no way rules out the possibility of alternative trajectories, as a result of interactions and interferences. However, if two events are *time-like* connected and an event P_2 can be indefinite with respect to event P_1 in the direction of its future light cone, fatalism does not seem to be a consequence of the Special theory. The Special theory is an extension of classical mechanics but its tentacles do not reach into the Social Sciences. If the Special theory gives rise to partial determinism, the question arises whether classical mechanics (CM) is as readily deterministic as the Laplacean Demon makes it out to be. Is it perhaps the case that there are also limits to classical determinism?

11.3 The Limits of Determinism

Omniscient as Laplace's Demon may be, he is mistaken about the extent of classical determinism. The determinism of classical mechanics has its limits, as Eddington's Demon seems to accept. Whilst in Newtonian physics it is perfectly possible, for instance, to compute the trajectory of two planets—their positions can both be predicted and retrodicted for a certain length of time—the motion of a three-body system poses computational difficulties. Although the motion of the planets seems to constitute a paradigm of a deterministic system, the so-called **three-body problem** in classical mechanics shows the limits of this assumption. The French mathematician and physicist Henri Poincaré realized that their orbits may impose limitations on predictive determinism. Firstly, it may not be possible to specify the initial conditions accurately.

> If we knew exactly the laws of nature and the situation of the universe at the initial moment, we could predict exactly the situation of that same universe at a succeeding moment. But even if it were the case that the natural laws had no longer any secrets for us, we could still only know the initial situation *approximately*. If that enabled us to predict the succeeding situation with *the same approximation*, that is all we require, and we should say that the phenomenon had been predicted, that it is governed by laws.

Secondly, slight errors in the specification of the initial conditions will lead to breakdowns in predictions.

> But it is not always so; it may happen that small differences in the initial conditions produce very great ones in the final phenomena. A small error in the former will produce an enormous error in the latter. Prediction becomes impossible, and we have the fortuitous phenomenon. (Poincaré 1952b: 68)

Although the orbits follow deterministic laws, they are not guaranteed to remain stable. Their stability and hence their predictability is not assured for eternity. In this sense, the motion of the solar system is 'chaotic' since it is not possible to compute and predict its evolution over a span of millions of years (Laskar 1989). The motion of the planets constitutes another example where Laplacean determinism and perfect predictability do not coincide (cf. Penrose 1989: 226). It is possible to assume metaphysical determinism and face predictive indeterminism. Should this instability bother the Laplacean Demon? The Laplacean Demon possesses exact knowledge of the initial conditions and the laws of planetary motion. For the Demon the motion is perfectly predictable. But for his human counterparts, as Poincaré pointed out, the orbits are not perfectly computable. The evolution of the solar system can be predicted over 10 million but not 100 million years. If a small perturbation in the initial conditions will result in '100 % discrepancy after 100 million years', this discrepancy constitutes an indeterminacy in the motion of the planets. If this discrepancy is not merely a limit in human predictability, which would not affect the Demon, but a genuine indeterminacy in the motions of the planets, it is to be expected that even the Demon would lose sight of the future trajectories of the planets. If the world is truly indeterministic, even on a large-scale scale, like planetary motions, the Laplacean Demon is powerless to restore deterministic behaviour. So the Demon poses a challenge: he assumes that the universe is ontologically determined but his human counterparts fail to compute and predict the exact long-term evolution of the solar system. Is it truly indeterministic or does it only appear indeterministic to human observers?

But it is not necessary to consider the planets' motion. Indeterminism can occur when **triple collisions** of rigid bodies are involved (Penrose 1989: 218–219; Earman 1986: Chap. III). The reason for their indeterministic behaviour lies in the order, in which collisions of the three bodies are imagined to occur. One possible order is that particle 1 and particle 2 come together first and particle 3 hits particle 2 immediately after this first collision. Another possible order is that particles 1 and 3 come together first and particle 2 collides with particle 1 immediately afterwards (Fig. 11.6). The resulting behaviour is indeterministic since it is not the case that the initial conditions of the particles (their positions and velocities) determine their future behaviour in a continuous way. The same past constellation leads to different futures. Yet the

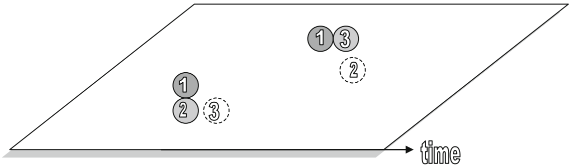


Fig. 11.6 Indeterminism as a result of the order of collisions, adapted from Penrose (1986: 219)

bodies at any one time only have a finite range of angles at which they can recoil. And hence if their initial conditions were known, the angle of reflection could in principle be determined. It would require a superhuman demon to make it computable or predictable. If the number of particles is increased, perhaps to infinity, indeterminism increases. Such systems may become non-computable, even for a demon. These examples emphasize that determinism and computability (predictability) may diverge. Chance events, by contrast, are indeterministic, yet computable (Part II, Sect. 11.4). The notions of indeterminism and non-computability do not always coincide either.

But predictive determinism can fail in classical mechanics even when a **single particle** is involved. As a concrete example consider the fate of the *Philae* lander, which landed on comet 67P/Churyumov-Gerasimenko on November 12, 2014, after a ten-year journey from Earth. The lander, after having separated from its mothership, *Rosetta*, bounced twice before it came to rest in the shadow of a cliff, where it receives insufficient sunlight to recharge its batteries. The two bounces were unplanned. They happened because the harpoons, which were meant to anchor *Philae* to the frozen surface of the comet, failed to fire. Once the lander had come to rest, scientists were able to calculate, retrospectively, where the lander could have landed and how far the two bounces had carried it away from its intended landing place. The lander's final resting place was unpredictable but it was determined by the local conditions of the uneven surface of the far-flung comet.

Unpredictability of a mechanical system is not just due to fuzziness in the initial conditions but may be expressed in the laws of motion themselves (see Hutchison 1993, 1995; Norton 2003: 12). Consider a particle with mass *m*, which moves inertially with constant velocity without friction from an initial position, x_0, towards a point x_1, where it is reflected elastically and returns to x_0; the motion repeats itself. Although the particle moves between two fixed points, x_0 and x_1, and its velocity is constant, it experiences a deviation $\Delta x = \Delta x_0 + t\Delta v_0$, which increases with time, *t*. Determinism rules out such a deviation $(\Delta x_0, \Delta v_0)$ and requires exact values for x_0, v_0. The fact is, however, that the final position will be undetermined, which means that after a critical time $t_c = l/\Delta v_0$, the particle could be found anywhere in the interval $x_0 < x < x_1$ (see Born 1955: 166–167; Penrose 1994: 22) (Fig. 11.7a, b). In this case, indeterminism and non-computability go hand in hand.

This example shows, once more, that determinism is based on an extreme idealization, which forbids any deviation from exact values, like $\Delta x_0, \Delta v_0$. In real physical systems, however, small deviations do occur and can amount, as the present example illustrates, to an indeterminacy, Δx.

In the 'reflected particle' example the ball moves back and forth at ordinary speeds. But a further problem for determinism arises when infinite velocities are involved. As we have seen the Special theory restricts the velocities, at which mass particles can travel to below the limit velocity of *c*. The Special theory prohibits any mass particle from travelling at the speed of light. The greater its mass, the greater the energy it would require to approach the limit velocity of *c*. Subatomic particles, like electrons and muons, however, do travel at speeds near the velocity of light photons. The additional problem for determinism arises from the fact that

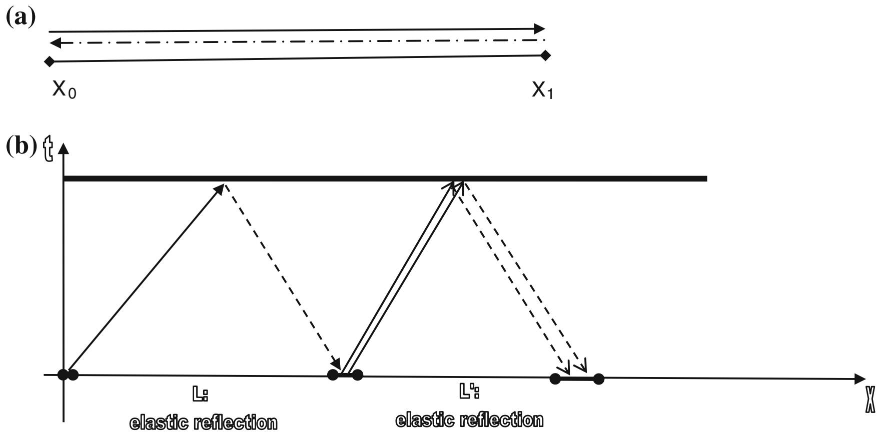


Fig. 11.7 **a** Straight-line motion of a particle and elastic rebounce. **b** Reflected motion of a particle between two endpoints, leading to an undetermined final position within a given interval

Newtonian physics does not impose any limit on the velocities at which bodies are allowed to travel. Although the finite velocity of light photons was discovered in Newton's time (by O. Rœmer in 1676) it was established as an empirical value. It was regarded as the fastest speed particles could achieve but it was not a limit speed in Newtonian mechanics. Newtonian mechanics therefore did not preclude the acceleration of a particle to potentially infinite speeds (v » c). According to Newtonian mechanics, a particle could therefore literally disappear from the universe in a finite amount of time. How far can such an imaginary particle travel once it reaches a cruising speed of $v \gg c$? As

$$distance = velocity \times time$$

theory permits it to travel an infinite distance in a finite amount of time. It can disappear behind the observable horizon, beyond the curtain of the farthest galaxies billions of light years away, in a fraction of a second. Earth-bound observers would lose sight of it in their visual field. In addition to the theoretical possibility of infinite velocities, classical mechanics is also time-reversal and time-translation invariant. Hence it permits the time-reverse of the disappearing particle—time symmetry allows the replacement of t by $-t$. Temporal symmetry means that Newtonian mechanics allows **space-invaders** at a time $t = 0$, say, who appear in the observers' visual field out of nowhere. But contrary to deterministic expectations, the initial conditions, prior to $t = 0$ (the sudden appearance of space-invaders) and the classical laws do not entail the appearance of space invaders. The initial conditions and the classical laws do not determine the appearance of space invaders from infinity (see Malament 2008, §3). In fact the initial conditions beyond the observable universe are not known. Even if the space invaders obey Newton's laws

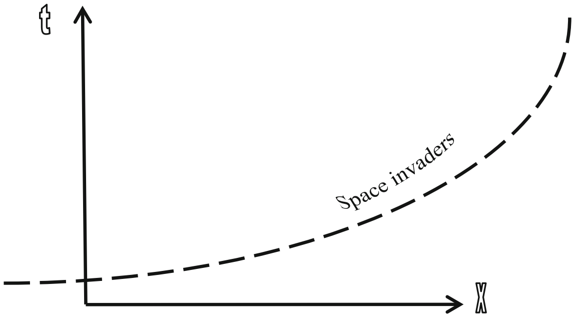


Fig. 11.8 Space invaders arriving from infinity

their arrival cannot be predicted, because the boundary conditions of their 'invasion' are unknowable (Fig. 11.8).

Laplacean determinism has limits both in classical physics and in the Special theory of relativity. What role then does chance play in the classical world?

11.4 Determinism and Chance

It was noted earlier that determinism—due to the difference between its metaphysical, predictive and causal versions—is not to be identified with computability or predictability. As the existence of a Laplacean Demon would demonstrate the world may be ontologically and causally determined—the present state becomes the cause of a unique future state of affairs—although this evolution may not be computable or predictable to the human mind. But the Laplacean Demon is lucky: no distinction apparently exists for the superintelligent Demon between the ontological determination of the world and its predictability. If, however, the world is not completely deterministic—or has elements of chance or even randomness—the Demon may fail to predict its evolution. Such randomness is often located in initial conditions.

In Part III we shall make the acquaintance of another demon—Maxwell's Demon—who is able to compute the motions of individual molecules in a volume of gas—their individual trajectories are taken to be governed by deterministic laws—although it is beyond the ability of a human researcher to compute the path of each molecule in gas containing, say, 10^{18} molecules. Maxwell's Demon is actually more modest than Laplace's Demon. He only claims to be able to manipulate individual molecules in a sealed container of gas!

It is useful to keep the notion of computability in mind because it helps to clarify the connotation of indeterminism with chance and randomness. If determinism means that, given initial conditions, the future and past state of affairs of a physical system can be uniquely determined—it is bi-directionally determined—then indeterminism should mean that the laws, in conjunction with the initial conditions, do not uniquely determine future trajectories. But this does not necessarily mean that

Fig. 11.9 A branching tree of probable effects from probable causes, compared with a deterministic cause

$$\text{'If } C \text{ is a } \textit{probable} \text{ cause, then } \left\{ \begin{array}{l} E_1 \text{ (Probability } \alpha) \\ E_2 \text{ (Probability } \beta) \\ E_3 \text{ (Probability } \gamma) \end{array} \right\}$$

'If *C* is a *deterministic* cause, then *E* (Probability 100%)'

and

'If non-*C*, then probably non-*E* (unless there is an alternative cause).

the trajectories are non-computable. The probability of the occurrence of chance events can be computed (just as the radioactive decay of atoms can be predicted on average). Consider some cause *C*, which instead of launching a unique chain of events, leaves a branching tree of probable events (Fig. 11.9).

In this situation, a cause *C* gives rise to three potential effects, the probability weight of which can be computed. It is also clear that in the absence of *C*, the effects may not arise. Such effects are commonplace. For instance in certain experiments in quantum mechanics, depending on the set-up, a particle has a 50 % chance of being registered deflected upwards and a 50 % chance of being registered deflected downwards on a recording screen (Fig. 17.1). A more mundane example is coin flipping where each side of the cube has a 1/6 chance of occurring. The statistical distribution of observable effects can be specified. In such situations there exists a lawful statistical dependence of the consequent condition on the antecedent condition. Probabilistic determinations can be associated with deterministic chances; that is chances which have a computable probability of occurring.

Such chance events, which can be expressed in probability functions (as in the example just given), should not be confused with randomness. Randomness is a function of unpredictability. Random events are patternless events, which lack predictability and hence computability. Randomness or non-computability implies an unpredictable order of occurrence, as in the decimal expansion of irrational numbers, like $\pi = 3.141421356237309505\ldots$, since it is not predictable when or whether the sequence might repeat itself. Randomness is sometimes associated with chaotic behaviour. But random behaviour does not depend on the previous history of a system, whilst the trajectories of chaotic systems are highly sensitive to initial conditions (see Gleick 1998; Peterson 1993). Although laws and initial conditions are available it is impossible to make long-term predictions as, for instance, in the case of planetary motions. In such cases slight changes in initial conditions lead to divergent trajectories.

The Laplacean Demon could easily cope with deterministic chances but would he be defeated by randomness? Laplace equips the Demon with three valuable attributes (Frigg et al. 2014):

1. *Computational omniscience*, i.e. he is able to calculate both the past and future history of all systems—both micro- and macro-systems—from their current states of affairs.

The Demon is able to enjoy computational omniscience because he possesses a second attribute.

2. *Dynamical omniscience*, i.e. he is able to determine the true, not an estimated, time evolution of the system under consideration. In the Demon's case ontological and predictive determinism coincide. The predictions he makes correspond precisely to the actual evolution of the system, the trajectory of which he calculates. There is no room for randomness.

Finally the Demon possesses:

3. *Observational omniscience*, i.e. he is not only able to determine the initial data of the current state of affairs from which he will compute both the past and the future. Laplace's Demon is a true inhabitant of the static Block Universe, since the whole vista of the course of the universe lies before his all-encompassing vision, like the frames of a filmstrip. Hence for the Demon Past, Present and Future seem equally real.

This Demon requires a commitment to a metaphysical version of determinism, for which there is little evidence in the real world of the human scientist. Yet Laplace's Demon poses a challenge: how far is real science from such an ideal and how close can the human apprentices of the Demon get to the master's vision?

Let us retrace our steps. We have seen that even in a deterministic world, the Demon could detect an arrow of time if we adopt the view of an Evolving Block Universe. We have discussed determinism and its limits both in classical mechanics and the Special theory of relativity. We have considered to which extent even the classical world may be a chancy world. We have reviewed the claim that the Special theory embodies fatalism, and hence a denial of free will. Now that we have at least a working notion of determinism, let us consider its connection to the thorny question of free agency.

Chapter 12
Determinism and Free Will

> *Man is already a masterpiece of creation because he believes that, despite all the determinism, he acts freely.*
> Lichtenberg, *Sudelbücher* (Band II, Heft J: §1491)

For the fatalist the question of free will does not arise. No amount of tampering with current conditions will have an effect on what fate has in store. Such a view may have serious implications for the way individuals may choose to lead their lives. If the time of my death is already pencilled in the diary of the universe, even before my own birth, then it makes no difference to how I lead my life. I may engage in extremely risky behaviour since a more cautious existence will not postpone my mortal end. As we have seen, fatalism is sometimes presented as a consequence of the Special theory of relativity.

> Einstein, Eddington and Jeans were right – the world is a four-dimensional block, without any room for free will. All events are eternally real – no need to be tormented by 'what might have been'. At no time are future events anything other than actualities lying in store for us, although unknown to us. (Paraphrased from Lockwood 2005: 68–69)

We have discarded fatalism as an acceptable *consequence* of the Special theory. But the question still remains whether determinism as it is generally understood still poses a threat to the idea of free will. Free will is characterized as the ability of rational agents to make a genuine choice between alternative courses of action, which are freely open to them prior to the moment of decision. Such a characterization of free will is difficult to make compatible with scientific determinism. According to Laplacean determinism every event has a prior cause, which uniquely determines the event. In fact for every chosen event, *E*, there is a unique chain of prior causes. There is no chance or randomness. A person's decision, then, to do X can only apparently be a 'free decision' because the resolve to do X must have been caused by some anterior condition in *P*'s life. But even such anterior conditions will have a previous cause so that what a person *P* decides now will have causes in *P*'s past stretching back to the beginning of the universe. On this view even a human being is imprisoned in a chain of cause and effect.

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_12

It lies in the nature of human decision-making that many decisions have a binary nature: either X is done ('go on holiday') or non-X is done ('do not go on holiday'). Once a decision is taken, its bifurcation will continue: either Y is done ('go on holiday *in France*') or non-Y is done ('do not go on holiday *in France*'). (A look at a restaurant menu seems to suggest that sometimes more options are available but this does not change the binary nature of the choice: choose item X at the exclusion of all other items.) But Laplacean determinism seems to rule out binary choice—a choice which a human seemingly makes has its cause in that person's past. A correlate of scientific determinism in the human mind is neural determinism. As we shall see in Part III, it explains the nature of the human mind, of which free will is a feature, in terms of neural wirings in the brain.

But for the rest of this chapter let us focus on the free will question. Science seems to tell us, controversially, that humans have no free will. What response is there? As it turns out the response depends on which version of determinism is adopted.

12.1 Responses to the Problem of Free Will

The age-old free agency debate has generated a cluster of responses, which can roughly be categorized as follows.

According to **compatibilism** or soft determinism (Th. Hobbes, G.W. Leibniz, D. Hume, I. Kant, J.S. Mill) free will and determinism are compatible. The physical world may be physically determined but the agent remains free to act because freedom resides in human agency. The action is free because the agent is the source of the action: the action originates from the agent. Soft determinism regards binary choice as genuine.

But **hard determinism** (B. Spinoza, P.H.Th. d'Holbach, J.O. LaMettrie, A. Schopenhauer, St. Hawking) denies the reality of free will since, due to the effect of prior causes, the agent could not have acted otherwise. The agent is bound in a chain of cause-and-effect relationships such that any impression of free choice between options is only apparent. Only hard (Laplacean) determinism rules out binary choice.

Libertarians (Aristotle, Lucretius, A. Eddington, D. Davidson, R. Kane), by contrast, deny the truth of determinism and affirm that free will is logically incompatible with the Laplacean image of a deterministic universe. This denial may be partly justified, because both classical mechanics and the Special theory are only partially deterministic. The agent is the source of freely chosen action; the agent's decisions obey a logic of reasons, not a logic of deterministic causes.

Hard determinists and libertarians are incompatibilists. The hard determinist holds that the postulate of a deterministic world means that it is incompatible with the assumption of free will. Libertarians claim that humans enjoy free will and that this freedom is incompatible with the assumption of a deterministic world. Both compatibilists and libertarian incompatibilists consider that human agents are free, to a certain extent, in the physical world. This freedom is due to a logical difference between *reasons* and *causes*. Human decision-making is inspired by reasons, not material causes, which are confined to the natural world. The principle 'same cause—same effect' seems to hold in the natural world but this principle cannot be applied to human actions. It is not the case that the same reason always leads to the same effects. Two people, say *A* and *B*, may both wish to act in the interest of their country, hence they share the same reason but this attitude may lead to contrary effects. Both have their chance at the next election. But *A* may cast a vote for party *C*, whilst *B* may think the country is best served by party *L* and votes accordingly. Both cite the same reason for their actions but with very different effects. Furthermore, causes and effects can be separated from one another in a way that actions and reasons cannot be kept apart. It is possible to consider a cause without concern for the effect and vice versa; but it is hard to consider an action without taking into account the reason for the action. (This is called the logical connection argument.) If *A* is friendly to *B* (despite their political differences) *B* will attribute to *A* the intention that *A* means to be friendly to *B* and that *A* acts in accordance with this reason. (*B* may nevertheless entertain, as a possibility, that *A* only pretends to be friendly but he would eventually be able to establish *A*'s true disposition.) On the other hand, if *A* acts in an unfriendly manner towards *B*, *B* will interpret the action as displaying a hostile motivation towards *B*. By contrast a scientist may be interested in studying the physical causes of rainbows without worrying about the effect it may have on human observers. A psychologist may wish to study the effect of rainbows on human well-being without worrying about what physical causes produce rainbows.

Nevertheless, a difficulty arises for the incompatibilist. Reasons are often the causes of our actions. Or at least reasons have causal components. If *A* means to be friendly to *B* then this intention causes *A* to act in an appropriate manner towards *B*. But if it is admitted that reasons have causal components, then the threat of hard determinism seems to return. The causal component has roots in antecedent conditions and hence human beings are not truly free after all.

But is the hard determinist right? If the reason truly originates in the agent, then its causal component has no antecedent cause. This possibility the determinist denies to preserve the 'causal closure' of the world. But even if it is granted that an antecedent cause is at play, it may only be a partial component of the cluster of reasons. To safeguard the 'freedom of the will' and the presence of the mind in a material world generally, compatibilists and libertarians may appeal, in continuation of the Darwinian research programme, to the notion of emergence (Part II, Sect. 12.2).

Laplacean determinism leaves out human agency. Or rather, it makes human agency simply a part of the anterior causal field, which will consequently determine future outcomes. More needs to be known about human agency before conceding victory to determinism. Note that the fact that the Laplacean Demon has a complete, unimpeded view of all the decisions *A* and *B*, for instance, will ever make, in no way implies that the human agent is unfree. The Laplacean Demon does not manipulate the agent. Physical determinism does not imply agential determinism (List 2014: §3; O'Connor 1971; Watson 2003). The Demon exercises no power over the agent in the way that Maxwell's Demon, for instance, manipulates gas molecules. The agent at any one point may consider various options but the Demon already knows which options the agent will choose. The Demon's prior knowledge has no impact on the agent's decision. The determinist may claim that the agent's decision regarding alternative options is not a free choice because, like a loaded dice, the agent's decision-making is biased towards a particular option. Unknown to the agent, this bias is determined by the conditions in the previous history of the agent and the universe. However, the binary nature of choice means that a human agent will always have to consider a cluster of conditions, which in themselves may be incompatible. In the face of incompatible causal conditions in the natural world—say water *and* combustible material—no fire can possibly occur. But in the case of incompatible conditions in the human world—say, the desire to stay at home or go out—the human agent is forced to make a choice between these binary alternatives.

A distinction must be drawn between, say, compulsive behaviour, on the one hand, and rational, willed behaviour on the other. Only rational behaviour has a moral dimension—hence an element of choice—and is adaptable to the natural and cultural environment—hence can be learned. Compulsive behaviour, however, crucially lacks this dimension of choice and adaptability. It is certainly true that certain types of compulsive behaviour—say schizophrenia and manic depression—have a strong genetic component. But they are not types of rational behaviour, for they neither change with the environment nor are they the result of learning processes. The same observation holds for aggression—a flagship of socio-biological explanations. A recent study (Brunner et al. 1993: 578–580) has linked abnormal behaviour—'impulsive aggression, arson, attempted rape and exhibitionism'—with a 'complete and selective deficiency of enzymatic activity of monoamine oxidase A (MAOA)', which affects five males in a family. Although this seems to be a case of a genetic-phenotypic link, it is compulsive, not rational behaviour, which is caused by the genetic structure. As the anthropologist Marshall Sahlins has urged against forms of genetic determinism: the symbolic event marks a radical discontinuity between culture and nature (Sahlins 1977: 12; cf. List 2014). Humans are symbolic creatures: they operate with symbolic tools, like language, which includes reasons. They are defined in terms of symbolic attributes, so that there is a basic indeterminacy between human nature and its cultural expressions (Sahlins 1977: 61). If culture is biology plus symbolic faculty, as Sahlins suggests, then biology is only a necessary, never a sufficient condition for social behaviour in humans (Sahlins 1977: xi, 65–67).

The British cosmologist Arthur Eddington suggested, on the one hand, a close correlation between a deterministic universe and a deterministic mind.

> A complete determinism of the material universe cannot be divided from determinism of the mind. There can be no fully deterministic control of inorganic phenomena unless the determinism governs mind itself.

On the other hand:

> (…) if we wish to emancipate mind we must to some extent emancipate the material world also. (Eddington 1932: 310)

Eddington concludes that there ‘appears to be no longer any obstacle to this emancipation.’ In terms of the debate about free will, Eddington’s suggestion amounts to the denial of determinism in the material world. As we have seen the classical and relativistic worlds harbour elements of indeterminism. But how can it be shown that the ‘mind’ is indeterministic, too? Refutation of determinism could be taken to be an affirmation of indeterminism, especially indeterminism in the atomic realm. As mentioned before, certain radioactive decay events occur spontaneously. According to Lande’s Demon, quantum events may be *causa sui* (see: http://www.informationphilosopher.com/solutions/demons/). Could quantum events serve as the basis for free will? Unfortunately, indeterminism in the realm of subatomic particles happens on a scale, which lies well below the range of human actions. The indeterminism of the quantum world cannot establish the free will of human beings, if it exists. Human actions fall squarely within the range of the classical world. Although determinism applies to the classical world, it is limited, as examples like the three-body problem, multiple collisions, space-invaders and a single particle oscillating between two end points demonstrate. How can we know that the determinism of the classical world, where it applies, does not affect human action? The determinism of the classical world could be mirrored in the determinism of the brain. And so hard determinists would be vindicated in their denial of free will.

Neuroscientists have a tendency to affirm, in Eddington’s words, that determinism governs neural pathways in the brain, which determines the mind itself. Whatever view one holds regarding the relationship, it is analytically convenient to separate the **mind** from the **brain**. In a computer analogy the mind is the software, the brain is the hardware. Descartes regarded the mind as an immaterial substance and the body as a material substance. The mind is the home of emotions and thoughts; the brain is the area of neural activity. But Descartes’s dualism faces the unsolved puzzle of how a material brain can affect an immaterial mind; and conversely how a mental phenomenon can have an effect on bodily functions. (How can anxiety and stress increase blood pressure?) In the light of this puzzle it is tempting to identify the mind with the brain.

The neuro-scientific view is that ‘brain mechanisms cause mental events.’ Neuro-science treats the ‘I’ as being synonymous with an individual’s brain (Haggard 2011: §2). Such a view amounts to the adoption of some kind of neural determinism, which makes it difficult to justify the human feeling of free will (Falkenburg 2012). This view is reminiscent of Freud’s deterministic view of the mind. The Freudian Demon controls, which hidden motives are allowed to enter conscious awareness. According to Freud there is nothing chancy in the mind. It is not the gene but the unconscious hidden *Id*, which determines our behaviour. Neuroscience replaces the Id with the gene. Neuroscience makes it look as if there were a 1:1 mapping of neuronal activity with mental states. Such views are widespread.

According to physicists, like Stephan Hawking,

> …the molecular basis of biology shows that biological processes are governed by the laws of physics and chemistry and therefore are as determined as the orbits of the planets. Recent experiments in neuroscience support the view that it is our physical brain, following the known laws of science, that determines our actions and not some agency that exists outside those laws…so it seems that we are no more than biological machines and that free will is just an illusion. (Hawking/Mlodinow 2010: 32; emphasis in original removed)

Neuroscientists hold quite similar views:

> Free will is an illusion. Our wills are simply not of our own making. Thoughts and intentions emerge from background causes of which we are unaware and over which we exert no conscious control. We do not have the freedom we think we have (Harris 2012: 5, emphasis in original removed).

Such views presuppose a mind-brain identity. The brain is a mechanical machine, which 'produces' mental events, like free will activity. Although Leibniz believed in the clockwork universe, he offered the following argument against a mechanical explanation of the mind:

> Suppose that there were a machine, big enough for a human to enter, so constructed as to produce thought, feeling, and perception…On going inside we would only see the parts impinging upon one another; we should not see anything, which would explain thought, feeling, perception. (Leibniz 1714: Sect. 17)

In other words, Leibniz appeals to some notion of emergence: the mind emerges from brain activity. But rather than being a machine, the brain is a biological organ. Evolution wires each brain slightly differently (apparently each knee is different too!). The brain must be studied in terms of biology. It is also subject to biochemical and thermodynamical processes, which are largely stochastic. It constantly rebuilds itself. At least some neuroscientists accept that 'the key to free will may be the complexity and flexibility of the mappings in the brain' (Haggard 2011: §10). The brain's layered structure is better captured in the analogy of a global workspace model (Swinburne 2011b) than a mechanical clockwork. There is no one-to-one correspondence between neural processes and mental phenomena in a global way. Yet neural processes and mental processes are linked. Damaged brains severely impact on mental functions. It was found, for instance, that genetic changes in the brain increase the suicide risk in soldiers and that genetic traits may have an effect on criminal behaviour.

What do damaged brains tell us about the mind? Damaged brains—split brains, blindsight—seem to demonstrate a close link between the mind and the brain. A dysfunctional brain leads to a dysfunctional mind. Does this not show vividly that the mind depends on the brain; in fact that the mind is just another expression for the brain? Would it not be correct to speak of the mechanical mind?

A well-functioning brain, however, is only a necessary, not a sufficient condition for a well-functioning mind (Bayne 2011). In the absence of a well-functioning brain, there can be no cognitive well-functioning processes; but the presence of a sound brain does not imply that the brain determines mental functions. It would be true to say of a clockwork, as a deterministic system, that a faulty cogwheel leads to a malfunctioning clock. And, conversely, the presence of a well-functioning clockwork implies a well-functioning clock. But to claim that a well-functioning brain causes proper cognitive functions is to presuppose determinism, which is precisely what is at issue. The point in the free will debate is the presence of human agency. A well-functioning brain is only a necessary condition for a well-functioning mind. It is the mind, which makes humans symbolic creatures.

> Actions are events that only exist in virtue of the fact that their *agent* has been the source of some input into the world. (Steward 2011: 144; italics in original; List 2014)

In terms of the distinction between reasons and causes, then, although reasons have a causal component they are not wholly determined by a causal chain, which stretches back to the agent's past.

The libertarian usually takes the view that 'agency refutes determinism' (Steward 2011). If agency plays a central part in our decision-making what relationship exists between the mind and the brain? If the brain is a biological organ and the mind is a symbolic operator, the question of Laplacean determinism is of secondary importance. The brain is an organic convolute, governed by stochastic processes. On the other hand, evidence strongly suggests that the mind and the brain are linked. The central issue, then, is the relationship between the brain and the mind. The free-will and mind-body debates should focus on this distinction. One interesting proposal is that the notion of emergence can serve as a link between brain states and mental states. Emergence will create room for the human mind, which is responsible for agency.

12.2 Emergence

The recognition of the importance of human agency and the distinction between causes and reasons can be upheld if mental processes are depicted as emerging properties of brain processes.

Emergence is usually understood as a higher-level phenomenon, which arises from a base, but acquires properties through the emerging process, which will no longer be reducible to the base from which they arose. The emergent phenomenon requires a separate level of description and is often governed by separate laws. The emergent property constitutes a qualitatively novel phenomenon. A water molecule consists of H_2O (2 hydrogen and 1 oxygen molecules) but water possesses a property—liquidity—which it does not share with its individual molecules. Electrons are subatomic particles, which have mass and other properties. They can travel near the speed of light. But electricity is an emergent novel property, which makes light bulb glow and radiators heat. An individual photon from the sun would not warm up the surface of the Earth and yet life on Earth is sustained by the streams of photons which arrive in abundance on its surface. Can this notion of emergence be employed to do work in the mind-brain debate?

A distinction can be made between strong and weak emergence. Weak emergence is also known as supervenience. There are many notions of supervenience (see McLaughlin/Bennett 2011). The notion of supervenience is often used in discussions of the philosophy of mind[1] (Kim 1993, 2000; Dennett 1992; Chalmers 1998).

> A set of properties *A* (e.g., mental properties) supervenes on a set of properties *B* (e.g., neural properties) if it is not the case that two objects can differ with respect to *A* properties without also differing with respect to *B* properties.

There is a co-variation between *A* and *B* such that 'there cannot be an A-*difference* without a B-*difference*' (McLaughlin/Bennett 2011). However, it is possible for there to be a *B*-difference without an *A*-difference. This latter case is called 'multiple realizability'. Two things can be the same at the supervenient level, without being identical at the base. For instance, two people can entertain identical thoughts or desires (*A* property) although they are physically distinct. This multiple realizability blocks a simple reduction of the supervenient level to the subvenient base. If a materialist explanation of the mind is sought, it is tempting to understand supervenience as a physical relation. Such a relation may be called *natural supervenience*. It would mean that the physical facts and laws of our brain structures entail the mental facts. But how the brain could 'entail' the mind is still a matter of debate (see Part III, Sect. 17.1.2). Natural supervenience requires (a) a co-variation of the properties of one domain, the physical base (the brain), with a supervenient domain (the mind) and (b) the dependence of the supervenient domain (mental states) on the base domain (brain states). The base constrains the supervenient domain (Rueger 2000). That is, any change in the physical base engenders a change in the supervenient domain. There are many examples of supervenience in a physical sense. For instance, the perception of colour in biological organisms is supervenient on the reception of different electro-magnetic waves. The experience of temperature is supervenient on the mean kinetic energy of air molecules. In a similar way, the mental may be said to supervene on the physical. An asymmetric dependence exists between the mental and the physical in the sense that any variation in the physical base brings about a variation in the mental domain, but not vice versa. Coincidence in the physical entails coincidence in the mental but a supervenient property may have alternative subvenient bases (Kim 1993: Essay 4). Changing the physical base—damage to the brain—will change the mental contents of any entity, which has that base. But changing the contents need not change the base. As above voting example shows, the same reason does not necessarily lead to the same behaviour. A voter may change her/his mind about which party to support. Even if the mental supervenes on the physical, an explanation of why this co-variation exists is still needed. A materialist explanation will seek to explain the supervenience relation in physical terms, in which case the mental will be reducible, in explanatory terms, to the physical base.

---

[1] The following section is based on material in my article 'Emergent Minds' (2009).

Supervenience (or weak emergence) may be an argument in favour of neural determinism—a theory which seeks to reduce the 'mind' to the 'brain' (Part III, Sect. 17.1).

By contrast *strong emergence* asserts that mental functions cannot be reduced to brain functions; hence mental functions truly emerge from neural functions as novel phenomena. Their base lies in brain functions but once they have emerged at a higher complex level they have acquired novel properties. As novel mental properties they require description in a different terminology. Whilst it will be correct to describe brain functions in terms of causal language (causes), mental functions must be described in terms of symbolic language (reasons). If mental functions are emergent properties in the strong sense of emergence, they retain a material base in neural activities but their 'freedom' derives from reasons, which agency requires.

In his *Descent of Man* (1871), Darwin explains the emergence of higher mental and moral functions as a result of natural selection amongst human groups.

> Judging from all we know of man and the lower animals, there has always been sufficient variability in their intellectual and moral faculties, for a steady advance through natural selection. (Darwin 1871: 168–169)

Darwin locates the roots of moral and social faculties in 'social instincts', which—with the increase of experience and reason in humans—are converted into cultural values. He argues that the cultivation of intellectual and moral capacities in one tribe—as against a neglect of such attributes in a rival tribe—will bestow an evolutionary advantage on the cooperative tribe.

> We can see, that in the rudest state of society, the individuals who were the most sagacious, who invented and used the best weapons or traps, and who were best able to defend themselves, would rear the greatest number of off-spring. (Darwin 1871:153)

Darwinism has become a research programme, which not only aims to explain the diversity of species but also how the 'brain could cause the mind'; how brain processes could 'cause' mental processes. The explanation had to remain tentative as long as Darwinians could not specify the 'causal' relationship between the brain and the mind. The Darwinians did not have at their disposal modern notions of emergence and embodied minds. But their efforts to construct a materialist theory of the mind, within an evolutionary context, paved the way for modern evolutionary, materialist approaches to the puzzle of mental phenomena (see Part III, Sect. 17.1.2).

What is of interest in the present context is that intellectual faculties produce cultural products, like the moral values—associated with freedom of the will—, which transcend their individual bearers. At first Darwin displays a strong tendency to attribute the spreading of civilized values to the operation of natural selection:

> All that we know about savages […] shew that from the remotest times successful tribes have supplanted other tribes. […] At the present day civilised nations are everywhere supplanting barbarous nations […]; and they succeed mainly, though not exclusively, through their arts, which are the products of the intellect. It is, therefore, highly probable that with mankind the intellectual faculties have been mainly and gradually perfected through natural selection […]. (Darwin 1871: 153; cf. Rosenberg 2005)

Although this quote reveals a disturbing hint of racism, Darwin later qualifies this statement with respect to ‘civilized nations’:

> With civilised nations, as far as an advanced standard of morality, and an increased number of fairly good men are concerned, natural selection apparently effects but little; though the fundamental social instincts were originally thus gained. (Darwin 1871: 163, cf. 80)

With this qualification Darwin finds himself in the company with T.S. Huxley who strongly emphasized that amongst human societies, natural selection had been replaced by cultural selection. Human values spread by cultural rather than natural selection. The development of civilization consists in the gradual deflection of the forces of natural selection. Huxley compares civilized life to a horticultural process, in which human gardeners modify their living conditions by deliberate choice.

> Laws and morality are restraints on the ‘struggle for existence between men in society.’ (Huxley 1894: 30)
>
> Society, then, differs from nature in having a ‘definite moral object’. (Huxley 1888: 202)

Thus it may be concluded that, on the Darwinian view, certain products of the intellectual faculties—cultural and scientific ideas as well as moral values—are emergent properties of our brain functions. They are qualitatively novel and are governed by cultural selection. But ultimately, the Darwinian challenge remains unanswered to the present day. As we shall see, the completion of the Darwinian programme of the emergence of the mind is still incomplete today. How do you explain the emergence of the immaterial (feelings and thoughts) from the biochemical activity of neurons in the architecture of the brain? The liquidity of water is an emergent property of the interaction of different molecules, where this emergent property can be understood by the operation of physical and chemical mechanisms working on the components. The Darwinian thesis of emergence has to explain how mental products, like novel ideas, may arise from the interaction of neural networks. The thesis entails the emergence of the immaterial from the material. What has been left open is the question of the physical correlate of entailment: how material processes in the brain can produce immaterial processes in the mind. But as long as no physical correlate can be found, notions like supervenience and emergence remain useful analogies. It is at this stage still unclear whether conceptions like cause, emergence, entailment or supervenience can provide a physical bridge to close the gap between the mental and the physical.

Despite this cautious assessment the history of science tells us that a good story is better than no story! Greek geocentrism remained a good story for some 2000 years until eventually it was replaced by heliocentrism. The geocentric model provided understanding, even though it fell short of a reliable explanation. Notions like emergence and supervenience also provide understanding. The main challenge remains to develop a model of the brain-mind relationship, which explains how appropriate physical action in the brain leads to the emergence of immaterial mental processes; how causes and reason interact to bring forth the mind and its ‘free will’. Two such proposals—one in terms of quantum physics, one in terms of evolutionary biology—will be reviewed in one of the following Chapters (Part III, Sect. 17.1).

Chapter 13

What Laplace's Demon Tells Us and Does not Tell Us About the World

This is a most important result of Boltzmann's investigation: the question of the direction of time as a whole must be separated from the question of the time direction observable to us.
Reichenbach, *The Direction of Time* (1956: 133)

Laplace's Demon is a denizen of a deterministic world, of a clockwork universe. He is a determinist not a fatalist. But Laplace's Demon vacillates between scientific and metaphysical determinism, between predictive and ontological determinism. His causal determinism seems to lead to the static Block Universe view, in which past, present and future are equally real. The t-invariance of fundamental laws seems to confirm this interpretation.

Laplace's Demon serves as an image of classical mechanics, which itself is not completely deterministic. The Special theory of relativity, as the extension of classical mechanics, is also only partially deterministic. In other words, neither classical mechanics nor the Special theory are deterministic in the Laplacean sense. Neither theory implies fatalism because both make room for indeterminism of at least the initial conditions. A change in initial conditions, under the same deterministic laws, can lead to branching futures. Hence the questions of free will and arrows of time arise. The Demon has failed to convince us that the universe is completely deterministic, in the Laplacean sense. Rather than being a static Block Universe, the universe, even according to causal determinism, may be read as suggesting the image of an evolving four-dimensional world. Even a deterministic world, due to its dynamical evolution, may display an arrow of time.

In the debate about the freedom of choice various versions of determinism have emerged; it is in fact only hard determinism, which denies the binary choice, characteristic of human agency. Compatibilists and libertarians can appeal to the distinction between cause and reason to argue in favour of free will. But this notion of free will does not break the circle of causal closure, because of the notion of emergence.

Scientific theories have philosophical consequences—like determinism—but they do not deductively follow from the principles of these theories. Determinism does not necessarily imply predictability, and indeterminism does not necessarily

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_13

imply non-computability. Non-Laplacean determinism is compatible with alternative histories. Determinism of this kind is compatible with arrows of time. Whilst Laplacean determinism poses a threat to the question of free will, other forms of determinism do not. However we found that the question of determinism is not as important for the issue of free will as the brain-mind distinction. It requires a reflection on the difference between causes and reasons.

In an essay on 'Science and Free Will', J.C. Maxwell drew a distinction between dynamical and statistical kind of knowledge:

> The discussion of statistical matter is within the province of human reason, and valid consequences may be deduced from it by legitimate methods; but there are certain peculiarities in the very form of the results, which indicate that they belong to a different department of knowledge from the domain of exact science. They are not symmetrical functions of time. It makes all the difference in the world whether we suppose the inquiry to be historical or prophetical – whether our object is to deduce the past state or the future state of things from the present state. (Maxwell 1873)

Not only does the statistical method give rise to the direction of time, in Maxwell's view it also saves the freedom of the will.

If determinism has its limits, its counterpart—indeterminism—deserves a fair hearing. It introduces a new demon: **Maxwell's Demon**. The entry of Maxwell's Demon helps us to introduce a distinction between local and cosmic arrows of time and to revisit the question of the nature of the mind.

Part III
Maxwell's Demon

Just as Laplace's Demon would have no use for probability theory, Maxwell's Demon would have no use for the science of thermodynamics.

Myrvold, 'Probabilities in Statistical Mechanics' (*PhilSci Archive* 2012: 26)

Laplace's Demon, despite his limitations, served as a useful guide to the conceptual ramifications and metaphysical commitments of classical physics. There are limits to what can be claimed in his name. And Laplace's Demon faces opposition from other demons, for instance Maxwell's Demon. Officially, Maxwell's Demon is employed to probe the status of the Second law of thermodynamics. But as in the case of Laplace's Demon, Maxwell's Demon, as we shall see now, serves a similar function in the philosophical investigation of the conceptual landscape of science. The notions, which come to the fore, are causality and indeterminism, the direction of time, emergence and the nature of the mind. We shall also make the acquaintance of a new demon: Loschmidt's Demon.

Chapter 14

Local and Cosmic Arrows of Time

Time in physics means Astronomer's Royal de facto time.
Eddington, *Nature of the Physical World* (1932: Chap. III)

As we saw in the previous Part, Laplace's Demon does not hold an unambiguous sway. The Demon's determinism is an extreme idealization. The theories of physics, which the Demon illustrates, are not strictly deterministic. There are lapses in the iron grip, which the Laplacean Demon supposedly holds over the material world. These uncertainties not only give hope to defenders of free will and believers in the mind. They make determinism compatible with the anisotropy of time: the past-future asymmetry, the passage of time and its arrows.

The question of the arrows of time has a time-honoured tradition, which dates back to the Greeks. Plato, as mentioned before, called the planets 'the instruments of time'. He identified time with 'the moving image of eternity.' The planets's motions were regular and unchanging since, according to the Greek worldview, they performed a perfect circular motion around the centre. Aristotle did not agree with Plato's identification of the passage of time with planetary motions. Motion is measured with respect to time but never time with respect to motion, since motion can be fast, slow or turn into rest, and time is still measured. Nevertheless Aristotle agreed that the motion of the planets constituted an ideal *criterion* for the anisotropy of time. The Greeks therefore used a physical feature—circular planetary motion—either to define the passage of time (Plato) or to identify a criterion (Aristotle), which served as a basis for the measurement of time. But then Nicolaus Copernicus introduced heliocentrism (1543). The Copernicans taught the Renaissance world that planets do not move around a central stationary Earth. Rather, the Earth, like the other planets, moves around the sun in elliptical and somewhat irregular orbits. What this story illustrates is that a characterization of the anisotropy of time needs to rely on certain physical criteria, which themselves are subject to change.

The Greeks, who believed in a geocentric worldview, correctly saw that for the measurement of time, two conditions are essential. To measure the passage of time, some regular process is needed, preferably a periodic regular process, because such a process allows the marking of finite intervals of time. The measurement of time must also be invariant, which means that a change in place and time must return the

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_14

same values for the measurement of a chosen temporal period. For the Greeks the Earth stood still at the centre of the universe. It did not undergo either a daily rotation on its own axis or an annual rotation around the sun. By contrast, according to the Greeks, the sun and all the other planets orbited the Earth in perfect circular orbits. Despite this mistaken assumption, Greek calculations of planetary orbits came close to modern values. Copernicus placed the planets in their correct order. These planetary data have remained (relatively) constant (invariant), despite the separation in space and time. The planetary motions, perfectly circular and ever-lasting, allowed the Greeks to infer the march of time. The anisotropy of time can be understood in a dual sense. It may refer to the subjective impression of the 'flow' of time, or *phenomenal* time. It may also refer to the objective passage of time or *physical* time. In the latter case it either indicates the lapse of time in our local cosmic neighbourhoods or a global motion in time of the whole universe. For conceptual convenience the local passing of time will be referred to as *local arrows* of time and the term *cosmic arrow* of time will be reserved for the global passage in time of the whole universe. In accordance with this distinction it becomes clear that the Greeks took the motion of the planets to designate the *passage* of time on Earth. It was a local arrow based on observable changes in the local cosmic neighbourhood. But would the Greeks have recognized a global, cosmic arrow of time of the whole universe? The answer is 'no' since for the Greeks the cosmos was eternal and perfectly symmetric. The planets drew their eternal circular orbits in a never-changing universe. As the Greeks would have agreed, where there is no physical change there is no passage of time. But change, according to the Greek worldview, only occurred in the decay and renewal, observable in local vicinities; essentially in the region between the Earth and the moon.

This distinction between the *local* and the *cosmic* arrows of time is also justified by modern considerations. Imagine living in a universe, which started in a Big Bang, reached a point of maximal expansion and from this point onward began to re-contract and eventually collapsed into a Big Crunch. For the purpose of the illustration, the physical conditions of the Big Crunch are set to be identical to the Big Bang. It is a Gold universe (Part II, Chap. 10). Compared to the lifetime of such a re-collapsing universe, the lifespan of each individual inhabitant would be very short. For the inhabitants of such a 'closed' universe the march of time, as judged by their experience, would always point in the future time direction. But if they lived as long as their universe they would eventually return to where it began. Whilst for the short-lived inhabitants time would pass 'normally' in one direction, for a cosmic demon, who could observe the whole universe return to its original state, there would be no cosmic arrow of time. Hence there is a need to distinguish local and global arrows of time (Fig. 14.1).

There are in fact many local arrows of time. One is the *psychological* arrow, namely the subjective feeling of the unidirectional flow of time from the past to the future. This is also known as phenomenal time. Humans remember the past and anticipate the future. The past is regarded as fixed, unchangeable whilst the future is still open. The past is frozen actuality, the future is open potentiality. It should be noted that our psychological sense of time is not necessarily identical with clock

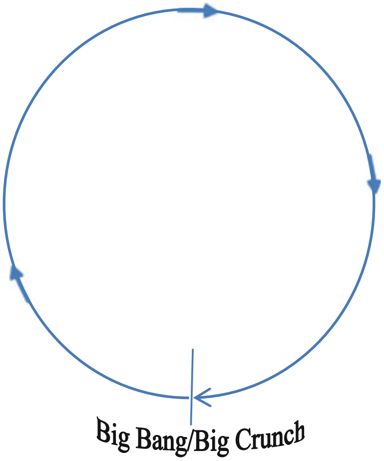


Fig. 14.1 A closed universe performs a circle from a beginning in the Big Bang to a collapse into a Big Crunch

time. Both age and drugs can affect the way individuals perceive time. It is well known that for young people time passes slowly whilst for old people it ‘flies by’.

Irrespective of our psychological feelings about the ‘flow’ of time, we are surrounded by many irreversible physical processes. Most people are born young and die old. Hot liquids grow cold. Apples rot. Unattended buildings fall into disrepair. Our experience never seems to show the reverse processes. Such irreversible processes constitute an objective *physical* arrow of time. These irreversible processes seem to describe a transition from more to less orderly states since they are based on the thermodynamic notion of entropy. It is also known as the *entropic* or thermodynamic arrow, indicating a transition from order to disorder. Other physical arrows are less familiar to us and yet constitute irreversible processes, which could serve as criteria for the anisotropy of time. One is the measurement process in quantum mechanics, according to which the state of quantum-mechanical systems is reduced to observable macroscopic results upon measurement.

The observation of familiar ripples and waves point to the existence of a *radiative* arrow. When a stone is thrown into the middle of a lake ripples diverge from the centre to its edge. But no-one has ever reported ripples coming from the edge of the lake and converging towards the centre to scoop up a stone from the bottom of the lake. A source emits radiation in an outward direction. Waves and radiation never seem to converge from scattered directions onto a concentrated source. Hence there is an asymmetry between diverging and converging waves. The origin of this asymmetry is controversial (see Price 2006 for an overview). One suggestion is that the radiative arrow is linked to the thermodynamic arrow. It is the energy stored in the environment, which is responsible for the preponderance of outgoing waves. The energy of the environment is not at equilibrium and hence has a lower entropy than a future final condition. The environment of the present universe enjoys large energy differences, which in turn are due to the entropic history of the universe, starting in a low-entropy Big Bang.

We are familiar with the *causal* arrow of time, i.e. the fact that seemingly effects always follow their causes but never the other way round. Immanuel Kant, for instance, explicitly argued that an objective rule of causality was needed to establish the succession of events in time. The law of causality has a temporal dimension, since for an effect to be caused, the antecedent causal conditions must be prior to the consequent effect conditions. This situation even obtains when cause and effect occur simultaneously, as in Kant's illustration of a ball, which presses a dent in a cushion. In his well-known example a lead ball is placed on a cushion, which causes a hollow in it for as long as the ball is kept there. In Kant's example cause and effect are simultaneous:

> If I view as a cause a ball which impresses a hollow as it lies on a stuffed cushion, the cause is simultaneous with the effect. But I still distinguish the two through the time-relation of their dynamical connection. For if I lay the ball on the cushion, a hollow follows upon the previous flat smooth shape; but if (for any reason) there previously exists a hollow in the cushion, a leaden ball does not follow up it. (Kant 1787: B248–B249)

Then, there is a *historical* arrow of time, which indicates biological and geological evolution. (Human history has suffered too many setbacks to discern a clear sense of progress towards, say, an enlightened society, as the Enlightenment philosophers confidently anticipated.) The geological arrow is related to the biological arrow in that the layers of fossil records reveal the evolution of life. Unlike his predecessor, Jean Baptiste Lamarck, Charles Darwin does not stipulate that evolution pursues a target. Lamarck believed in a necessary progression from primitive micro-organisms to complex macro-organisms. The aim of evolution was to give rise to human beings as the most complex creatures in the biological realm. They were the crown of creation. Darwin does not teach, unlike Lamarck, that evolution is a necessary progression from lower to higher forms of life. Darwin describes evolution as a contingent process in the colonization of ecological niches. This occupation of ecological niches by different organisms may produce very complex organs, like the eye. But Darwinism makes no room for a one-directional arrow from simplicity to complexity of organic forms.

All these local arrows of time must be distinguished from the *cosmic* arrow of time. A cosmic arrow of time cannot simply be inferred from local arrows, since the passage of time is compatible with a universe, which lacks a global arrow (Fig. 14.1). One criterion for the global arrow of time is the expansion of the universe. For this cosmic arrow to exist the initial conditions, at the Big Bang, must be physically different from the final conditions in what is sometimes called the Big Chill (or the heat death in the language of the 19th century). A universe, which returns to its initial condition after a period of expansion, does not possess a cosmic arrow of time. But if the initial and final conditions of the evolution of the universe differ, the universe displays such an arrow. This condition also obtains for local arrows of time.

The current Part on Maxwell's Demon will be concerned with the many manifestations of local arrows of time. In the next Part, on Nietzsche's Demon, the question of the cosmic arrow of time will be addressed. In this connection many questions arise. How are all these arrows of time related? Is the thermodynamic arrow more fundamental than the other arrows? How is the cosmic arrow of time related to the thermodynamic arrow of time? If the universe expands to a maximal point and then begins to re-contract does this mean that time itself reverses? It seems at first that the thermodynamic arrow differs from the other arrows, which appear to be involved in the generation of order. The entropic or thermodynamic arrow, however, seems to lead to the destruction of information and the loss of order.

The notion of entropy therefore was destined to play a special role in considerations of the arrows of time. Both the Austrian physicist Ludwig Boltzmann and the British astronomer Arthur Eddington had a tendency to identify the arrow of time with increasing entropy, according to the Second law of thermodynamics. But both also had reservations about this identification. Although Boltzmann characterized the Second law as a ‘steady gradation of energy’, he did not think that the motion towards a ‘heat death’ applied to the whole universe (Boltzmann 1886: 19, 1898: Sect. 89). In fact he assumed that the whole universe existed in a state of equilibrium, with individual pockets of disequilibrium, where life may evolve. Under this assumption ‘one can understand the validity of the second law and the heat death of each individual world without invoking a unidirectional change of the entire universe from a definite initial state to a final state’ (Boltzmann 1897: 242). Boltzmann, then, accepted local arrows but no global arrow of time, since he saw the whole universe as existing in a state of equilibrium.

Arthur Eddington’s thinking, too, underwent a significant development. In his early essays on the theory of relativity he embraced the notion of a static Block Universe. The physical universe had no temporal dimension (Weinert 2004: Chap. 4.2). The passage of time was a mental not a physical phenomenon. Whilst there was phenomenal time, there was no physical time. But in his later work he argued that in order to ‘express the one-way property of time’, an arrow of time needed to be added to the representation of the four-dimensional world. The Special theory of relativity encourages the representation of the physical world as a map of events, without temporal flow. Eddington came to be dissatisfied with this characterization. In order to invest the physical world with dynamic properties, he assigned the Second law a special place amongst the otherwise time-reversal invariant fundamental laws of physics. The increase in randomness or entropy would allow us to distinguish the past from the future. In *The Nature of the Physical World* Eddington grants the Second law an apparently unassailable, exceptionless ‘supreme position amongst the laws of Nature’ (Eddington 1932: 74) and adds that as ‘far as physics is concerned time’s arrow is a property of entropy alone’ (Eddington 1932: 80). With the introduction of entropy into the physical worldview a transition from a static to a dynamic view had been made possible (Eddington 1932: 110). But in his later work, Eddington expressed misgivings about the identification of increasing entropy with the arrow of time. In his book *New Pathways in Science* (1935) he introduced further refinements to these ideas. He acknowledges that the Second law is a statistical law, but does not contemplate that it may lose its supreme position. The increase in entropy now becomes a ‘signpost’ for the arrow of time. Eddington explicitly moves away from his earlier view, which seems to suggest that time's arrow should be identified with the Second law. Although the Second law is a statistical law, he continues to regard entropy as a 'unique local signpost of time' (Eddington 1935: 68–71). But at thermodynamic equilibrium, 'our signpost for time disappears and time ceases to go on'; nevertheless it continues to exist and like space, it extends since, for instance, atoms still vibrate (Eddington 1932: 79). Like the Greeks, Eddington moves from a Platonian to an Aristotelian position: although the Second law is statistical in nature and should not be used to identify the arrow of time, it is nevertheless a useful criterion for the anisotropy of time. Eddington also makes a distinction between local and the global arrows of time. Apart from entropy increase Eddington offers a second argument for the cosmic arrow of time: the expansion of the universe.

Why did both Boltzmann and Eddington move from an identification of the arrow of time with the increase in entropy to the latter's use as a criterion for the anisotropy of time? The reason is that the Second law had been turned from a deterministic into a statistical law. This switch in understanding was due to the interference of a demon: Maxwell's Demon.

Chapter 15
Maxwell's Demon

> *Maxwell's immortal Demon proved that Victorian whimsy could relieve some of the gloom of the Germanic Heat Death.*
> Brush, 'Irreversibility and Indeterminism: Fourier to Heisenberg', *Journal of the History of Ideas* 37 (1976: 603–30)

In a famous thought experiment, involving 'a being with superior faculties', James Clerk Maxwell attempted to show that the Second law of thermodynamics only possessed statistical validity (Maxwell 1875: 328–329; cf. Earman/Norton 1998, 1999; Leff/Rex 2003; Hemmo/Shenker 2010, 2012; Daub 1970; Reichenbach 1956; Zeh 1992: Chap. 3). This being, later dubbed 'Maxwell's Demon', is able to 'follow every molecule in its course'. In an appropriate setup such a being would be able, says Maxwell, to sort the molecules according to their respective velocities. The setup is simply a container, divided into two chambers by a partition, in which there is an opening (Fig. 15.1). The Demon's only work involves the opening and closing of the hole 'so as to allow only the swifter molecules to pass from A to B, and only the slower ones to pass from B to A.' Maxwell concludes:

> He will thus, without expenditure of work, raise the temperature of B and lower that of A, in contradiction to the second law of thermodynamics.

If it can be broken, the Second law cannot enjoy deterministic validity. It must be a statistical law, which holds with overwhelming probability for systems composed of many particles. It is a law for typical cases, without forbidding atypical cases. It follows that findings, which apply to the macroscopic level, may not apply to the microscopic level:

> This is only one of the instances, in which conclusions which we have drawn from our experience of bodies consisting of an immense number of molecules may be found not to be applicable to the more delicate observations and experiments which we may suppose made by one who can perceive and handle the individual molecules which we deal with only in large classes.

Following this insight, Maxwell spells out the classical view of probability, which results in knowledge of averages over micro-states (of molecules) but leaves us ignorant about their individual properties (position, momentum):

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_15

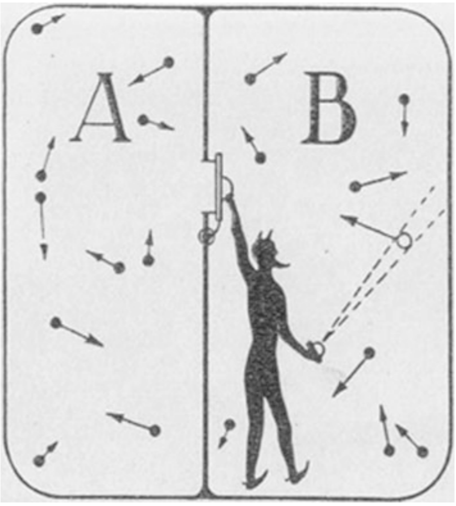


Fig. 15.1 Maxwell's demon at work. *Source:* Wikimedia Commons

> In dealing with masses of matter, while we do not perceive the individual molecules, we are compelled to adopt what I have described as the statistical method of calculation, and to abandon the strict dynamical method, in which we follow every motion by the calculus. (Maxwell 1875: 329)

Since the creation of Maxwell's Demon there have been many attempts to show that the Demon cannot achieve his aim; that the Demon himself is subject to the Second law.[1] If devices, built according to the Demon's instructions, were possible

> (m)achines of all kinds could be operated without batteries, fuel tanks or power cords. For example, the demon would enable one to run a steam engine continuously without fuel, by keeping the engine's boiler perpetually hot and its condenser perpetually cold. (Bennett 1987: 88)

But the Demon or his physical manifestation cannot escape the effects of the Second law for the Demon, by carrying out his task, would heat up. In order to sort the molecules, he must be in contact with them and thus is affected by their thermal motions. The Demon would absorb more heat from the molecules than he can expand. In other words, he would warm up. He would begin to shake from the Brownian motion of the molecules inside his body, which would make him unfit to perform his task. As physicist Richard Feynman points out, Maxwell's intelligent Demon can be replaced by simpler devices.

---

[1] As Earman and Norton (1998, 1999) discuss, different scenarios can be envisaged: whether the demon is an intelligent being or a physical system does not, however, affect his ultimate failure. Cf. Szilard (1983); von Beyer (1998); Leff/Rex (2003); Maloney (2009). But see Hemmo/Shenker (2012) for a defense of Maxwell's Demon.

> It turns out, if we build a finite-sized demon, that the demon himself gets so warm that he cannot see very well after a while. The simplest demon, as an example, would be a trap door held over the hole by a spring. A fast molecule comes through, because it is able to lift the trap door. The slow molecule cannot get through, and bounces back. But this thing is nothing but our ratchet and pawl in another form, and ultimately the mechanism will heat up. If we assume that the specific heat of the demon is not infinite, it must heat up. It has but a finite number of internal gears and wheels, so it cannot get rid of the extra heat that it gets from observing the molecules. Soon it is shaking from Brownian motion so much that it cannot tell whether it is coming or going, much less whether the molecules are coming or going, so it does not work. (Feynman 1963: §46.3; Zeh 1992: §3.3)

The Maxwellian Demon must be able to acquire information about the physical state of the individual molecules: whether they are fast enough to justify the opening of the trapdoor or whether they should be confined, on account of their slowness, to chamber A. In either case—whether an intelligent demon or a simple mechanical device is used—the Maxwellian Demon illustrates two traditional understandings of the Second law of thermodynamics.

The Second law can be understood either in terms of a loss of order or a loss of information. Before the Demon sets to work to separate the slower and faster molecules into chambers A and B, respectively, all the molecules are randomly distributed across the whole container. In an intuitive sense, this distribution is less orderly than the neat separation the Demon aims to achieve. The random order the Demon faces could be reduced to a more orderly distribution if the Demon achieved his task. Such a change from a random to an ordered distribution of the molecules would also increase the information of the whereabouts of the molecules. Order and information, as well as disorder and disinformation are linked (see Caticha 2014). Before the Demon sets to work information about the location of the fast and slow molecules is uncertain: any particular molecule could be in chamber A or B, respectively. By restoring order the Demon also increases the information content: if the Demon succeeds all the slow molecules will be confined to chamber A, and all the fast molecules to chamber B. If the Demon is then asked where the molecules are the answer will be unequivocal. However, imagine the Demon takes a rest after his gargantuan task but forgets to close the opening. The order of the molecules—fast ones in A, slow ones in B—will be destroyed, which leads to information loss. The uncertainty of the information—or the relative unpredictability of the whereabouts of the molecules—can be linked to an information-theoretic notion of entropy. The less information is available about the outcome the greater is the information entropy. The Demon reduces the information entropy of his set-up by performing his entropy-defying task; as soon as the Demon neglects his duty, he increases the information entropy.

In addition to these two traditional meanings of entropy—in terms of disorder and information loss—this Chapter will introduce a modern understanding of entropy, in terms of the occupation of phase space. For the moment let us allow the Maxwellian Demon to rest and ask whether the trajectories of the individual molecules and their ensemble are irreversible or reversible. This requires the introduction of a new demon.

Chapter 16
Loschmidt's Demon: Reversibility and Irreversibility

> *If seconds were counted without clock, according to a certain feeling, the big question is whether these seconds would all be equal at one time or another.*
> Lichtenberg, *Sudelbücher* II, Goldpapierheft: §34; translated by the author

What would happen to a volume of gas, which was left to expand in a sealed container without the interference of a demon? If a perfume bottle was placed in a sealed container and then opened, the molecules would spread throughout the available volume (Fig. 16.1). We would not expect the molecules to return to the bottle in a finite amount of time. This expectation is based on our experience: a cup of hot coffee will eventually cool but it will not, as far as we know, spontaneously reheat; a fresh apple left on a table will eventually rot but we do not expect it to return to its fresh state; a person is born young but we do not expect an old person to return to youth; we expect ripples to expand outwards from the centre of a lake but we do not count on converging waves from the edge of the lake towards the centre. The passage from order to randomness, from high energy to low energy, is an irreversible process and yet it is not forbidden by the laws of nature. Each individual molecule in the large volume of gas follows deterministic laws. In theory it has as much chance of wandering away from the bottle as it has to return to it. Even though it does not happen in practice, it is allowed in theory. Either a demon would have to reverse the molecules or it would take an infinite amount of time for this return to happen. Physicists have estimated the amount of time it would take for a volume of gas, containing 10^{18} molecules, to return to its initial state (position and momentum variables). It is assumed that each molecule (with an average molecular velocity of 5×10^4 cm/s in both directions) would return to within 10^{-7} cm of each initial position variable and within 10^2 cm/s of each velocity variable. The estimated time for a return to such a configuration would require $10^{10^{19}}$ years, which is well beyond the estimated age of the universe ($\sim 10^9$ years) (see Schlegel 1968: 52–53). According to some recent estimates a return of all the particles in a two-chamber system to just one chamber has a probability of $10^{-6x10^{22}}$ and the

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_16

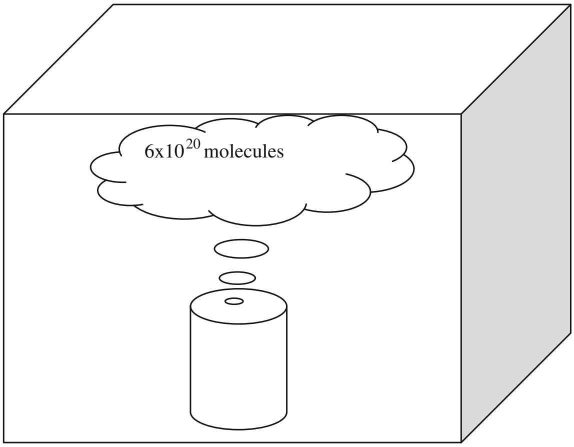


Fig. 16.1 Expanding molecule cloud in a three-dimensional coordinate space towards a less ordered state

mean time <T> for such an occurrence is of the order of $10^{6x10^{22}}$ s, the estimated age of the universe (see D'Abramo 2012).

Within the finite amount of time in our experience we are justified, at least on practical grounds, to generally expect transitions from smooth and orderly to less orderly and fibrillated states (Fig. 16.2).

It is not humanly possible to reverse all the motions of the gas molecules from a larger space, like the sealed container, to a smaller space, like the perfume bottle. Nor can humans wait for the end of time to experience the return of gas molecules.

But the service of a demon could be employed. Let this new demon, dubbed **Loschmidt's Demon**,[1] be the opposite of Laplace's Demon. Laplace's Demon was able to predict, in theory, all the motions of all the particles in the universe from the present to both their past and future states. But Laplace's Demon is a passive observer of a 'deterministic' world. Maxwell's Demon is a mere sorting demon, a restorer of order. Loschmidt's Demon is an active demon: he is able to destroy the fibrillated state and reverse all the velocities of all the particles to a less fibrillated, a smoother state. How could the Demon achieve this feat? In Part II, Chap. 10 the time reversal of fundamental laws was introduced. What is needed is the reversal operator **R**, which reverses the instantaneous state of a system, characterized by its position, p, and its momentum, m: $R(p,m) = (p,-m)$. The temporal operator, **T**, converts time t into $-t$. Loschmidt's Demon would have to reverse the momenta of each and every single molecule at the final state, t_f, to return them to the initial state, t_i. Would Loschmidt's Demon achieve the reversibility of the molecules, which Nature seems to withhold from human experimenters?

---

[1] This Demon is based on J. Loschmidt's reversibility objection to Boltzmann's deterministic understanding of the Second law. The context of Loschmidt's objection is the 19th century concern about the eventual heat death of the universe. Note that Loschmidt generalizes from a sealed container to the whole universe (Loschmidt 1876: 139).

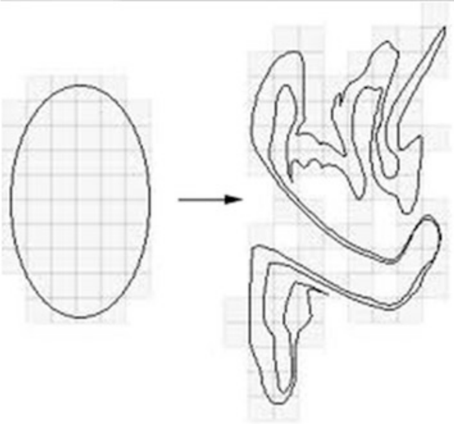


Fig. 16.2 From order to randomness as a typical evolution. The final state is more 'fibrillated' than the original state

Of course Loschmidt's Demon has to operate within the laws of the natural world. For isolated systems the Demon's task seems achievable since the fundamental laws are time-reversal invariant. If the Demon opened a perfume bottle in a sealed container, he should in principle be able to return every molecule to its original ordered state. But the larger the phase space, the more gargantuan the task.

> To construct a robot with a reversed psychological arrow of time it would be necessary to reverse the thermodynamic arrow, not only of the robot, but also of the local environment it is observing. That is possible in principle. However, since we have a system of matter coupled to electromagnetic radiation, it would be necessary to deal with every molecule and photon within a radius of 2×10^{10} km to reverse the system for a day (Hartle 2005: §IV).

But the effects of energy and entropy have not yet been taken into account. There is a theorem—Liouville's theorem—which states that if we take a volume of a bundle of trajectories and let it evolve over time, then its volume element along a flow line remains invariant although its shape may change (Fig. 16.3). This theorem, which holds in classical mechanics, has an interesting corollary. An immediate consequence of it is that even though the *volume* is preserved the *shape* of this phase space region is not preserved and this implies a dynamic evolution of the

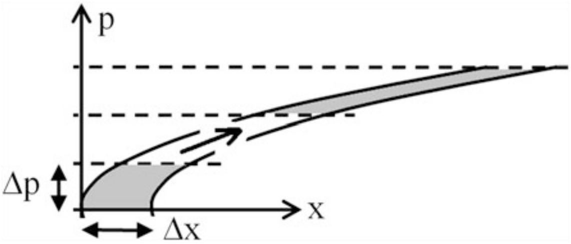


Fig. 16.3 Illustration of Liouville's theorem for the evolution of classical systems. The motion of a collection of closely packed orbits associated with free fall obeys Liouville's theorem. Though the shape area remains constant, the shape is distorted. *Source*: Wikimedia Commons (phase space free fall); cf. Zeh 1992: Chap. 1

trajectories within this region (see Earman 2002). For two shapes cannot differ from each other without an evolution of the trajectories. Such an evolution from smooth to fibrillated conditions would occur of the Demon dropped a bottle on the floor. This theorem also implies that a *reversed* evolution of the trajectories will preserve the volume but not necessarily the shape and hence that reversed trajectories need not be invariant with respect to the shape of the phase space region. Could the Demon nevertheless return the wandering molecules or broken pieces to their original ordered state? Loschmidt's Demon does not claim that the same initial conditions lead to equal numbers of entropy-increasing and entropy-decreasing evolutions. Rather, 'all Loschmidt is pointing out is that there are equal numbers of increasing-entropy and decreasing-entropy evolutions overall, when we consider every possible initial condition' (Carroll 2010: 398 Fn 141). In theory Loschmidt's Demon has the ability to reverse all the motion to exactly the same distribution from where they started. Loschmidt's Demon can secure perfect reversibility of the motions, which one would expect to follow from deterministic assumptions. But the physical world we live in is not that of Loschmidt's Demon. The Demon would need energy to flick the broken pieces back into place. If the bottle accidentally dropped from the top of the table, there would be an asymmetry in the expenditure of initial and final energy. If furthermore, in analogy with Eddington's shuffling cards, the floor receded slowly away from the top of the table, the Demon would need longer and longer to reassemble the bottle. In other words, if we go beyond classical mechanics and add further perspectives, the Demon's task becomes more difficult in practice, though not impossible in theory. Even if a system could be reversed to its initial position, it would not be a reversal of time but a return to a copy of initial conditions. (As discussed in Part IV, this objection also holds against Nietzsche's thesis of the cosmic recurrence of all events.) The Demon would destroy the cosmic arrow of time but not the local arrows because the reversal follows a sequence of events:

$$\textbf{Initial conditions}_1 \rightarrow \textbf{final conditions} \rightarrow \textbf{initial conditions}_2.$$

The difficulties of Loschmidt's Demon reflect the limits of Laplacean determinism in classical mechanics. However, if the classical world does not display Laplacean determinism, if Loschmidt's Demon may fail and even the long-term evolution of the solar system cannot be predicted, should it be assumed that indeterminism, rather than determinism, rules the universe?

Chapter 17
Indeterminism

We become developed into advocates of necessity or of free-will.
Maxwell, 'Essay on Free Will' (1873: 434)

Indeterminism is usually associated with quantum mechanics—the physics of the atom. In the simplest case, for instance, a beam of electrons, sent through the inhomogeneous magnet of a Stern-Gerlach apparatus, will split into two beams, which leaves each electron with a 0.5 probability of either being deflected upward or downward (Fig. 17.1).

As observed above, indeterminism also exists in the long-term evolution of the solar system. Although the planets do not move erratically away from their periodic orbits, their orbits do not follow the strict periodicity, which Laplace assigned to them. Planetary orbits move within a certain chaotic zone, of varying size, which is dependent on whether they are inner or outer planets. Their orbits within these zones are indeterministic but the zones themselves are stable. But determinism claims that the trajectory of given mechanical systems can be exactly determined from a given current state both into the future and the past. For short time spans determinism rules the planetary motions. For instance, calculations of Neptune's orbit show that Galileo must have perceived the planet 234 years before its discovery in 1846. But the determinism of planetary motion does not last into the distant future. Indeterminism therefore characterizes the impossibility of determining the exact spatio-temporal trajectory of both classical and atomic systems (cf. Brush 1976). As in the case of Laplace's Demon, indeterminism can be understood in both an ontological and epistemological sense. Indeterminism, however, is not to be confused with complete randomness. As discussed in Part II, Chap. 9, the probability of the occurrence of certain events on the basis of the knowledge of the current state of the system is predictable. To illustrate, consider de Broglie's (1892–1987) thought experiment. A beam of electrons is targeted at a crystal, which represents the current state of the system, *C*. The encounter of the beam with the surface of the crystal will lead to diffraction effects, E_1, E_2, E_3, which

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_17

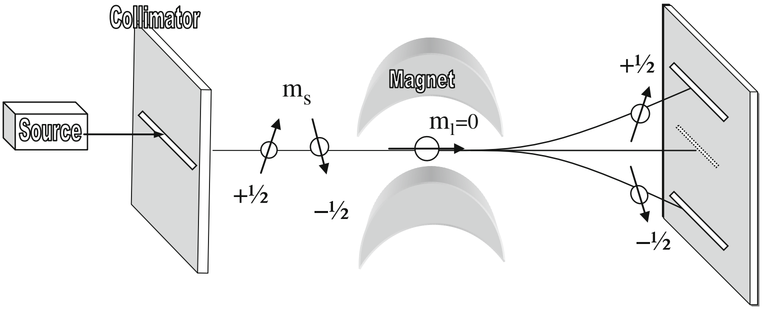


Fig. 17.1 The Stern-Gerlach Experiments (1921–1925): In the simplest case an atomic beam is split by an inhomogeneous magnet into upper (+½) and lower deflections (−½), as shown on the recording screen

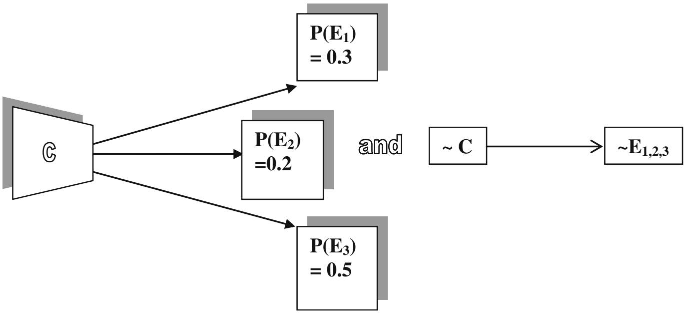


Fig. 17.2 de Broglie's thought experiment, with probability weightings for the various effects

will be recorded as scintillations at different points on a recording screen. The diagram also shows what happens when *C* is absent: If *C* does not occur in this situation, then the effects $E_{1,2,3}$ are unlikely to appear. *C* will be a necessary condition for the appearance of the effect, since in its absence the effect will be absent (Fig. 17.2).

In terms of the causal arrow of time, indeterminism therefore means that from a current state of affairs, only the probability of future events can be calculated (Fig. 11.9). Thus, it is not the case, as one scientist reacted, that

> (f)rom the state of the matter at one instant, it is impossible in principle to discover what the state will be at a future instant. (Jeans 1943: 149–150)

Indeterminism does not mean complete randomness or unpredictability. A statistical prediction of future events from a given current state is still possible. But what does this mean for retrodiction? Given certain effect events at a given time, is it possible to determine their cause, which precedes them? (Fig. 17.3).

$$\text{If} \left\{ \begin{array}{l} E_1 \ \textbf{(Probability } \boldsymbol{\alpha}\textbf{)} \\ E_2 \ \textbf{(Probability } \boldsymbol{\beta}\textbf{)} \\ E_3 \ \textbf{(Probability } \boldsymbol{\gamma}\textbf{)} \end{array} \right\} \text{then } \mathbf{C}_{1,2,3}?$$

Fig. 17.3 Instead of inferring effects from a cause, as in Fig. 11.9, probable causes are inferred from effects

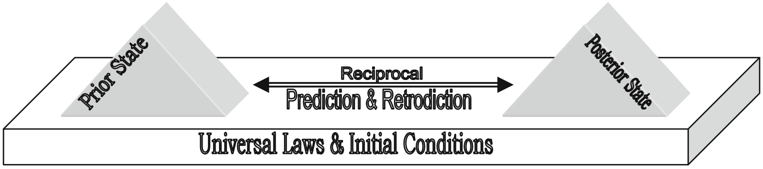


Fig. 17.4 Temporally symmetric inferences in a Laplacean deterministic system

In a Laplacean deterministic system inferences are temporally symmetric: from a prior state it is possible to predict the posterior state (effect) and vice versa (Fig. 17.4).

But in an indeterministic system, such symmetric inferences are no longer possible. From a given effect, E_1, only probable causes C_1, C_2, C_3 can be inferred.

$$\mathbf{If\ E_1, then} \left\{ \begin{array}{l} \mathbf{C_1} \ (\textbf{Probability}\, \alpha') \\ \mathbf{C_2} \ (\textbf{Probability}\, \beta') \\ \mathbf{C_3} \ (\textbf{Probability}\, \gamma') \end{array} \right\}$$

Some causes, however, are more probable than others. Consider room temperature, which according to thermal physics is a macro-state. This macro-state is made up of numerous micro-states, i.e. the individual air molecules and their properties. There are many ways, in which a particular room temperature can be achieved by the combination of its micro-states. According to classical physics, room temperature is nothing but mean kinetic energy of the air molecules, which fill up the room. At room temperature air molecules have an average speed of 2000 mph, but the velocity of individual molecules can range from 0 to 4000 mph. At a particular room temperature, say T_1, there is a distribution of molecule velocities (or micro-states), which correspond to a particular macro-state (temperature). The average room temperature can be achieved in many different configurations of the micro-states. For instance, if half the molecules are at rest and the other half all move at 4000 mph then their average velocity is 2000 mph, which corresponds to the average room temperature. However this state is extremely unlikely; it is atypical. There will be a more typical velocity distribution of the molecules (the micro-states), which is likely to result in a given room temperature. By a method of elimination—or inference to the most plausible account—the atypical cases can be eliminated in favour of the typical cases.

If α is greater than β and γ, eliminate C_2 and C_3 and keep C_1 as the most probable cause.

If the system is indeterministic, it may not be possible to identify a unique cause. The remaining cause is only more probable than the eliminated causes. Consider, again, the future evolution of the solar system within a well-defined chaotic zone. The trajectory of the planets of the solar system does not stray from the chaotic zone but within the zone their trajectory is indeterministic. Then it may not be possible to determine the exact past state of the planets from which they evolved into their future state. And if Loschmidt's Demon took it upon himself to reverse the motion of the planets from their future position to return them to their current position, he would not be able to achieve this task because of their indeterministic evolution.

In Part II we saw that determinism affects both the question of free will and the arrows of time. Does indeterminism have a similar effect? On an abstract level, indeterminism provides a basis for an arrow of time, because in an indeterministic system past and future states will differ. In fact, the past-future asymmetry is one of the fundamental experiences of human existence. But can indeterminism save 'free will'? Does it teach us something about the nature of the mind? What effect does it have on causality? It behoves us to return to these questions. Let us first address the question of free will and ask whether indeterminism, as Eddington believed, can 'emancipate the mind'.

17.1 Indeterminism and Free Will

The existence of free will is sometimes associated with the indeterminism, which governs the micro-world. Indeterminism, as observed above, involves branching histories, with different probability weightings for the permissible outcomes. According to indeterminism one particular causal event may engender different possible histories, with different degrees of probability. As observed in Part II, Chaps. 11, 12, such branching events are reminiscent of the binary character of choice in human action. However, in the Stern-Gerlach experiment (Fig. 17.1) the electrons are deflected by a magnet, which by all accounts is a physical device, whilst the freedom of human action is taken to be based on reasons, at least according to some accounts. Can indeterminism save 'free will'?

On the one hand, human beings (in terms of classical physics) are macro-objects, which are not governed by micro-physical laws. As Arthur Eddington observed long ago, 'indeterminacy of a few atoms does not guarantee free will' (Eddington 1935: 86–91; cf. Earman 1986: Chap. XII, §3). Indeterminism at the microscopic scale gets simply washed out at the macroscopic level.

On the other hand the indeterminism in the macroscopic world, which arises under certain circumstances, cannot account for free will either: it affects macro-objects, like planets and billiard balls, which are very different from the human brain. The brain is an indeterministic biological system of great complexity, in which chemical, quantum-mechanical and thermodynamic processes play their part. The brain is clearly linked to the mind, since damaged brains lead provably to damaged minds. In the opposite direction, the mind does not seem to be a mere epiphenomenon—a passive by-product of brain activity but barred from acting on the brain. Whatever the scientific credentials of psycho-analysis,[1] Freud certainly insisted on the effect of mind-stuff on body-stuff: the effect of mental events on a patient's bodily demeanour in terms of neurotic behaviour, slips of the tongue or dreams. There is no need to summon Freud as a witness. Every human's personable experience testifies to the effect, which mental processes (like intuitions and wishes) can have on physical processes (like hunger and thirst). An interaction seems to exist between mind and body or between neural and mental events. Conscious free will decisions seem to be permanently lodged in an overlapping region, where mind and body cohabit. Even though this interaction remains a mystery, it is reasonable to expect that some scientific account will play a part in its final solution. But it will be a solution about the nature of the mind and its interrelation with the brain. As argued earlier, notions like 'emergence', 'entailment', supervenience' can be elaborated into plausible stories. But such notions remain mere 'stopgaps' as long as no physical correlate for them is found. There have been several attempts in this direction. One approach comes from physics and has been proposed by the mathematical physicist Roger Penrose and the neuroscientist Stuart Hameroff. It depends on the existence of quantum coherence over large brain areas. Biological evolution supposedly provided the brain with a structure, which would enable it to produce conscious experiences. The other approach relies on biology and is due to the biologist Gerald Edelman, who introduced a theory of 'neural Darwinism'.

17.1.1 Quantum Coherence

According to the Penrose-Stuart account, consciousness is part of the universe and therefore must be amenable to scientific explanation (Penrose 1994: Chaps. 1, 4, 7). Conscious thinking cannot be reduced to an algorithm. Its explanation requires quantum oscillations in the microtubules inside brain neurons. Microtubes are hollow cylindrical tubes (polymers) of varying lengths. They are capable of sustaining 'orchestrated' quantum-coherent states. For this to hold, large areas of the

---

[1] Throughout his career Freud claimed the psycho-analysis was a science, comparable to physics, but worries of its falsifiability and the competition of alternative explanations cast a permanent shadow of doubt over Freud's confident claims (see Weinert 2009: Part III.3).

brain must find themselves in a collective quantum state, which will however be reduced, by a special mechanism (OrchOR), to discrete conscious events.

> Biology evolved a mechanism to orchestrate such events and to couple them to neuronal activity, resulting in meaningful, cognitive, conscious moments and thence to causal control of behaviour. These events are proposed specifically to be moments of quantum state reduction (intrinsic quantum 'self-measurement'). (Hameroff/Penrose 2014: 40)

Consequently consciousness is also a global phenomenon so that it is unlikely that there exists a one-to-one correspondence between mental and brain states.

> The unity of a single mind can (…) arise only if there is some form of quantum coherence extending across at least an appreciable part of the entire brain. (Penrose 1994: 372; cf. Hameroff/Penrose 2014)

Quantum coherence describes the fact that in quantum mechanics particles can be entangled in such a way that they have to be treated as a single system. In the case of the brain this single system remains unentangled with its environment. As human brains are biological systems, which have evolved and increased in capacity over millions of years, it must be assumed that such quantum coherence exist in biological systems. However, such quantum coherent states must undergo a state reduction (OR) inside the brain in order to render conscious states possible. Although this model, involving quantum mechanics, provides a physical basis for consciousness it does not address the central puzzle, namely how conscious states can emerge from quantum coherent states in the brain. Penrose accepts that

> (l)arge-scale coherence does *not*, in itself, imply consciousness, of course—otherwise superconductors would be conscious. Yet it is quite possible that such coherence would be *part* of what is needed for consciousness. (Penrose 1994: 408; italics in original)

Quantum-coherent states may provide a physical prerequisite for consciousness but by the authors' own admission quantum-mechanical processes in the brain do not tell us how consciousness actually results from quantum oscillations in the microtubules. As observed previously, physical processes are a *necessary* but not a *sufficient* condition for the emergence of consciousness. The authors claim that conscious events are 'terminations of quantum computations in brain microtubules.' But 'computation' is a logical term, whilst 'oscillation' is a physical concept. The puzzle of consciousness is, however, how the transition from quantum oscillations can give rise to conscious expressions. The authors admit that 'the mechanism by which such neuronal computations may produce conscious experience remains unknown' (Hameroff/Penrose 2014: 40). That is, in the absence of such quantum oscillations normal consciousness may not be possible but the presence of such quantum computations is not sufficient to explain the existence of conscious states.

> 'Orchestrated objective reduction ('Orch OR') is a theory which proposes that consciousness consists of a sequence of discrete events, each being a moment of 'objective reduction' (OR) of a quantum state (…), where it is taken that these quantum states exist as parts of a quantum computation carried on primarily in neuronal microtubules. Such OR events would have to be 'orchestrated' in an appropriate way (Orch OR), for genuine consciousness to arise. (Hameroff/Penrose 2014: 73)

The Penrose-Stuart account explicitly accepts that evolution has formed the human brain, although their theory relies on quantum mechanical principles in an effort to solve the brain-mind problem. An appeal to explicit Darwinian principles forms the backbone of another proposal, due to Gerald Edelman, to find the physical correlate of consciousness. But it too does not go beyond the claim that consciousness has a physical or biological basis.

17.1.2 Neural Darwinism and Emergence

The Darwinian T.S. Huxley assumed the existence of a non-linear relationship between small changes in the 'nervous system' and the vast functional changes in the states of consciousness. As noted in Part II, Sect. 12.2, it is typical of emergent properties that they constitute qualitatively novel phenomena, which are no longer reducible to the base from which they emerge. Here the base is constituted by the neural networks in the brain, which produce novel, higher-order mental functions. Mental processes become the emergent properties of interacting neural networks. The integration of neural networks and the links, which exist between areas of the human brain, means that mental functions tend to be distributed across the cortex, although, depending on the activity involved, certain brain areas will be more at work than others. Emergent properties tend to be higher-order properties of the whole system (Weinert 2009: Chap. II.V).

The notion of emergence helps to avoid a false dichotomy: either the human mind is causally determined or it is causally undetermined, hence random. In both cases human action would be unfree. However indeterminism does not mean randomness or indeed a denial of causality.

The possibility of emergence has led Mario Bunge to the thesis of emergentist materialism, which is characterized by three tenets (Bunge 1977: 506).

1. All mental states (events, processes) are states (events, processes) in the central nervous system of vertebrates (CNS).
2. These states, events and processes are emergent relative to those of the cellular components of the CNS.
3. The so-called psychophysical relations are interactions between different subsystems of the CNS or between them and other components of the organism. There is no one-to-one mapping between brain states and mental states.

Emergence requires that every emergent property of a system can be explained in terms of properties of its components *and* the interactions between them. To be scientifically defensible this set of philosophical hypotheses needs to be translated into empirical research and this has been done in a number of approaches. For instance, the neuroscientist G. Edelman aims at completing Darwin's research programme through the development of a biological theory of consciousness (Edelman 1992, 2004).

> The theory must show how the neural bases for consciousness could have arisen during evolution and how consciousness develops in certain animals. (Edelman 2004: 3)

Such a task, however, requires a much greater knowledge of the 'molecular arrangements' of the brain than was available to Huxley and his contemporaries. But the key to such a materialist approach is still to find the 'neural correlates of consciousness' (Edelman 2004: 13). Edelman proposes a global theory of the brain, called neural Darwinism or theory of neuronal group selection. It has three basic tenets (Edelman 2004: 39–41):

1. Developmental selection leads to a highly diverse set of circuits; 'the dynamic primary processes of development […] lead to the formation of the neuroanatomy characteristic of a given species' (Edelman 1992: 83).
2. Experiential selection leads to changes in the connection strengths of synapses, favouring some pathways and weakening others, resulting from 'variations in environmental input during behaviour.'
3. 'Reentry—during development, large numbers of reciprocal connections are established both locally and over long distances. This provides a basis for signalling between mapped areas across such reciprocal fibres. Reentry is the ongoing recursive interchange of parallel signals among brains areas, which serves to coordinate the activities of different brain areas in space and time. Unlike feedback, reentry is not a sequential transmission of an error signal in a simple loop. Instead, it simultaneously involves many parallel reciprocal paths and has no prescribed error function attach to it.' […] reentry is the central organizing principle that governs the spatiotemporal coordination among multiple selectional networks of the brain.'

The particularly Darwinian aspect arises when an evolutionary event occurs that connects 'previously evolved capacities with new structural and functional features that emerge as a result of natural selection' (Edelman 2004: 48).

Leaving aside the technical details of the brain structure, which—like quantum oscillations—may underlie the observable physical and mental behaviour, the crux of the theory is still how 'the brain can cause the mind.' Edelman appeals to the notion of entailment. His thesis is that the 'phenomenal transform, C, is entailed by the neural activity, C′ (Edelman 2004: 78). Although conscious states (C) accompany neural states (C′), it is the neural correlate C′ that is 'causal of other neural events and certain bodily actions' (Edelman 2004: 78). If there is mental causation, if mental states can act on physical states, this causation must, on Edelman's view, occur via the mechanisms embedded in neural activity, since the 'world is causally closed'.

> The consequences of this line of reasoning is that evolution selected C′ (underlain by the neural activities of the dynamic core) for the efficacy in planning conferred by its activity. At the same time, however, such C′ activity entailed corresponding C states. Indeed, there is no other way for an individual animal to directly experience the effects of C′. (Edelman 2004: 79–80)

Does Edelman's theory achieve the completion of the Darwinian programme? One central problem is that 'entailment' is as much a logical notion as ‘computation’. Entailment is a relation between propositions, not states of affairs. Edelman’s notion of ‘entailment’ $C' \rightarrow C$ still leaves open the question of the physical correlate of entailment. Two much discussed candidates are ‘supervenience’ and ‘emergence’, as discussed in Part II, Sect. 12.2. On the Penrose-Stuart account quantum oscillations provide the bridge between the brain and the mind, although they are only a necessary condition for the emergence of consciousness. On Edelman’s account neural activity does the job.

One difficulty is that an appeal to biochemical or physical examples (like quantum oscillation and neural activity) can rely on lawlike regularities in the physical world but it has not yet been established whether psycho-physical laws exists between the mental and the physical realm. Furthermore, the Darwinian materialists were eager to grant a certain independence to the mind and its products. The Darwinians considered non-adaptive change in their explanation of intellectual and moral faculties. They treated organisms as integrated systems, which implies both that there is no direct mapping of single brain states to single mental states and that brain and mind capacities have been subject to evolution. But weak or strong emergence has no evolutionary dimensions. Recall that *weak* emergence (supervenience) requires (**a**) a co-variation of the properties of one domain, the physical base (the brain), with a supervenient domain (the mind) and (**b**) the dependence of the supervenient domain (mental states) on the base domain (brain states). *Strong* emergence requires the emergence of novel properties from an underlying base, to which the novelty can no longer be reduced. The Darwinian version of emergence also implies that the mental domain is not exhausted by conscious states alone. Consciousness, the world of subjective experiences, may be emergent from brain states. The subjective feelings, which accompany sensations and perceptions, may be ‘entailed’ by the existence of brain states. Though it is true that consciousness—the subjective awareness of ourselves in the world—disappears with the death of the body, it does not follow that all manifestations of the conscious mind vanish with the disappearance of the base. The Darwinians were not concerned with subjective ‘qualia’ but rather with objective results of mental activity. For instance, language, moral values and cultural achievements can survive the demise of individuals and societies. Ideas live on in other people’s minds, in books and computer memories. Ideas can take on a material existence in social and cultural institutions and channel social actions in particular directions. Some mental products may therefore not be sufficiently explained by emergence and call for an additional explanation. For a long time the mind led a rather ethereal existence in philosophy. Under the Cartesian influence, the mind was depicted as a separate entity, confined to its own realm. William James proposed to view the mind as a process, so that today neuroscientists tend to think of the mind ‘as what brains do’ (Blackmore 2003: 13). In addition to the view of mind as a process, recent developments have emphasized strongly that the mind is enmeshed with the world. This has led to the concept of *embodied minds* (Clark 1997; Edelman 1992). The embodied mind interacts with the environment and uses symbolic props—symbolic language, cultural institutions, and memory devices—to go about its problem-solving activities. The embodied mind also leaves publicly available documents, like mathematical theorems, physical equations and many other cultural products in the public sphere. The extended mind is not likely to be simply supervenient on brain states. Given the Darwinian emphasis on correlated variation and the possibility of unselected mental functions, it is likely that they would have regarded the mental faculties as emergent properties.

Both Huxley and Darwin came to emphasize the emergent aspects of mental properties. Thus Darwin stresses the importance of intellectual and moral faculties in the progress of civilization:

> Of the high importance of the intellectual faculties there can be no doubt, for man mainly owes to them his predominant position in the world (Darwin 1871: 153).

Darwin sees in the intellectual faculties an evolutionary advantage. What is of interest in the present context is that intellectual faculties produce cultural products, like moral values, which transcend their individual bearers.

These Darwinian programmes provide plausible models for a possible physical basis of conscious states. But no physical mechanism has been identified, which would provide the necessary and sufficient conditions to show how conscious events actually result from brain events. The development of these programmes is nevertheless an important achievement, since they are a step towards a materialist theory of the mind and a contribution to the Darwinian programme of inserting the mind into the physical world. But whatever materialism can achieve, it seems that a theory of human behaviour with its appeal to human agency will be required to account for the mind. Materialist theories have a long tradition, which predates Darwin's work. To cite an example, Jean Baptist Lamarck who can be regarded as Darwin's precursor, remarked at the beginning of the 19th century:

> What is the mind? It is a mere invention for the purpose of resolving the difficulties that follow from inadequate knowledge of the laws of nature. Physical and moral have a common origin; ideas, thought, imagination are only natural phenomena (Lamarck 1809: Part II, Introduction).

Mechanical views of nature were widespread throughout the 19th century. Towards the end of the century Ludwig Boltzmann asked himself, which philosophy would best characterize his age. He answered without hesitation that it would be called 'the century of the mechanical view of nature.' But he added that it could also be known as Darwin's century (Boltzmann 1886: 15).

Yet this characterization is also surprising since Maxwell invented his Demon to show that the Second law of thermodynamics possessed only statistical validity. Boltzmann took note of Maxwell's Demon. Through his work on the notion of entropy Boltzmann could, with equal credibility, be dubbed the instigator of the 'statistical age'. In fact Darwin's theory of evolution is such a statistical theory, since it fundamentally holds that natural selection, the mechanism, which drives evolution, has a *tendency* to preserve favourable characteristics of an organism—favourable with respect to a given environment—and to eliminate unfavourable characteristics. That is, Darwin's theory is indeterministic.

At first it appears that the evolution of species—and indeed the universe—as well as the increase in complexity run counter to the claim associated with the Second law, namely the universal dissipation of energy, the loss of information and the increase in disorder. This link between evolution and entropy needs to be explored, not least because evolution is related to the arrow of time. As noted in Part I, Sect. 7.3 more than an analogy exists between Maxwell's Demon and the evolution of life and the universe.

Chapter 18
Entropy and Evolution

> *I do not wish to gloss over the fragmentary state of our present knowledge; but the subject of the expanding universe seems to me to deserve prominence as one that it is of the utmost importance to continue investigating.*
>
> Eddington, *New Pathways in Science* (1935: 228)

At first blush it seems that the evolution of species and galaxies runs counter to the seemingly unstoppable increase in entropy and randomness towards a state of equilibrium. Evolution seems to act like a Maxwellian Demon. Recall that Boltzmann characterized the Second law as proclaiming 'a steady degradation of energy until all tensions that might still perform work and all visible motion in the universe would have to cease' (Boltzmann 1886: 19, 1898: §89). He does not entertain any hope that the universe could be saved 'from this thermal death.' But biological systems exist far from equilibrium and since the Big Bang galaxies have formed ordered clusters across the sky. If the whole universe evolves towards a 'heat death'—a total dissipation of all energy differences, making life anywhere impossible—how can nature on Earth have produced the biological complexities, at which Darwin marvelled, and how can planets and galaxies have formed ordered patterns in the sky, which filled Kant with awe? If the Second law is understood as a statistical law, then anti-thermodynamic processes are not forbidden by the laws of nature.

The formation of order in one subsystem of an overall system must be paid for by the increase of entropy in other parts of the system. Living systems on Earth are open systems, which compensate for their entropy production by the constant exchange of energy with their environment (Mainzer 1996: 529). Organic systems on Earth remain far from equilibrium because they consume forms of energy, orderly stored in meat and plants, which increases the overall energy of the system. As discussed earlier, ultimately all forms of life on Earth are possible because the sun provides them with energy in a low-entropy form. The visible sunlight received from the sun consists of low-energy photons. This energy is consumed and transformed from useful to useless energy, both in the form of waste products and high-energy photons, which are radiated back into space (Penrose 1989: 413–414).

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_18

The formation of order in the universe—the clusters of galaxies with their planetary systems—is also paid for by an overall increase of entropy in the whole universe. The original material from the Big Bang clumps into stars and galaxies but all matter in the universe will eventually collapse into black holes; the latter provide some of the highest source of entropy in the universe. The evolution of the universe from smooth initial to messy final conditions may therefore provide us with a cosmic arrow of time.

Could biological evolution serve as a criterion for local arrows of time? According to Dollo's law biological evolution is irreversible. The Belgian palaeontologist, Louis Dollo, stated his principle in 1893: 'An organism is unable to return, even partially, to a previous stage already realized in the ranks of its ancestors.' According to this hypothesis a function, organ or structure which was lost or discarded during the course of evolution will not reappear in exactly the same form in that lineage of organisms. Unlike the law of entropy, Dollo's principle turned out to be only an empirical generalization, since it has since been shown that it is possible to reverse evolution. As the evolutionary biologist Ernst Mayr states clearly, Dollo's 'law' of the irreversibility of evolution is an empirical rule, with numerous exceptions, and 'quite fundamentally different from the universal laws of physics' (Mayr 2001: 227). The fact that evolutionary specialization can be undone means, from a philosophical point of view, that Dollo's 'law' states a mere trend. Although evolution tends to fill ecological niches and thus contributes to an increase in complexity, a change in the environment can lead to changes in ecological niches or even their disappearance. Such ecological changes will force organisms to adapt to new conditions, in the process of which specific organs, like the eye in cave fish may disappear. A disappearance of body elements also appeared in snakes with their loss of eyelids, forelimbs and the loss of external and internal ears. Such organs require inordinate amounts of energy to sustain them but become useless in environments, in which they do not aid the organism's struggle for survival. Evolution accounts, to a certain extent, for the *historical* arrow of time but it also shows that notions like order and complexity, are too imprecise to ground the anisotropy of time.

The contingency and indeterminism of evolutionary and human histories is one reason why Dollo's 'law' is not a reliable factor to serve as a criterion for the passage of time. Other reasons are that evolution is best presented in terms of an evolutionary tree, with both dead-end and live evolutionary branches. Several mass extinctions of whole species have taken place over evolutionary history and may take place again under the impact of global warming. Although large fluctuations in entropy levels are theoretically possible, they are not probable in the lifetime of our type of universe. Entropy is therefore a better criterion for the arrows of time than evolution. It is not a trend like Dollo's 'law' but a proper law of nature, which, although statistical, cannot be reversed. Although the law allows both short- and long-term fluctuations, it is itself irreversible.

Can the law of entropy explain any of the other local arrows of time? Many physical systems on Earth are subject to a physical or *entropic* arrow of time. As mentioned above, entropy can be characterized in three ways: in terms of loss of order, loss of information or as a change in the shape of phase space volume. If entropy is characterized in terms of an increase in disorder, many local systems are affected by it if no external energy is provided to keep dissipation at bay. But the maintenance of many systems, from computers to gardens and zoos, is only possible because energy is provided from outside these systems to either maintain or restore order.

If entropy is characterized in terms of information loss, there are many examples when information about the location of particular items can easily be lost. Consider, again, the perfume bottle sitting in the middle of a concealed container. As long as the bottle is kept closed, knowledge of the location of the molecules is more precise than after the opening of the bottle. Or consider knowledge of the whereabouts of three shoppers in a department store with two floors. As long as they are confined to the ground floor we possess more knowledge of their whereabouts than when they are distributed over the two floors. The phase space notion involves spreading, as seen in the molecule cloud, which spreads through a larger phase space volume, when the perfume bottle is opened. When the second floor becomes accessible to the shoppers, our knowledge of their location in the building decreases in proportion to the increase in their spreading over the coordinate space. There are 2^3 ways in which they can be distributed across the two floors. As they occupy more of the available (phase) space our information decreases in proportion to their spreading and the information entropy increases. This loss of information is easily reversible. In many cases, then, an input of energy can restore order and information. Similarly, the volume of phase space can be controlled by manipulation, as Maxwell's and Loschmidt's Demons show.

In some case, however, it is impossible to keep the system away from equilibrium. Human beings are born young and mostly die old but no amount of energy taken from outside can prevent the final disintegration. Food can be kept for a limited time in refrigerators and freezers but its eventual decay can only be delayed at a considerable expenditure of energy. Hot coffee will eventually grow cold but is unlikely to heat itself again in the lifetime of the universe.

The dominance of the entropic arrow can be understood in various ways. The advantage of entropy is its ubiquity. Entropy is a universal feature, which is felt on both the local and global scale. It should therefore be possible to relate entropy to some of the other local arrows of time, like the *causal* arrow and the *past-future* asymmetry. (Part IV will raise the question whether entropy can serve as a master arrow of time.)

18.1 Entropy and Causality

In Part II the consideration of Laplace's Demon revealed the need for a distinction between determinism and causality. In the present chapter the notion of indeterminism showed that an antecedent cause can have a number of differentially weighted effects, $E_{1,2,3}$. Conversely, a given effect, E, can be related to some antecedent causes, $C_{\alpha',\beta',\gamma'}$ with different probability weights. As has been emphasized the notions of cause and effect can be broken down into a distinction between antecedent and consequent conditions. This distinction will prove useful for the description of a conditional model of causality and its relation to entropy.

18.1.1 Causation

The notion of causation has had a long, if not altogether distinguished career in philosophical and scientific thinking, harking back to antiquity. Over time, several different models of causation have been proposed. To mention but a few, Aristotle distinguished four notions of causation. If the effect of construction is the building of a house, then the architect's plan is the *formal* cause, the building materials are the *material* cause, the builders' work is the *efficient* cause, and the finished product—the warmth of the house—is the *final* cause. The transition from the medieval organismic to the modern mechanistic worldview reduced this matrix of causal notions to just efficient causation. Laplace understood Leibniz's 'Principle of Sufficient Reason' as affirming the universal statement that 'every event has a unique cause'. The 16th century discovery of mathematical laws of science, first introduced by Johannes Kepler with his three laws of planetary motion and the 17th century development of differential equations promulgated a functional view of causation. The functional view equates causation with the existence of differential equations, which describe the rate of change of some physical parameter with respect to time. The mathematization of the physical sciences, the invention of the calculus and differential equations, strongly encouraged the Laplacean Demon's identification of the notion of causation with that of causal and predictive determinism. But the Laplacean account was not the only one, on which scientists could draw. Kant's notion of causation as an a priori category of the mind exerted a strong influence on some continental scientists. Hermann Helmholtz, for instance, one of the co-discoverers of the principle of conservation of energy, stated that

> the law of causation, by virtue of which we infer the cause from the effect, has to be considered also as being a law of our thinking which is prior to all experience.[1]

Following in Kant's footsteps the Danish physicist Niels Bohr speculated in 1929 that 'causality may be considered as a mode of perception by which we reduce our sense impressions to order' (Bohr 1929: 116). By contrast the pioneering physicists Werner Heisenberg and Max Planck both insisted that the recent discoveries in quantum mechanics had shown that causation was not a necessary category of

---

[1] Quoted in Warren/Warren, *Helmholtz on Perception* 1968: 201, 228. The views expressed by von Helmholtz in his Introduction to *Über die Erhaltung der Kraft* (1847), are also strongly Kantian in flavour. However in a Footnote, added in 1881, von Helmholtz distanced himself from this earlier Kantian influence and equated the principle of causation with lawfulness (i.e. determinism).

thought. For if causation is identified with determinism, the indeterminism of subatomic particles remains a mystery. As de Broglie's thought experiment shows (Fig. 17.2), subatomic events can be both *indeterministic* and *causal*.

How can a philosophical model of causation be made compatible with such scientific discoveries? In an attempt to understand empirical results, a scientist considers the causal conditions, which would produce the specific effects observed in the experiments (Figs. 11.9 and 17.3). Under these specifiable conditions there is a lawlike dependence of the effects on the antecedent conditions. For a human observer, this situation may give rise to a probabilistic notion of causation because no specific predictions are possible about the path of individual particles. For such events, even a Laplacean Demon may have to accept his limits. The suggestion could be made to analyse these situations by reference to a number of causal conditions or variables, which bring about the effect. Of these conditions some may be *necessary* and others *sufficient*. Of the variables some may be *dependent*, others *independent*. But the basic idea is that there will be a conditional dependence of the consequent on the antecedent conditions.

A suitable account of probabilistic causation may be developed from a consideration of the causal conditions involved in the generation of some phenomenon. As mentioned in Part II, Chap. 8, an invariable succession of events—like day and night—qualifies as a correlation rather than a causal relation. A causal relation requires that an effect become conditionally dependent on the existence of causal conditions.

Hans Reichenbach was one of the first philosophers who attempted a conceptual model of causation, which would be compatible with the indeterminism of quantum mechanics (Reichenbach 1920, 1931). Reichenbach sought to achieve this compatibility by associating the notion of causation with that of probability. The first step in any physical situation, in which a causal connection may be suspected between antecedent and consequent conditions, is the recognition that the antecedent (or cause) consists of a number of factors, controlled and background parameters. In the physical sciences it is customary for some of the background factors to become measured parameters; other background factors may be negligible. A closer analysis may turn some of these rest factors into measured parameters but the rest factors can never be exhausted. For Reichenbach causation is concerned with the relation between *individual* measured parameters. Probability has to do with the distribution of the background factors. The principle of causation cannot be formulated without the principle of a statistical distribution. Causal claims take the form of an implication ('If C, then E'). But it is known that C consists of observable measured parameters and background factors, which may equally have an influence over E. In some cases, especially in the physical sciences, such background factors can be neglected but in others, especially in the biological and social sciences, background factors may play a non-measurable part, to which, perhaps, only a demon would have access.

> What we know of C, can only be expressed in terms of a statistical statement: we know that subsequent situations, *with great probability*, differ little from C. (…) We predict E only with probability, not certainty. *Every causal statement, applied to the prediction of a natural event, has the form of a statistical statement. (…) If an event is described by a finite list of parameters, the future evolution of the event can be predicted with probability. This probability tends towards 1, the more parameters are taken into account.* (Reichenbach 1931: 715–716, italics in original; cf. Bunge 1959: §2.2. Author's own translation; Reichenbach's letters A, B have been exchanged for C, E).

This notion of probabilistic causation, which is implied in Reichenbach's combination of probability and causation, can then be regarded as a generalisation of the classical notion of causation. Reichenbach anticipated a *conditional* view of causation, which is close to a modern philosophical version of the conditional view of causation: John Mackie's ***INUS*** account.

In *The Cement of the Universe* (1974) Mackie made no attempt to measure the adequacy of his *INUS* account against classical physics, let alone quantum mechanics. Rather, Mackie tried to develop a general model of *physical* causation. Causation as it works in the real world is the cement that holds the universe together. This model stands in the tradition of D. Hume and J.S. Mill. Any questions of the existence of a causal bond in the physical universe between correlated events, over and above their succession, are treated with caution. In the physical world, causation is only regular succession of events.

> 'Causation', Mackie holds, 'is not something *between* events in a spatio-temporal sense, but is rather the way in which they follow one another.' (Mackie 1980: 296; italics in original)

For Mackie, a cause is an ***INUS*** condition, an ***I***nsufficient but ***N***on-redundant Part of an ***U***nnecessary but ***S***ufficient condition for some *E* (Mackie 1980: 62). Thus there is a cluster of factors, making up the cause *C*, which brings about the effect *E*. Unlike Mill, however, who took the cause to be the sum total of the conditions, Mackie makes a distinction between *necessary* and *sufficient* conditions.

If *X* is a *necessary* condition for *Y*, then in the *absence* of *X*, *Y* cannot happen. In the absence of oxygen (*X*), fire (*Y*) cannot occur. Note that the mere presence of *X* does not mean that *Y* will occur since a fire needs other factors than oxygen to start. It is just that, in Mackie's words, 'whenever an event of type *Y* occurs, an event of type *X* also occurs' (Mackie 1980: 62).

If *X* is a *sufficient* condition for *Y*, then in the *presence* of *X*, *Y* will occur. Rain is a sufficient condition for the street to get wet, since in its presence the streets get wet. But rain is clearly not a necessary condition for the streets to get wet, since in its absence, flooding or sprinklers could achieve the same effect.

One obvious disadvantage of this minimal 'conditional' view of causation presented so far is that it makes no distinction between conditions, whether necessary or sufficient, which are physically efficient in the production of the effect and non-operative conditions, which are merely in the background of the cause-effect relationship. This is sometimes expressed by making the distinction between 'causal conditions' and 'contributing conditions.' Mill regarded all conditions as constituting the cause. To reflect this distinction, Mackie introduces the concept of a *causal* *field* (Mackie 1980: 35, 63). A causal field comprises the background conditions, which make the normal running of things possible. But the background conditions are different from the causal conditions. The antecedent conditions, *C*, must make a *difference* to the consequent conditions, *E*, within a field to establish a cause-effect relationship. This condition implies that for *C* to make a difference, *E* must be conditionally dependent on *C*. In the Stern-Gerlach experiment (Fig. 17.1), for instance, a beam of atoms travels through the magnetic field to leave traces on a recording screen. Antecedent causal conditions are statistically relevant if their inclusion in the causal account affect the probability of the outcome, otherwise they are statistically irrelevant. In the Stern-Gerlach experiment it is statistically relevant whether the magnetic field is non-uniform but the motion of atoms in the magnet itself is statistically irrelevant. Whether a condition belongs to the causal field (the statistically irrelevant conditions) can often be measured. For instance, the atoms in the inhomogeneous magnet are known to play no causal role in the experiment. If a causal condition is to make a difference to an effect in a causal field, its efficient operation will also require energy. The question arises: What is the link between entropy and causality?

18.1.2 Causality and Entropy

Causality has been characterized in terms of a cluster of antecedent and consequent conditions. The consequent cluster of conditions cannot just lie temporally later than the antecedent conditions. A causal account must go beyond the idea of an invariant sequence of events. The consequent condition must be *conditionally dependent* on the antecedent condition. For a distinction must be drawn between a simple invariant correlation (like day and night) and a causal relation (like rain and wetness).

Nevertheless causal relations involve temporal relations. David Hume pointed out that a cause-effect relation involved a temporal succession of events: the cause is temporally prior to the effect. But Kant realized that a cause can be simultaneous with its effect. If a lead ball is placed on a cushion it will cause a hollow in it for as long as the ball is kept there. By contrast, a hollow in a cushion does not cause a lead ball to rest there.

How do we know in this and similar examples of simultaneous causality that it is a genuine causal relation and not a mere correlation? The hollow in the cushion is conditionally dependent on the gravitational force, which the ball exerts on the cushion. A cause is not necessarily earlier than an effect in a merely temporal sense but it is always, in our experience, conditionally prior to the effect. We do not normally endeavour to provide a causal explanation for the ‘normal running of things’. Causal questions arise when the normal running of things is interrupted: Why does the car refuse to start? Why does the light switch fail?

Whether the cause is temporally earlier than the effect or simultaneous with it, in both cases an expenditure of energy is required for the causal conditions to exert an effect on the subsequent conditions in a causal field. In Kant's example it is gravitational energy. Maxwell's Demon needs to expend kinetic energy to separate the slower from the faster molecules. Loschmidt's Demon also requires kinetic energy in his attempt to reverse the motions of the particles from a final, fibrillated state to a smooth, initial state. (By contrast Laplace's Demon does not require energy—except for his computations—because he does not causally interfere with the running of the deterministic world.)

Causal conditions are operative in the physical universe. Such causal relations affect the energy balance of the systems at hand. Recall that high-energy photons from the sun sustain life on Earth but are ultimately radiated back into space as low-energy, high entropy photons. Causality therefore not only has temporal and conditional but also entropic connotations. It has been proposed to associate the earlier-later direction of causation with the direction of entropy:

> (t)he t-direction from cause to effect is necessarily the same as the t-direction of entropy increase. (Eckhardt 2006: 16)

This suggestion is reminiscent of Eddington's view that

> (t)he discrimination between cause and effect depends on time's arrow and can only be settled by reference to entropy. (Eddington 1932: 129)

But whether this thesis is correct depends on the three different meanings of entropy.

- The most common understanding in terms of a *decrease* in order. A brief reflection, however, shows that causality can produce order *as well as* disorder. The sun's high energy photons allow life on Earth but the sun's rays can also be harmful and destructive. The association of entropy in terms of (dis-)order and causality is therefore not helpful because there is no uniform link between the direction of entropy and the temporal direction of causality. Causality is associated with either the destruction of order—a bridge is demolished—or the creation of order—a building is constructed. In both cases the cause is prior to the effect; but only in the latter case is the entropy of the local system reduced. There is no unambiguous link, in a local system, between the direction of causation and a disorder-based account of entropy.
- If entropy is understood in terms of loss of *information*, the link with the temporal direction of causality is equally ambiguous. In some cases an increase in entropy is associated with a loss of information. For instance when a demon removes the top from the perfume bottle, which is placed in a sealed container, the perfume molecules will soon occupy the whole available phase space. Information about their whereabouts has been minimized. But the job of Maxwell's Demon is to increase the information. The Demon could, instead of separating fast and slow molecules, separate air from perfume molecules and thus trap the perfume molecules back in the bottle. The Demon would reduce

information-entropy and reverse the direction of causation. There is no unambiguous link, in a local system, between the direction of causation and the information-based notion of entropy.

- If, however, entropy is to be characterized in terms of *phase space* volumes, would the direction in causal relations coincide with the increase in occupied phase space? Would the temporal order of cause and effect in this way be related to the thermodynamic arrow of time? Recall that thermodynamic systems have a tendency to occupy more of the accessible phase space. If the cause is earlier than the effect, the phase space occupied by the antecedent condition must be smaller than the phase space of the subsequent conditions. Intuitively this idea may work when a causal interference is 'destructive' and creates disorder. When a demon, like Loschmidt's Demon, opens the perfume bottle, the perfume molecules undergo a transition from a relatively small region of the phase space to a less ordered state in a larger region, in which they come to occupy the whole volume of the sealed container and mix with the air molecules. In this case it could be said that 'the direction of causality derives from that of entropy increase' (Eckhardt 2006: 2). If causal asymmetry is indeed based on the direction of increasing entropy, it would be much easier to understand why backward causation—from future to past—is so difficult to achieve. However, what happens when a causal interference is 'constructive' and restores order? After having released the molecules Loschmidt's Demon returns them, by judicious separation from the air molecules, to the bottle. Locally, the Demon has reduced the occupied phase space of the perfume molecules to a smaller volume, and hence has decreased their entropy. In this local system, then, the entropic arrow and the causal arrow are not aligned. On the phase space argument the phase space volume of the causal conditions must be smaller than the phase space volume of the effect conditions. But in the case of constructive causal interference, the effect conditions are not more fibrillated than the causal conditions, i.e. the Demon has reduced the phase space volume of the perfume molecules from a larger to a smaller phase space.[2]

In causal relations a temporal, a conditional and an entropic part have to be identified. To which component does the thesis refer?

Does the thesis refer to the conditional Part of causal relations? According to the conditional view an effect is conditionally dependent on a causal antecedent, for instance in terms of a force exerted. The exertion of force requires energy and therefore an increase in entropy. This increase in disorder is manifest in destructive causal interference in a local system but is not visible in constructive causal interference. As was pointed out in the case of biological and cosmological evolution, an increase in entropy can be accompanied by an increase in order, even on a global scale.

---

[2] In this connection, think of volume as the region, which a number of molecules occupy.

This thesis cannot apply to the entropic part either because causal interference can restore order or reduce the occupied phase space volume of the effect conditions in comparison to the causal conditions. Loschmidt's Demon who reverses the motion of all particles achieves precisely this feat.

The most natural interpretation is therefore to say that the thesis refers with the temporal aspect of causal relations: it seems that a cause is always earlier than an effect. If Loschmidt's Demon is banned from operating, it is a common experience that backward causation does not occur. On the other hand, the local scale must be distinguished from the global scale. In each case considered, it was always the *local* system, in which the association between the *t*-direction of causality and the *t*-direction of entropy increase could be broken. The entropy of a particular future state is highly likely—but not necessarily so—to be higher than the entropy of a particular past state. But in each case it is also true that globally the entropy increases, according to the Second law of thermodynamics. A cluster of causal conditions is always part of an earlier state of the universe than a cluster of subsequent conditions. And it follows from the Second law—in its statistical version—that the entropy of the future state of the universe is likely to be higher than the entropy of its past state. Note that this claim is compatible with local decreases in entropy due to constructive causal interference. But if the entropic arrow can explain the earlier-to-later direction of causation, it is natural to ask whether it can also explain the past-future asymmetry, with which humans are particularly familiar?

Chapter 19
The Past-Future Asymmetry

(...) energy is Nature's currency (...).
Alvarez, *T.rex and the Crater of Doom* (1997: 8)

It is a commonplace experience that we can influence the future but not the past; that we have knowledge of the past but not of the future. It is also well-known that most accounts of time travel into the past engender paradoxes—like the grandfather paradox—, which do not arise from time travel into the future. There exists therefore a *past-future asymmetry*. The asymmetry is, on reflection, also puzzling, since no such problems exist in the spatial analogue. We can revisit a place an indefinite number of times; we can alter a past place and an architect's plan incorporates knowledge of future places. Yet, in the temporal analogue, we are powerless with respect to the past, if not with respect to the future. There is no *here-there* space-asymmetry in the same way as there is an *earlier-later* time asymmetry. But why can we not visit past times in the same way that we can visit past places? A simple answer is that the past no longer exists and the future has not yet arrived. Only the present is real. But this simply postpones the puzzle: why does the past no longer exist? And when does the present exist? Saint Augustine may have a point when he says that the present moment is infinitely small for any event that is Now—say, the utterance of a word—can be subdivided into past and future moments.

Note that the simple answer in the temporal case does not hold for the spatial case. For it could reasonably be said that the past exists in fossil records and old monuments. I can lay my hands on the Sphinx in a way that I cannot lay my hands on Bismarck. It should also be noted that the simple answer presupposes a common view of time: all events glide from the future into the present and fade irretrievably into the past. It contradicts, as we have seen, the Special theory of relativity, which has destroyed the notion of a universal Now and seems to imply the so-called static

---

This Chapter is partly based on my paper 'The Past-Future Asymmetry' (2013b), by permission of Brill.

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_19

Block Universe, in which past, present and future are equally real. The puzzle of the past-future asymmetry has intrigued many thinkers, and some of their solutions will be reviewed. But the question is whether the notion of entropy—and its 'universality' in the familiar macro-world—can provide a satisfactory answer. This chapter will argue that the human experience of the past-future asymmetry can indeed be explained by appeal to the notion of entropy.

19.1 Some Attempts to Explain the Past-Future Asymmetry

Let us look at three different attempts to explain the past-future asymmetry. They rely on fundamental considerations other than entropy.

1. Lockwood (2005: Chap. 12) makes use of the above-mentioned conditional model of causation, according to which the cause, *C*, of an event, *E*, can be broken down into a set of necessary and sufficient conditions. Jointly, the antecedent necessary and sufficient conditions can explain the occurrence of *E*, because *E* is conditionally dependent on *C*. Lockwood introduces the term 'nasic conditions', which is an acronym for Millian necessary and sufficient conditions in the circumstances. A causal set of antecedent factors, *C*, is a nasic condition for the occurrence of event *E*, where appropriate background conditions are taken into account. A cause introduces a change in the normal running of things. Lockwood furthermore embraces Reichenbach's observation that 'one can infer the total cause from a partial effect but one cannot infer the total effect from a partial cause' (Lockwood 2005: 253; see Fig. 19.1).
   Reichenbach pointed out that the '*inference from the partial effect to the total cause* is typical of all forms of recording processes' (Reichenbach 1956: 180; italics in original). Partial effects are typically sets of conditions that obtain locally in conjunction with other partial effects and general background conditions. Consider a pond analogy: when a stone is thrown into the centre of a pond it will create divergent waves that will break on the shore of the lake (Popper 1956: 538). From the arrival of divergent waves and the breaking of the waves on the shoreline, one can infer that a disturbance at the centre caused the divergent waves. The analysis of the divergent waves may even tell us something about the physical properties of the object. But knowing only partial

Partial effect → total cause → past (records)
Partial cause ↛ total effect ↛future
Total cause = common cause of multiple effects. (Reichenbach 1956: 180)

Key to symbols: → (inference permitted); ↛ (inference not permitted)

Fig. 19.1 Partial effects and total causes causal conditions prevents us from inferring what the total effect of these conditions will be. Thus knowing that a stone has been launched at an angle α, and that some additional parameters are satisfied, various characteristics of its trajectory can be calculated; but if we do not have more information we cannot predict that it will hit the middle of the pond, that it will startle a duck that flies off into the sight of a hunter's rifle. This conclusion does not necessarily hold under conditions of uncertainty about partial effects and when the cause-effect relationship is a many-to-one relation, as it may occur in irreversible thermodynamic processes. For instance, the present room temperature may have multiple microscopic configurations as a cause, since the air molecules can have a whole range of velocities and still make up the macro-condition, which is the temperature of the room. Equally, indeterministic relations between cause and effect make inferences from the cause to the effect and vice versa more complicated (Figs. 11.9 and 17.3).

Using these notions, Lockwood attempts to explain the past-future asymmetry. He holds that 'events frequently have highly localized nasic conditions in the past' but they do not tend to have 'highly localized nasic conditions in the future' (Lockwood 2005: 253). Lockwood's view on the difference between past and future can be summarized as follows:

(a) Past outcomes are genuinely overdetermined by their currently prevailing partial effects. The totality of various partial effects overdetermines the total cause. Hence, in line with Reichenbach's conclusion, we can infer the total cause from their various partial effects. Note that we have to learn a lot about the partial effects to be able to exclude alternative causes and ensure that the total cause identified has a much higher probability of explaining the effect than alternative causes. Such painstaking detective work went into the identification of the cause of the dinosaurs' extinction 65 million years ago. According to the so-called Alvarez hypothesis, it was due to an impact from outer space rather than volcanic activity; but it has not yet been established whether the impact was due to an asteroid *or* a comet (see Alvarez 1997).

(b) By contrast, events do not have highly localized nasic conditions in their future (Lockwood 2005: 253). Whilst effects radiate out from some highly localized causal event, it is not the case that some future events are so highly localized that they could serve as nasic conditions for present events. This fact has been labelled the "law of conditional independence": incoming influences emanating from different directions in space are uncorrelated (Penrose and Percival 1962; cf. Price 1996: 118; Weinert 2013a: Chap. 4.4.4). Consider again Popper's pond analogy to illustrate this difference. When a stone is thrown into the middle of the pond, its waves will diverge towards the shore. From the divergent waves it can be inferred at least that some local disturbance caused the ripples on the lake's surface, although further investigation may be required to determine the nature of the disturbance. By contrast, there are no highly localized nasic conditions on the shore of the lake that would cause convergent waves towards the centre. Without the help of Loschmidt's

Demon, we cannot expect a disturbance in the middle of the lake to be caused by uncorrelated (fibrillated) conditions on the shore of the lake. It is highly unlikely that conditions on the lakeshore will conspire to produce converging waves towards the centre sufficient to lift a stone from the bottom of the lake. Thus Lockwood's explanation of the past-future asymmetry boils down to the statement that present events are conditionally dependent on past events, while present events are conditionally independent of future events. But seen in this light Lockwood's analysis is more like a redescription than an explanation of the asymmetry between past and future events. We can always ask for a physical mechanism why 'nasic conditions' are typically highly concentrated in the past and why future actions cannot be regarded as nasic conditions for current events.

2. Albert (2000: Chap. 6) introduces a distinction between the prediction of future states and the retrodiction of past states and distinguishes retrodictions from records of past events. Both predictions and retrodictions are inferences from the present state of affairs to either a future state of affairs (say, the occurrence of a solar eclipse) or to a past state of affairs (say, the identification of an unidentified celestial object in Galileo's notebook as the planet Neptune). The ability to make predictions and retrodictions depends on the use of appropriate equations of motion and the availability of boundary conditions. But according to Albert most of our knowledge about the past cannot be derived from retrodictions; we have it by means of indelible records. Such past records are irreversible data of past events. Albert argues that we have epistemological access to the past other than by means of retrodiction, because records are related to a certain feature of the Second law of thermodynamics. This feature involves the so-called Past Hypothesis, which states that 'the world first came into being in whatever particular low-entropy highly condensed big-bang sort of macro-condition' (Albert 2000: 96; see also Novikov 1998: 204ff). Albert's thesis is that we have records of the past because our experience is 'confirmatory of the past hypothesis but not of any future one.[1]' (Albert 2000: 118)
   It seems implausible to relate the existence of past records to the beginning of the universe and its particular energy state. Rather, as Murray Gell-Mann and James Hartle observe, 'a record is a present alternative, that is, with high probability, correlated with an alternative in the past (Gell-Mann/Hartle 1990: §10). The current state of the world is branch-dependent on the past state, where the past state may have happened quite recently: it is 'contingent on events that

---

[1] It should be mentioned in this connection that both Lockwood's assumption that future events are uncorrelated with respect to the present and Albert's assumption that the entropy of the Big Bang is lower than the entropy of the Big Crunch have been questioned in the literature (see Price 1996, 2002; Schulman 1997 and the previous discussion of Gold universe models).

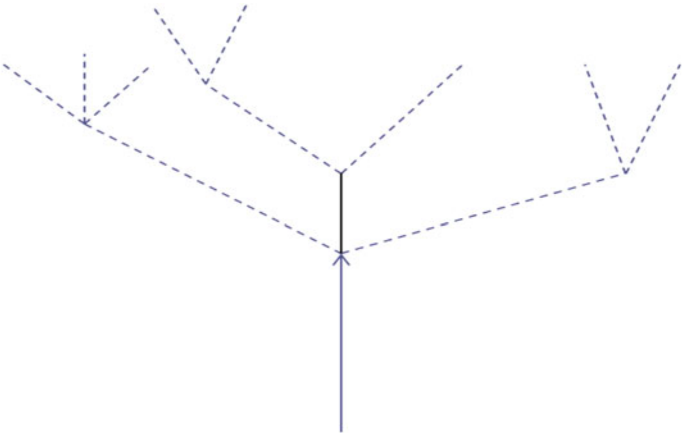


Fig. 19.2 Branch attrition in the branching tree model, indicating the probabilities of the occurrence of future branches

have happened' (Gell-Mann/Hartle 1993: 3346). Lockwood seems to be closer to the mark with his suggestion that nasic conditions are highly concentrated in the past but not in the future. But Lockwood redescribes the asymmetry rather than explaining it. By contrast Albert hints at a connection between the past and the Second law of thermodynamics but implausibly explains our experience of the past as confirmatory of a low-entropy Big Bang.[2]

3. Finally, McCall (1994: Chaps. 1, 2; cf. Horwich 1987: Chap. II.3) defends the view that the difference between past and future is that the future comprises many possibilities, while the past is actual and unique. His explanation is not based on whether we can travel to the past, or whether we can change the past, or whether entropy increases, though it does imply that the past is fixed. His model is incompatible with the static Block Universe since if the world is indeterministic (or probabilistic) this view requires that every physically possible future be situated on a different 4-dimensional branch, and that the probability of any future event is specified by the proportion of future branches on which that event occurs (cf. Part II, Sect. 11.2.1). The model is also incompatible with the view that the march of time is a purely subjective phenomenon. The model is in the shape of a tree, the past being a single trunk, the future a multiplicity of branches, and the present the first branch point (Fig. 19.2).

---

[2] It is now commonly assumed that our universe started in a low-entropy Big Bang and has expanded ever since. This expansion, which actually accelerates, as well as the effects of gravity, establishes a connection between the thermodynamic and cosmological arrows of time. Many cosmologists now believe that the universe will end in what the nineteenth century termed a 'heat death,' that is, a total dissipation of energy that will make life impossible (see Carroll 2009; Penrose 2010). In this sense, there is a connection between the expansion of space and the spatial spreading of energy states (see Part IV).

> Objective time flow is represented by the vanishing or ‘falling off’ of branches, the one branch remaining being the ‘actual’ one that becomes part of the trunk. The present moves stochastically up the tree. It would continue to do so even if all conscious beings in the universe vanished. This view attributes the evolution of the universe, from past to future, from possibility to actuality, to the process of ‘branch attrition’.

In the tree model, branch attrition seems to describe the past-future asymmetry in an objective, observer-independent way. But branch attrition is not a known physical mechanism. The model only serves a descriptive purpose. It describes metaphorically why the frozen past is different from the potentialities of the future, but it does not *explain* the past-future asymmetry. For instance, it does not explain why we cannot descend down the tree trunk and return to the past. Branch attrition should not prevent us from climbing down the tree. What is needed to make McCall’s tree model plausible is a known physical mechanism that can account for ‘branch attrition.’ Such a mechanism was also the missing element in Lockwood’s re-description. In the recent physical literature the notion of decoherence has been proposed as a mechanism to explain the emergence of quasi-classical sets of histories, i.e. the individual histories of our classical world, which obey with high probability, ‘effective classical equations of motion interrupted continually by small fluctuations and occasionally by large ones’ (Gell-Mann/Hartle 1993: 3345, 3376). Decoherence signifies the emergence of classical macro-states from their underlying quantum states as a result of measurements by their environments. Decoherence leads to different alternative histories for the universe, to branch-dependence of histories and the permanence of the past (Gell-Mann/Hartle 1993: §10). Branch-dependence means that individual histories are ‘contingent on which of many possible histories have happened’ (Gell-Mann/Hartle 1993: 2246–2247). The permanence of the past expresses the feature

> of a quasi-classical domain that what has happened in the past is independent of any information expressed by a future projection. Neither the decoherence of past alternatives nor the selection of a particular past alternative is threatened by new information. (Gell-Mann/Hartle 1993: 3354)

Decoherence is an irreversible process, for all practical purposes. Since decoherence can be understood as the carrying away of phase information into the environment, leading to noise (Gell-Mann/Hartle 1993: 3364, 3376), and as a form of continuous measurement of quantum systems by the environment, leading to entropy (Schlosshauer 2008: 41), decoherence leads to irreversible past records. But irreversibility is a feature of the Second law of thermodynamics; hence it is appropriate to associate the past-future asymmetry with the increase of entropy.[3]

---

[3] Although the Second law of thermodynamics was discovered in the middle of the nineteenth century, it is a minor scandal in physics that no unanimous agreement exists on its precise meaning. For the purpose of the following discussion, which uses the notion of entropy to discuss a mechanism for the past-future asymmetry, it will be sufficient to use the established sense of the notion of entropy as it is discussed in statistical mechanics. As human awareness of the past-future

19.2 Entropy and the Second Law of Thermodynamics

All of the approaches considered so far are engaged in obtaining a clearer conceptual understanding of the past-future asymmetry, but they still fail to provide a dynamic explanation. Albert alluded to a dynamic reason for the past-future asymmetry in terms of the Second law of thermodynamics, while decoherence encodes a physical mechanism to obtain the classical domain from the quantum realm.

It is therefore appropriate to consider a dynamic reason for the past-future asymmetry, one that makes full use of the Second law of thermodynamics and the notion of entropy increase. It leads to a general thesis: the entropy of past states (including causal states) is fixed and we cannot change them, but we can influence the entropy of future states.[4] The discussion will be confined here to classical notions of entropy because humans experience the past-future asymmetry in the macro-world.

It is customary, in standard textbooks, to describe the increase in entropy in terms of an increase in disorder or loss of information. For instance, the physicist Kip Thorne (1994: 424) illustrates the notion with the case of a father who arranges the toys of his child's playroom in such a way that all the toys are stacked

---

(Footnote 3 continued)

asymmetry concerns macroscopic systems, a quantum-mechanical version of entropy, which refers to quantum states, appears to be less useful than the statistical version of entropy. Furthermore, the notion of entropy can be used either to describe the asymmetries in the familiar environment—in which case we should speak of the *local arrows* of time—or in a cosmological context, in which case we should speak of the *cosmic arrow* of time. In the cosmological context, entropy is used to consider both the origin of time's arrow and the ultimate fate of the universe. As we shall see, the universe itself can be understood either as 'our' universe or as the 'multiverse,' of which our universe would be only one universe in many parallel universes (cf. Carroll 2008, 2010; Krauss/Scherrer 2008). Although the arrow of time and its origin are challenging tasks for cosmologists, the present discussion will be limited to the passage of time in our familiar environment because this familiar environment challenges us with the past-future asymmetry.

[4] In the present book, entropy is taken to be an important feature of physical systems, in line with classic statements in Davies (1974) and Penrose (2005, 2010). We also assume the validity of the notion of entropy as defined in Statistical mechanics and as an accepted part of standard textbooks. For instance, Carroll (2010: 32) calls it Nature's most reliable law. It should be noted that entropy is difficult to measure experimentally and that it is actually inferred from macroscopic parameters, like temperature, pressure, and work (Čápek/Sheehan 2005: 26). However, one cannot pretend that 'entropy' is an uncontested notion. There exists, for instance, a variety of different definitions of entropy in the literature; and the Second law can be stated without reference to entropy, since the original Clausius statement explains that all work can be transformed into heat but not all heat into work (cf. Čápek/Sheehan 2005: Chap. 1; Uffink 2001; Leff 2007). Furthermore, while the Second law lacks a thorough theoretical proof, "its empirical support is vast and presently uncracked" (Čápek/Sheehan 2005: 42; Sklar 1993: Chap. 7). Although its theoretical foundation is uncertain, it rests on firm empirical evidence, which justifies its use in the present argument. Despite the disagreement about the theoretical grounding of entropy, it will be assumed that a spreading descriptor, introduced above (Leff 2007: 1744), is an appropriate tool to support the argument in favour of the past-future asymmetry.

in one corner of the room. It is clear where the toys are. Then the child comes to play and scatters the toys all over the room. There is loss of information as to the location of the toys. When the child has grown tired and is put to bed, the father returns and restores the toys to their former ordered state. Through the father's constructive causal interference, the information is restored. According to the Second law the overall entropy of the playroom situation, regarded here as a closed system, has increased because the father needs to spend energy to restore order in the playroom and the child has spent energy playing (scattering the toys). However, this widespread illustration can also be very unhelpful because an increase in entropy can be accompanied by an increase of order. For instance, the overall entropy of the universe since the Big Bang is commonly understood to have increased, but the universe today exhibits a great deal of order as seen in the formation of solar systems, galaxies and clusters of galaxies (Carroll 2008; Davies 1974: §4.6).

A better way to understand entropy, as mentioned above, is in terms of *phase space volumes*: the micro-states of a system spread into the available phase space and thereby increase the entropy of the system. To illustrate, consider again a sealed container, in which a perfume bottle is placed. There are about 6×10^{20} molecules in this bottle (Fig. 19.3). When the lid is removed from the bottle, the molecules will spread and mix with the air molecules. The amount of spreading is a function of time, since the perfume molecules begin to occupy the co-ordinate space in the container. If we take a snapshot of the spreading at time t_1 ($t_1 > t_0$), we observe a certain amount of spreading; at a later time, t_2 ($t_2 > t_1$), the spreading will have increased. This spreading can serve as a primitive clock. Statistically speaking the amount of spreading will increase with high probability until the system reaches equilibrium. Humans can only stop the spreading if they decide to interfere with the system from outside, for instance by partitioning the box before the perfume molecules have filled all the available space. Then the spreading and increase in entropy are halted by the injection of external energy. Loschmidt's Demon could of

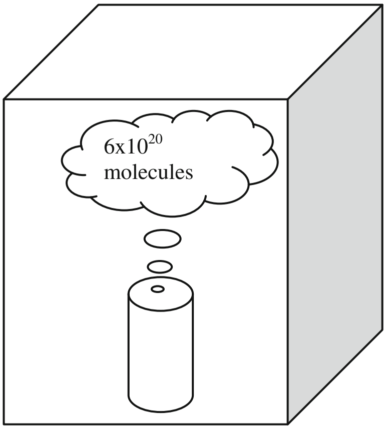


Fig. 19.3 Expanding molecule cloud in phase space, here a three-dimensional coordinate space; spreading as a function of time

course take up residence in the container and attempt to return the molecules to the perfume bottle.

This spreading metaphor may be made more precise in the following manner:

> An increase of entropy may be said to correspond to a "spreading" of the system over a large number, *W*, of occupied quantum states. Alternatively one might say that entropy is a measure of the extent to which the system in question is unrestrained; the less constrained it is, the greater the number of its accessible quantum states for given values of those constraints which exist. (Denbigh/Denbigh 1985: 44)

The authors refer to *W* as the number of occupied quantum states but in the present context it will be understood as the thermodynamic probability. *W* then expresses the number of ways in which a macroscopic state, like pressure or temperature, can be realized by its micro-states. Recall that temperature in physics expresses the mean kinetic energy of the molecules. Air molecules at room temperature, with an average speed of 2000 mph ($\cong 3.47 \times 10^4$ cm/s), can have speeds anywhere between 0 and 4000 mph.

W is related to entropy by the Boltzmann relation: $S = k_B \log W$. In order to appreciate this relation, actually the definition of entropy in statistical mechanics, more must be said about the distinction between the macro-states of a system, and its micro-states, which make up the constellation of the microscopic constituents of the system. The molecules from the perfume bottle are the microscopic elements, while the pressure of the molecules on the container wall is a macroscopic state. It generally holds that any macro-state corresponds to a large number of micro-states, but that any given micro-state corresponds to one given macro-state. A useful analogy is to think of the number of ways in which you can pay for an item that costs say, £10. You may tender a ten-pound note, ten one pound coins, twenty fifty pence coins, one thousand one pence coins and so on. The coins and notes correspond to the micro-states, whilst the value of £10 corresponds to the macro-state. Thus any given macro-state in a physical system, like pressure or temperature, can be realized by a certain number of microstates. *W* is then the number of ways in which a macro-state can be realized. If $W = 1$ the macro-state is in a state of least spreading, but as *W* increases the spreading of the micro-states increases, and so does entropy. The spreading of the energy states of the system is at all times accompanied by a distribution of the energy states in a phase space. More precisely, *W* is related to the number of arrangements in Γ-space, a six-N dimensional space, which specifies the microscopic configuration of an N-particle system. This description is needed to account for the motions of the individual molecules, in terms of momentum. Although it is customary to describe the increase in entropy as an increase in disorder, it is more appropriate to think of the increase in entropy in terms of the increase of *W*:

> A single system's phase point traverses the phase space in time as particles move and exchange energy. This traversal provides a graphic image of a system's particles spreading and exchanging energy through space, as time progresses. Energy, space, and time are all explicitly involved. In this sense, entropy can be thought of as a *spreading* function. (Leff 2007: 1750; italics in original; cf. Eddington 1932: Chap. IV)

To explain this spreading function further both the spatial spreading of processes and the temporal spreading of system states over accessible states must be considered (Leff 2007). For instance, in the case of the child's playroom the spreading of the toys resembles a spatial disintegration of a neat pile, but spatial spreading makes no explicit reference to energy balances. By contrast temporal spreading refers to the expansion of system states into the available phase space, as expressed in Boltzmann's statistical definition of entropy.[5] Boltzmann's statistical definition refers directly to the 'colonization' of the available phase space by the micro-states, since the Boltzmann entropy is a measure of the number of micro-states that are compatible with a given macro-state. Even though the equations of motion are time-reversible invariant, the spreading, say, of gas molecules into the available phase space occurs because there are many more ways of populating a larger than a smaller space. Only Loschmidt's Demon would be able to return the perfume molecules to the exact configuration in the perfume bottle. Given the spread of a system at a later stage, there is a negligible probability, in the lifetime of the universe, that the final conditions could be reversed. The physical realization of the temporal inverse of typical trajectories is negligibly small. Demons are generally unavailable.

The idea of the spreading of energy states can be applied to the past-future asymmetry. Any past moment corresponds to an amount of spreading over the available phase space, but the present moment from which the past is observed corresponds to a greater amount of spreading. As the present is an open system, we cannot manipulate the ensemble of micro-states to return them to their past configuration.

As we have seen, records are evidence of the branch-dependence of current states on past states. If past records are considered as decohered states—i.e. the result of an interaction with their environment, technically described as a loss of interference terms—then decoherence leads to an increase in entropy. Records reflect a differential in energy states between the present and the past. Records are evidence of past entropy states, but these are not necessarily ordered states—like a well-preserved fossil—but can also be disordered states—like a decomposed organism. One immediate problem with this suggestion is that, on purely statistical grounds, records of the past are as likely to have emerged from a higher entropy state by a fluke fluctuation as from a lower energy state (cf. von Weizsäcker 1937; Earman 1974: §7; Carroll 2010: Chap. 9). The notion of decoherence, which is an umbrella term describing various physical mechanisms, leads to different alternative histories and branch-dependence.

How can this proposal to ground the past-future distinction in differential entropy states be made compatible with the *t*-symmetry of the fundamental equations and the statistical character of the Second law? In order to deal with the

---

[5] In order to emphasize the importance of the spreading function, recall the above-mentioned Liouville's theorem, which states that the temporal evolution of classical dynamic systems preserves volumes of phase space regions but not their shapes. A rather uniform region in the initial stage can become very fibrillated after the spreading of the system.

objection just mentioned, it is important to realize that 'what is theoretically possible is not thereby practically probable'. In other words, some possible histories have a negligibly small probability of occurring, whilst others, like the familiar histories of classical systems, have a high probability. Thus, while all histories are equally possible, they are not all equally probable. Without the help of a demon, most systems have a negligibly small probability of returning to even their approximate initial states. The probability of a history depends on the level of its entanglement with other systems that act as a measuring environment. And decoherence needs to be paid for by an increase in entropy (Schlosshauer 2008: §2.8).

Why concentrate on entropy? First, it is a universal relation in the sense that no empirical violation of the Second law is known (Čápek/Sheehan 2005: 13, 42; Carroll 2010: 284). Even though it is only statistical in nature, according to statistical mechanics, it affects many other arrows of time. Second, it provides a dynamic explanation of the past-future asymmetry that was missing from other accounts. The fact that the Second law is statistical in nature may invite certain reservations. It means that, in theory, a state in higher entropy may spontaneously return to a state of lower entropy, and such an unobserved reversed process would be compatible with the Second law. Theory allows a broken cup to spontaneously reassemble itself. In fact, according to Poincaré's theorem a finite mechanical system, whose state S_o is characterized by the position (q_k) and momentum (p_k) variables of its micro-constituents, will return as closely as possible to the initial set of variables in Poincaré recurrence time. These considerations tell us that the Second law, in its statistical version, gives rise to *de facto* (not *de jure*) irreversibility.

De facto irreversible processes are to be understood as complex processes whose time-reverse is highly improbable in the history of the entire universe, although they remain theoretically possible. If they occurred, they would not violate the laws of the micro-processes in terms in which the macro-processes are understood. The fact that the reversibility of physical processes, if it occurs, is highly unlikely and does not violate the Second law, has been called weak t-invariance (or de facto irreversibility).

This weak t-invariance must satisfy the 'requirement that its time inverse (although perhaps improbable) does not violate the laws of the most elementary processes in terms of which it is understood' (Landsberg 1982: 8). This take on things implies that the t-invariance of physical laws is compatible with asymmetric solutions, if appropriate boundary conditions are taken into consideration (Price 1996: 88–89, 96; Denbigh 1981: Chap. 6.2). Thus it is not a violation of the Second law that a cold cup of coffee spontaneously reheats itself at some stage in the future history of the universe—through a fortunate self-rearrangement of the molecules—but such behaviour has never been observed. Nor is it expected to be observed in the remaining course of the history of the universe. But how improbable is the reversal of such processes? One aspect of an answer to this question is that the Poincaré recurrence time only exists for isolated systems in classical physics, but modern quantum physics emphasizes the importance of open systems:

> we know the reason why decoherence is considered irreversible for all practical purposes. To actually "relocalize" the superposition at the level of the system (i.e., to effectively time-reverse the process of decoherence), we would need to have appropriate control over the environment, which is usually impossible to achieve in practice. (Schlosshauer 2008: 255; cf. Griffiths 1994: 149–150)

Such arguments relate the probability of occurrences to the available number of realizable states, i.e. the inequality in the topology (or shape) of the initial and final states. If W is much greater for the later state, it reduces the probability of the coherence of final condition to return the system to its initial lower-W state.

To illustrate the negligibly small probability of the recurrence of a system to its initial state—in the absence of Loschmidt's Demon—note that the time scale for a Poincaré recurrence is of the order of $10^{10^{25}}$ years for a gram mole (Denbigh/Denbigh 1985: 140; Ambegaokar/Clerk 1999). Or to give an example on the cosmic scale, recall that according to cosmological estimates the amount of time it would take for a volume of gas containing 10^{18} molecules to return to its approximate initial state (position and momentum variables) would require $10^{10^{19}}$ years, which is well beyond the estimated age of the universe ($\sim 10^9$ years). As Eddington came to realize, an increase in entropy should not be used to identify the arrow of time, precisely because of the theoretical possibility of recurrence, say at the hands of a Loschmidt Demon. The empirical world displays a de facto irreversibility. But the temporal symmetry of the fundamental equations of motion means that physical systems exhibit de jure reversibility. It remains nevertheless true that de facto irreversible processes are so overwhelmingly probable that they can serve as a dynamical explanation of the past-future asymmetry in our familiar world.

If the increase in entropy is a de facto irreversible process, should the other arrows of time be grounded in entropy increase? Does entropy provide a master arrow of time? Would the arrow of time not reverse if a given system implausibly returned to its initial state? Answers to such questions will be discussed in Part IV on Nietzsche's and Landsberg's Demons. It will turn out that there is no master arrow of time. Time is multi-fingered. The discussion so far makes clear that it would be a mistake to *identify* arrows of time with entropy increases. It would be a mistake because the arrow of time is uni-directional but entropy is a statistical notion, which allows Boltzmann fluctuations. It was the purpose of Maxwell's Demon to show that the law of entropy is statistical in nature, since at least the Demon could reverse the inexorable increase in entropy. But the statistical nature of entropy does not prevent us from regarding entropy—or more precisely, the thermodynamic probability W—as one of the *criteria* from which the anisotropy of physical time can be inferred. There are many parallel processes in the universe—a rise in the entropy gradient of many systems, the cosmological expansion of the universe, the emergence of classical systems through decoherence mechanisms, the measurement of quantum systems—all indicating a past-future distinction. In addition, it must be recognized, as Popper's lake analogy illustrates, that the boundary conditions of the universe—in the present case the low-entropy initial conditions of our universe—must be taken into account when we seek an explanation of the arrows of time. As a matter of fact, boundary conditions are mostly asymmetric. It is no longer the case—as Part IV will show—that boundary conditions are merely stipulated since present-day cosmology is precisely concerned with explaining events like the Big Bang and its low entropy initial state. This research also strongly suggests that there is no evidence of a future low-entropy state of the universe. Even proponents of temporal symmetry and the Block Universe concede that *t*-symmetric laws may have *t*-asymmetric solutions (Earman 1967: 548; Price 1996: 88). When all these processes are taken into account, we possess reliable criteria from which to infer the anisotropy of past and future and arrows of time.

On this entropy account, then, we can explain several aspects of the past-future asymmetry:

1. A time traveller cannot go back to the past in ordinary time machines (Weinert 2013a: Chap. 4.7). Even as a non-participant observer he would interfere with the entropy balance of the past, which is fixed. In other words, time travellers cannot manipulate the conditions that would allow them to roll back the amount of spreading that has taken place between, say, 1921 and 1957. At the time of grandfather's youth, in 1921, the occupation of the accessible phase space had reached a certain configuration: going back to the past, even as an innocent bystander, would necessarily change this configuration. This cannot be done because what separates grandfather's view of the accessible phase space from that of the time traveller's—his grandson in 1957—is a change in the distribution of states. One could say that the change in the distribution of states—for which grandfather was causally responsible—made grandson possible. The trajectory leading from grandfather to grandson is not a spatial trajectory, on which we could travel up and down a certain number of times. It is a spreading of energy states, which cannot be undone, because the energy spent on the trajectory has changed the entropy balance of the states and this state cannot be regained by a manipulation of the more fibrillated later states. The spreading is also sequential. The time traveller cannot weave his way back through the multiple sequences of energy states because none of the energy spent between grandfather's time and grandson's time is recoverable, since only part of this energy will have been available to do useful work (as is clear from First Law of Thermodynamics). But even if it were possible for a Loschmidt Demon to reverse the spreading and restore a former energy state, the Demon's actions would not constitute a return to the past but a return to a copy of the past. It would reemphasize the essentially linear nature of temporal processes in the natural world.
2. While the state of spreading of the past cannot be changed, the future state can be affected, as illustrated in the branching tree model. From the point of view of the present state of the universe, we can change the spreading of states into the future because we are free to use the available energy today to channel it in particular directions. If we decide to make a cup of coffee we use the available

energy; if we do not, the energy remains available to do other useful work. We can choose not to open the perfume bottle or to stop its spreading by inserting a partition in the container. It will affect future states.

3. We have records of the past but not of the future because records are manifestations of the entropic state of the world of the past. Records may be either of ordered systems—like footprints on a beach—or of disordered systems—like a dilapidated building. When Lockwood stresses the concentration of nasic conditions in the past, for which there is no equivalent in the future, he presumably has such records in mind. We do not have such records of future states because (a) present spreading is contingent on past but not on future records; and (b) spreading is a progressive process contingent on the past history of the system.

The past-future asymmetry has thus been grounded in a dynamic explanation that refers to the Second law of thermodynamics. According to the entropy view, the past-future asymmetry exists objectively in the physical world because the energy balance of past stages of the world exists at a different level of 'spreading' than present and future energy balances.

Chapter 20
What Maxwell's Demon Tells Us and Does not Tell Us About the World

> *But I think the most important effect of molecular science on our way of thinking will be that it forces on our attention the distinction between two kinds of knowledge, which we may call for convenience the Dynamical and Statistical.*
>
> Maxwell (1873)

Maxwell's Demon introduces indeterminism and irreversibility into the conceptual landscape. Through the notion of indeterminism, Maxwell's Demon opposes the Laplacean identification of causality with determinism. Physical systems can be indeterministic and still obey a causal order. The two notions have to be kept apart. The notion of indeterminism changes causality from a deterministic to a probabilistic concept. Deterministic causality becomes a limiting case of probabilistic causality. As causal relations require energy to bring about the desired effects, entropy can be linked both to causality and the past-future asymmetry. Once determinism falls under a cloud of doubt, it becomes possible to review the mind-body question. Microscopic indeterminism does not have the resources to 'save' free will. But indeterminism may support a distinction between brain states and mind states. For the physicist the question arises how indeterminstic quantum processes may give rise to mind states. For the biologist the brain is a biological stochastic system, which may give rise to a Darwinian theory of the mind. For both approaches the mind is not simply identical with the brain: human consciousness may be the result of 'emergent' processes. As a materialist would require, the mind would then still be correlated with brain states, without being identical with them. But what does 'emergence' mean? So far all physical correlates between mind and brain, which have been proposed, provide at best only necessary conditions. Science does not (yet) tell us that the mind is an illusion.

What about the arrows of time? Irreversibility of initial conditions is an essential component in consideration of the temporal arrows. Recall the activities of Loschmidt's Demon. If only time reversibility of fundamental equations is taken into account, the Demon can in theory return all trajectories to the initial conditions, from which they started. But if the spreading from accessed to accessible phase space and energy considerations are taken into account, a return to initial conditions

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_20

in the real world becomes increasingly improbable. Maxwell's Demon destroys the 19th century identification of the arrow of time with entropy, since it shows that entropy is not a deterministic relation. But Maxwell's Demon does not show that entropy is a useless criterion for inferring arrows of time on a local and global scale. In fact once the focus has shifted to *criteria*, from which arrows of time can be inferred, other criteria come to mind: the accelerated expansion of the universe, the emergence of the classical world through decoherence, gravity, radiation and the disappearance of superpositions in the measurement of quantum systems.

Science does not tell us that time is an illusion, that we are trapped in a static Block Universe, that local arrows of time—the passage of time—are mental constructs. Only a particular interpretation of the theory of relativity leads to such conclusions. The very fact that different criteria may lead to different interpretations of the arrows of time shows that considerations of the nature of time are philosophical consequences of scientific theories. They do not follow deductively from their principles. Yet enough physical criteria exist to justify a distinction between past and future; these criteria tell us that there are local arrows of time.

Local arrows of time do not reveal the existence of a cosmic arrow of time. As remarked at the beginning, our awareness of asymmetric phenomena is compatible with different topologies of time. For instance, if the universe—our universe—expanded from a Big Bang event to a maximum point of expansion and then began to re-contract to end in a Big Crunch (which may be similar to the Big Bang event in terms of entropy), the inhabitants of such a universe would still experience the passage of time as asymmetric. The topology of such a universe would be a closed circle—a closed time-like curve—but every section of it would display the familiar asymmetry of our experience. However, the universe does not seem to close in on itself. The current understanding in cosmology is that the universe will expand forever; it is in fact accelerating towards its ultimate fate which the nineteenth century termed 'heat death'—the total dissipation of all energy gradients. This scenario suggests a linear topology, that our universe had its beginning in a Big Bang event but will expand forever. After having considered local arrows of time—in the present Part—it behoves us to consider global arrows of time. Does the universe itself exhibit an arrow of time? What does the expression 'the universe' refer to? It may refer to our familiar Milky Way or to the multiverse, of which our universe would only be one branch. Then the question repeats itself: does our universe display an arrow of time or does the multiverse exhibit an arrow of time? The difference drawn here between the local passage of time and the global arrow of time can be described as that between an asymmetry in local regions of space-time and the whole of space-time (Davies 1974: §2.1). Only a superhuman demon will be able to decide whether the Milky Way is part of a multiverse and whether this multiverse is static or dynamic. According to Nietzsche's Demon the universe is trapped in an eternal recurrence of events. It is a cyclic universe, in which events simply repeat themselves. But the notion of a cyclic universe is philosophically incoherent. If the universe is in fact expanding, with the prospect of a cosmic arrow of time, Nietzsche's Demon will have to make room for a more competent demon: Landsberg's Demon, who is able to survey the whole cosmic landscape.

If the universe is expanding it will exhibit an arrow of time, as is indeed suggested in newer cosmological models of the universe. In the following Part we will therefore weigh the prospect of an eternal return of events against the scenario of an ever-expanding universe, and its cosmic arrow of time.

Part IV
Nietzsche's Demon

Physics is an attempt at the conceptual construction of a model of the real world*, as well as its lawful structure.*

Einstein, Letter to Schlick 1930, quoted in A. Fine, *The Shaky Game* (1986: 97)

At the beginning of Part III, a model of a universe was considered, which—after a period of expansion—begins to re-contract and to return to its initial conditions. The inhabitants of such a universe would experience local arrows of time. But such a closed universe would not exhibit a global arrow since the final conditions are assumed to be identical with the initial conditions. In this respect, a closed universe resembles the static cosmos of the Greeks (geocentrism), which did not give rise to a cosmic arrow of time. Greek cosmology was split into two parts: between the moon and the 'fixed' stars the universe existed eternally, without change, displaying perfect symmetry. The passage of time was nevertheless associated with the motion of the planets, which could serve as a clock—for instance, by way of sun dials—to measure change in the immediate vicinity of the Earth. Other cultures and times favoured models of cyclic universes, which stipulate an eternal return of events.

Chapter 21
The Eternal Recurrence of Events

But it is interesting to think that our subjective distinction between future and past can ultimately be traced to the cosmological boundary conditions that distinguish the future and past of the universe.

(Hartle 2005: §IV)

Scenarios of cyclic universes have repeatedly been contemplated in the history of ideas. For instance, according to Nemesius, Bishop of Emesa, in the 4th century A.D. time performs a cycle:

> Socrates and Plato and each individual man will live again, with the same friends and fellow citizens. They will go through the same experiences. Every city and village will be restored, just as it was. And this restoration of the universe takes place not once, but over and over again – to all eternity without end. (…) For there will never be any new thing other than that which has been before, but everything is repeated down to the minutest detail. (Quoted in Whitrow 1989: 42–43)

An important consequence of such a view is that its proponents hold little regard for historical developments and evolutionary changes. In fact, there is no room for the notion of history in such a worldview of the eternal return of events. Since the Enlightenment, history has been conceptualized as a man-made irreversible process, which confronts us with new events on a daily scale. This view of history is very much a consequence of a dynamic view of the universe. Individual events and historic figures, which are regarded as particularly important for the progress of society, are singled out. But many archaic societies share the older view of a recurring universe. The collective memory in archaic societies works according to different structures: categories, instead of events, and archetypes instead of historic figures. 'The historical personage is assimilated to his mythical model (a hero etc.) and the event is included into the category of mythical actions' (Eliade 1954: 58).

The ontology of the archaic mind differs markedly from contemporary views as a consequence of the cyclical conception of time. An object or an action only become 'real' if they imitate or repeat an archetype (such as hunting, marriage, birth, death). That is, reality is acquired exclusively through repetition of or participation in an

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_21

archetype. Everything, which does not share an exemplary model, is deprived of meaning, i.e. it has no reality (Eliade 1954: 48).

Through the imitation and repetition of archetypes, Man is projected to a mythical epoch, in which the archetypes were first formed. This projection into mythical time only happens during rituals and important ceremonies. Profane time and the history of ordinary events are suspended when an object or gesture acquires a certain reality through repetition of paradigmatic gestures. Every action with a precise meaning (such as hunting, marriage) takes part in the sacred world, since it is the repetition of a model or archetype. Profane activities have no mythical significance, i.e. they lack exemplary models.

Thus two notions of time can be found in archaic societies. Profane time, in which people spend most of their lives, is devoid of meaning. But, thanks to rituals, they are able to leave profane time and enter mythical time (Eliade 1954: 50). What happens during these rituals is an imitation of an archetype. This imitation alone confers 'reality' on events.

The notion of a cyclic universe is not restricted to archaic societies. It has made several appearances since the 17th century Scientific Revolution.

In his book *The Gay Science* (1882), Friedrich Nietzsche introduces a demon who announces to the world an eternally repeating universe.

> What, if some day or night a demon were to steal after you into your loneliest loneliness and say to you: 'This life as you now live it and have lived it, you will have to live once more and innumerable times again; and there will be nothing new in it, but every pain and every joy and every thought and sigh and everything unspeakably small or great in your life must return to you, all in the same succession and sequence – even this spider and this moonlight between the trees, and even this moment and I myself. The eternal hourglass of existence is turned over again and again, and you with it, speck of dust.' Would you not throw yourself down and gnash your teeth and curse the demon who spoke thus? Or have you once experienced a tremendous moment when you would have answered him: 'You are a god and never have I heard anything more divine.' (Nietzsche 1882: §341; cf. Tipler 1980)

The eternal recurrence of all things echoes the ancient views of Nemesius. The universe is likened to an hourglass, which can be turned over indefinitely. In this respect it differs markedly from Eddington's use of the hourglass image of space-time. As we shall see shortly, Eddington argued that four-dimensional space-time harbours an arrow of time. Not so Nietzsche:

> Your whole life, like a sandglass, will always be reversed and will ever run out again,—a long minute of time will elapse until all those conditions out of which you were evolved return in the wheel of the cosmic process. And then you will find every pain and every pleasure, every friend and every enemy, every hope and every error, every blade of grass and every ray of sunshine once more, and the whole fabric of things, which make up your life. This ring in which you are but a grain will glitter afresh forever. And in every one of these cycles of human life there will be one hour where, for the first time one man, and then many, will perceive the mighty thought of the eternal recurrence of all things: and for mankind this is always the hour of Noon. (Nietzsche, 'Notes on the Eternal Recurrence' 1881)

Nietzsche refers to the ‘wheel of the cosmic process’, but he fails to draw a distinction between local and global arrows of time. Only his Demon could witness the ‘cosmic wheel’, whilst short-lived inhabitants of the hourglass universe would experience merely local arrows of time.

If local observers will never see the whole universe, from beginning to end, it requires the skills of a demon to survey the whole cosmos. Is it static or dynamic? Laplace’s Demon—with his belief in determinism—was able to perceive a unique chain of events, stretching both into the future and the past. But Laplace’s Demon can see neither the beginning nor the end of the universe, since the clockwork universe is infinite in time. Although the events within this Newtonian universe are dynamic, since they are subject to the laws of motion, Laplace’s Demon cannot inform us about the ultimate fate of the universe. The Laplacean Demon is unable to tell us whether the universe undergoes any evolution from, say, a beginning in a Big Bang to an end in a Big Chill. Nor did Newton, his contemporaries and classical successors envisage the possibility of a multiplicity of universes—a **multiverse**—a scenario which gives rise to the birth and death of many universes. Nietzsche’s Demon is committed to a cyclic view of the universe: the same universe repeats itself over and over again, down to the minutest detail. It will become apparent that Nietzsche’s Demon faces two problems. Philosophically a cyclic universe is incoherent. And factually, the universe evolves. In order to deal with the question of the fate of the universe and in fact alternative universes, the services of another demon are needed: **Landsberg’s Demon**. This Demon is able to stand outside and observe the workings of the multiverse. Landsberg’s Demon would be able to witness the birth and death of individual universes, as part of the ongoing story of a larger multiverse. Landsberg’s Demon would be able to see whether a universe expands, contracts and returns to its original state, whether it oscillates or evolves towards a heat death (a Big Chill). Landsberg’s Demon would be able to determine whether there exists a cosmic arrow of time: either for the whole multiverse or for individual universes, to which it gives birth. If such a Demon existed, residing outside of space and time—like Popper’s Demon (Part II, Sect. 11.2.2)—what would the universe look like to him? Human cosmologists can make some educated guesses.

One answer is provided by the so-called *Gold universe* model. It is a closed universe but with the additional feature of a flipping arrow of time. The physicist Thomas Gold (1962, 1974) argued that the thermodynamic arrow of time is closely coupled to the cosmological expansion of the universe. A Gold universe scenario assumes that initial and final conditions of the universe are identical. In the usual model, the Gold universe starts in a low-entropy past state and ends in a low-entropy future state, with the possibility that a future collapse could influence cosmic events today. It is usually assumed that a Gold universe undergoes a cycle from a Big Bang to a Big Crunch (Fig. 21.1a). But a Gold universe could equally start and end in a high-entropy state (Fig. 21.1b). In both scenarios it is assumed that the arrows of time reverse with the expansion or contraction of the universe. In a Gold universe model the cosmological arrow determines the entropic arrow. There are several concerns about such a Gold-type universe. First, there is no

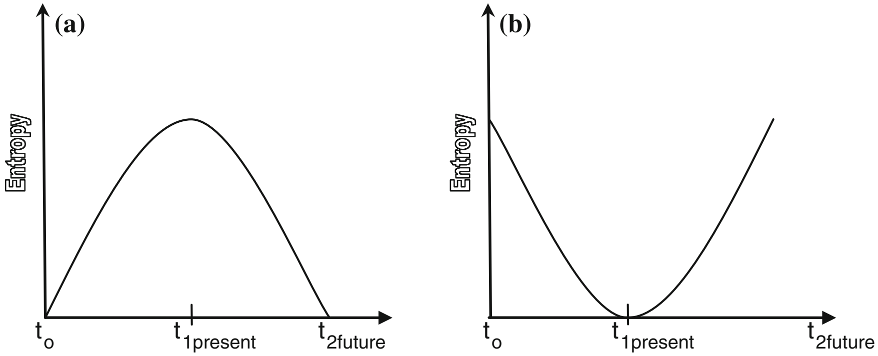


Fig. 21.1 **a**, **b** The Gold universe and its inverse

empirical evidence of the existence of a future low-entropy state, which would 'cause' the switch-over from an entropy-increasing to an entropy-decreasing universe (or vice versa). Such a switch-over would require the genius of a Loschmidt Demon. Loschmidt's Demon is a creature of classical mechanics; he can only reverse the velocities of the particles' trajectories. In theory the trajectories will return to their original state. But in an expanding universe, this return to the exact initial configuration is extremely improbable. Furthermore, a Gold universe of either type assumes that the arrow of time tracks the cosmological evolution. This assumption may make sense for the expanding phase but physicists now generally assume that entropy will increase even if the universe enters a contraction phase. The cosmic arrows will not reverse (Tipler 1980; Hawking 1994).

Gold universes are not usually associated with scenarios of eternal return. But it is easy to see that a Gold universe can be made compatible with a sequence of expansion and contraction phases. For the question arises whether, say, a low-entropy future would give rise to a new expansion phase or whether a high-entropy future would give way to a new contraction phase. Although one cycle of a Gold universe would, according to Gold, see a reversal of the arrow of time, does the postulated identity of past and future conditions mean that the universe would start repeating itself, and would such a recurring universe display a cosmic arrow of time? Or would the rather disheartening prospect of an ever-returning universe mean that there is no cosmic arrow of time? An answer will emerge if we contrast Nietzsche's Demon with a lesser known, but more competent demon: **Landsberg's Demon**.

Chapter 22
Landsberg's Demon

> *If time and space were not real features of the world it would have severe implications for us as agents.*
> (LePoidevin 2003: 245; paraphrase)

According to Bishop Nemesius and Nietzsche eternal recurrence means that the universe will repeat the events and their patterns indefinitely. A temporal analogy is the return of the seasons on Earth. If the cycles of the seasons are idealized, making every spring, summer, autumn and winter exactly alike, an analogy of eternal recurrence is obtained. The seasons perform a cycle, which goes through a finite number of steps from beginning to end; the end of one cycle constitutes the beginning of the next cycle. A spatial analogy is a runner who runs around a circular track a number of times. These analogies alert us to an incoherence in cyclic models, i.e. the fact that even a cyclic universe, as imagined by believers in an eternal return, will display an arrow of time. The seasons on Earth are due to the annual orbit of our planet around the sun and its tilt of 23.5°. Even if the seasons were perfectly identical, they would still succeed each other, as a demon stationed on the Sun could easily observe. The sun-inhabiting demon could count the number of orbits of the Earth around the sun against the background of the 'fixed' stars. Similarly, a spectator in the stadium could count the number of times a runner crosses, say, the finishing line. Although the Earth and the runner go through the same (idealized) phases, and thus the events appear as a recurrence, the occurrences are temporally ordered and hence constitute a sequence of happenings. If the universe performed an eternal recurrence, as stipulated by Nietzsche's Demon, 'the restoration of the universe' would in fact be a sequence of seemingly identical events. Even though the events are identical, as Nietzsche's Demon announces, they are not identical in a temporal sense. They are sequential. An archaic-like mind may believe the Demon's claim that life is just a never-ending cycle of identical happenings. Hence there is no dynamic evolution, no sequential history of events. But the archaic tribesman would do better to listen to Landsberg's rather than Nietzsche's Demon. He would realize that the return of events is temporally sequential from the point of view of a cosmic observer. This discovery may not be

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_22

of great comfort to the archaic mind but from a cosmic point of view even the eternal recurrence of events is based on a linear conception of time (Fig. 22.1).

Events may look identical to a local observer who may resort to mythical time if the reassuring words of Nietzsche's Demon fail to comfort him. But from the global point of view of Landsberg's Demon the seeming recurrence of events, as perceived by a local observer, turns out to follow a temporal succession. For a cosmic demon Nietzsche's 'eternal' recurrence displays an arrow of time (Fig. 22.2). This insight is important, for Landsberg's Demon surveys the large-scale evolution of the universe. Landsberg's Demon is able to ask such questions since he has moved on from the time of Laplace's Demon and realized that the universe is no longer Newtonian in character.

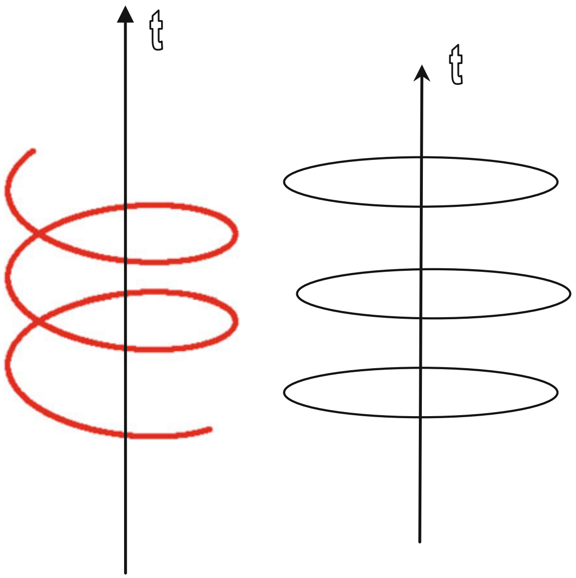


Fig. 22.1 Two temporal representations of the eternal recurrence of events. The events may be identical in every respect but their temporal location along a temporal axis, *t*, has a sequential character

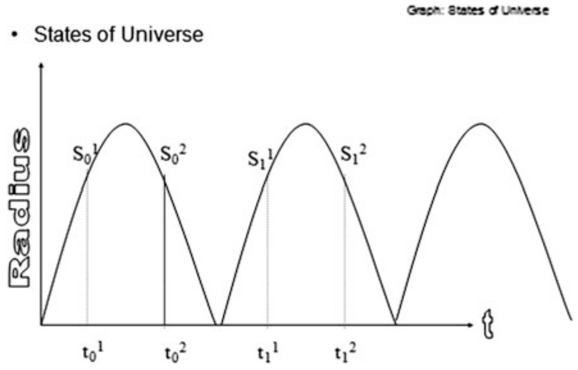


Fig. 22.2 Even though the events, $S_{0,1}$, seem to repeat themselves, they are in fact sequential in time

Inspired by relativistic cosmology, Landsberg's Demon would be able to tell whether the universe was contracting, oscillating or expanding. But what does the term 'the universe' mean? The expression may refer to 'our' familiar universe—called 'the Milky Way'—, which is commonly taken to have started its existence in a Big Bang, some 13.7 billion years ago. Or it may refer to the 'multiverse', according to which the Milky Way is one of numerous universes, which coexist on a higher cosmic plane (Fig. 22.3). Landsberg's Demon enjoys a panoramic view of the whole 'multiverse', rather than just the Milky Way. In the multiverse scenario individual universes—like the Milky Way—are born and die but the life cycle of the birth and death of universes may never reach an end.

22.1 The Multiverse

The question is whether Landsberg's Demon, a resident of the cosmic multiverse plane, would observe a cosmic arrow of time. The answer to this question depends on the nature of the multiverse. Models of the multiverse remain speculative, since Landsberg's Demon is reluctant to share his knowledge with human cosmologists. The latter have conceived the multiverse in a number of ways.

- The multiverse may be a self-reproducing, eternally existing cosmic landscape.
- The multiverse may consist of a succession of oscillating universes.
- The multiverse may be the cosmic mother of 'baby universes'.

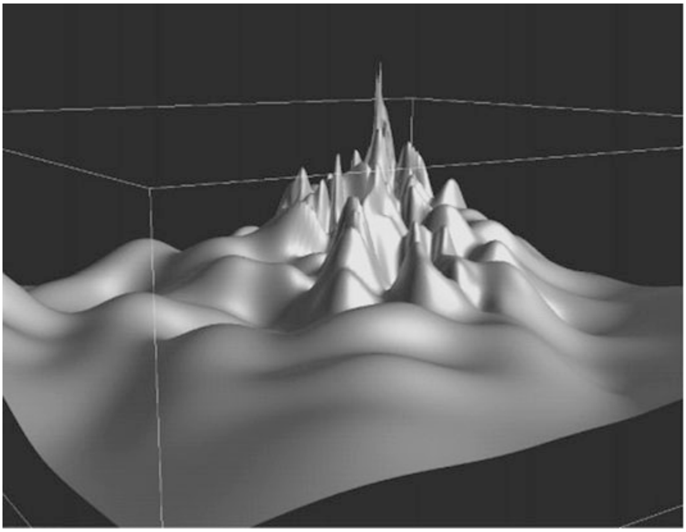


Fig. 22.3 A computer simulation of the eternal inflationary universe. The observable universe lies in the valleys, whilst inflation is still continuing at the peaks. *Source*: http://www.jrank.org/space/pages/2287/cosmological-inflation.html

In each case the question of the cosmic arrow of time arises.

(A) If the multiverse is an eternal, self-creating universe, with no identifiable initial or final conditions (Fig. 22.3), then Landsberg's Demon shares the fate of Laplace's Demon. Reminiscent of Plato's cosmology, both demons observe the local passage of time, in the birth and death of individual universes, but they have no way of establishing a cosmic arrow of time of the whole multiverse. The observation of a cosmic arrow of time requires that the initial and final conditions of the universe differ, which implies that dynamic change would have taken place between the two points in time. It is true that Landsberg's Demon, who surveys the whole of the multiverse witnesses the birth and demise of individual universes. This Demon sees volcanic activity. There is dynamic change, which is one of the prerequisites for the passage of time. But the change has no direction; there is no sequence of measurable events, which lead from an identifiable initial state to an identifiable final state. However, if Landsberg's Demon cares to focus on the fate of individual universes in this seismic landscape of the multiverse, he will see the birth and death of individual universes, like the Milky Way. He will see that the beginning of our universe is markedly different from its presumed final state. He will see the dynamic changes; hence he will see a cosmic arrow of time in individual universes.

The standard cosmological model of our universe is the *Big Bang* model, which makes the universe spatially homogeneous and isotropic. The latter property indicates that the Milky Way appears to look the same in every spatial direction. The Big Bang model assumes that the Universe had a definite beginning, and started from a singularity (a state of infinite temperature, density) in a gigantic explosion, approximately 13.7 billion (1.37×10^{10}) years ago. At the very beginning the universe was very hot (10^{110} °C), so hot that neither atoms, nor atomic nuclei or molecules could form; only quarks and other fundamental particles were present. But the universe began to cool very rapidly so that after the first 3 min, the early universe was filled with elementary particles (electrons, neutrinos, protons, photons). Over the next millions of years, the universe began to expand and cool so that ordinary matter—the first stars (after 150 million years), and galaxies (after 800 million years)—could form out of the soup of elementary particles. It took three billion years for our Milky Way to take shape. It is generally assumed that the entropy of the early universe was low so that its expansion is in accordance with the Second law of thermodynamics (Penrose 2005: §27.7, 27.13).

Modern cosmologists would like to ask Landsberg's Demon a number of questions: If the universe had a definite beginning, what will its ultimate fate be? Will it expand forever and end in a heat death—a Big Chill—or will it grind to a halt and re-collapse to a Big Crunch? According to the Standard Model its evolution depends on the amount of matter in the universe. As a threshold, cosmologists specify a critical density, ρ_c, which is the largest density, which the universe can have and still expand. Cosmologists then define a parameter, Ω, as a ratio of actual and critical mass density:

$$\Omega = \frac{\rho_o}{\rho_c} = \frac{actual - mass - density}{critical - mass - density} \tag{22.1)}$$

The value of Ω can be smaller than 1 (<1), equal to 1 or greater than 1 (>1). This value indicates the future fate of the universe (Fig. 22.4). If Ω is <1, then the critical mass density is greater than the actual mass density and the universe will expand forever. The question of expansion or contraction essentially depends on the amount of dark matter and dark energy in the universe. It cannot be directly observed, but its gravitational pull could be strong enough to bring the expansion of the universe to a halt and start the contraction phase. Current data indicate that the universe is actually accelerating so that at present the expansion seems set to continue forever. The value Ω = 1 signifies that the universe will expand but at a more steady rate. This value implies the total dissipation of energy, such that no energy differentials would be left to perform useful work and sustain life. The value of Ω < 1 indicates that the actual mass density in the universe is greater than the critical mass density, which implies that the universe would collapse to what is sometimes called the Big Crunch.

As illustrated in Fig. 22.5 an important question from the point of view of the asymmetry of time is whether the Big Crunch will resemble or even be identical

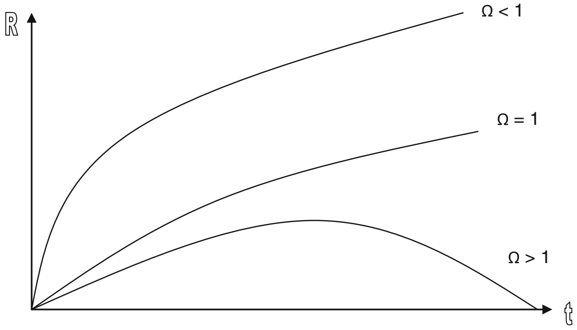


Fig. 22.4 Critical values for Ω and the corresponding evolution of the universe

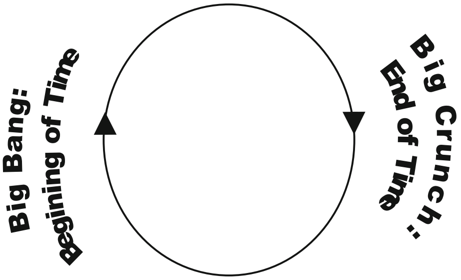


Fig. 22.5 The Big Bang-Big Crunch cycle

with the Big Bang. The Big Bang-Big Crunch scenario differs markedly from Nietzsche's eternal recurrence because the universe is destined to collapse to a Big Crunch without a chance of a resurrection.

Latest cosmological data practically rule out the return of the universe to a Big Crunch but the model remains interesting from the point of view of the asymmetry of time. If the Big Crunch means indeed a return to the conditions of the Big Bang, such that even a demon cannot distinguish the Big Crunch conditions from the Big Bang state, then no cosmic arrow of time could be detected. But this scenario is very unlikely, given the evidence. The evidence points to a cosmic evolution which will lead the universe—our Milky Way—to what the 19th century pessimistically called the 'heat death'. Such a heat death will surely befall our solar system. The solar system formed approximately 4.5 billion years ago. The sun, which makes life on Earth possible, has reached the half-life of its existence. The sun will provide energy to the solar system for another 5 billion years. It will then expand and eventually engulf all the planets before it collapses to become a white dwarf. All life in the solar system will be wiped out. Hence the beginning of the solar system will be in stark contrast with its end and this difference, supported by the dynamic changes, will constitute an arrow of time. Translated to the level of the Milky Way, this means that if there is a stark difference in the initial and final states of the universe, which Landsberg's Demon would be able to observe, our universe will display a clear cosmic arrow of time.

However, Landsberg's Demon surveys not just 'our' universe—whether Newtonian or Einsteinian—but the whole vista of the multiverse. Instead of seeing a static universe, Landsberg's Demon may observe an oscillating universe (Fig. 22.6).

(B) An oscillating universe is an extension of the Big Bang-Big Crunch sequence such that a dying universe, which collapses into a Big Chill, gives rise to a new universe in a sequence of time-ordered events. Globally this model seems to resemble Nietzsche's model of eternal recurrence but the events in an oscillating universe model clearly trace an arrow of time, since the earlier universe gives birth to a new universe rather than a copy of the old universe. The model of an oscillating universe faces the same challenges as that of a closed universe: what forces bring about the switch-over at the point of maximum expansion so that the universe can return to a low-entropy end point? And how is such a model compatible with the assumption of a universal increase in entropy over aeons of time?

Such conceptual difficulties have given rise to alternative scenarios of the multiverse: 'cyclic' universes and 'baby' universes. These speculative models do not agree with each other in respect of temporal symmetry but they adhere to the universal validity of the Second law of thermodynamics.

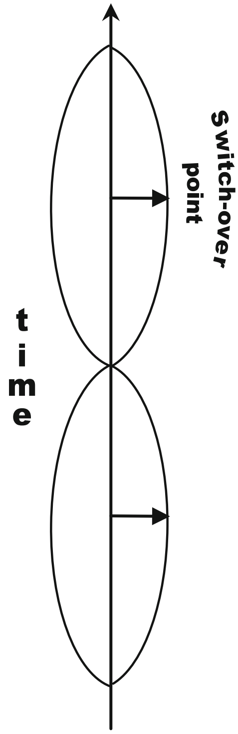
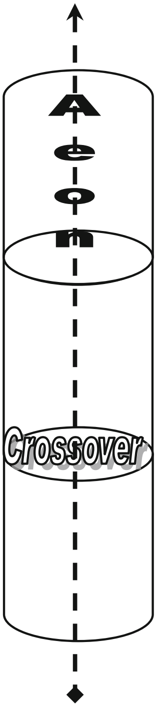


Fig. 22.6 An illustration of an oscillating universe

(C) A cyclic universe (Fig. 22.7) is not to be understood in Nietzsche's sense of the eternal recurrence of events. A cyclic universe displays a clear arrow of time. It consists of a sequence of aeons of time, not a rerun of an earlier universe. This whole conception of 'cycles of time' assumes temporal asymmetry, not just within one universe—one aeon—but along the timeline of the whole multiverse. What distinguishes the aeons from each other are 'crossovers', which separate an earlier from a later evolution. In accordance with the Second law, the model assumes that entropy increases without limit throughout the sequence of possibly infinitely many aeons. The end of one constitutes the beginning of another aeon. Each cycle, each aeon, begins with a morass of massless particles, rather than a singularity, as in the standard Big-Bang picture. Eventually it collapses into a massive Black Hole. These black holes themselves evaporate after 100^{100} years, leaving a sea of massless particles (like photons) and fields. This is not quite the heat death of 19th century physics since these particles and fields are capable of crossing the boundaries between aeons in order to begin a new cycle of time. Incidentally, in a universe occupied by massless particles and fields observers would lose the ability to make measurements of the passage of time. This cyclic universe model displays a cosmic arrow of time both for the multiverse and its individual universes since entropy increases without limit, possibly to infinity.

Fig. 22.7 Penrose's cyclic universe: time increases with the increase in entropy



The problem with this model is that the demand for ever-increasing entropy pushes the beginning of the universe far back into the infinitely distant past. As in the self-creating multiverse the universe needs to live forever. More troubling in this model is, however, that the beginning of each new universe requires a sufficient amount of fine-tuning to make the entropy small enough to serve as the starting point of a new universe (Carroll 2010: 345–53; Tipler 1980).

> It is the assumption of temporal asymmetry, which marks a decisive difference with the alternative view: *baby universes* (Carroll 2010; Gott 2001: 173–92). In this model, the universe as a whole is temporally symmetric but temporal asymmetry reigns within each baby universe (Fig. 22.8). Entropy increases within each baby universe but not over the whole multiverse. Recall that these models, however speculative, try to avoid the postulation of a Big Bang singularity, at which the laws of physics break down. In standard cosmological models the Big Bang was 'put in by hand' but the very existence of a Big Bang was the major conundrum.

The newer cosmological models attempt to avoid the Big-Bang singularity by introducing a pre Big-Bang phase of the universe. This phase then 'explains' the emergence of an individual universe, like the Milky Way. In Penrose's 'cycles of time' model the transition to a new universe happens at the crossover sections. But entropy increases without limit, providing a consistent cosmic arrow of time. By contrast, Carroll's 'baby universe' scenario adheres to the principle of temporal symmetry. Almost all the fundamental laws of physics display temporal symmetry

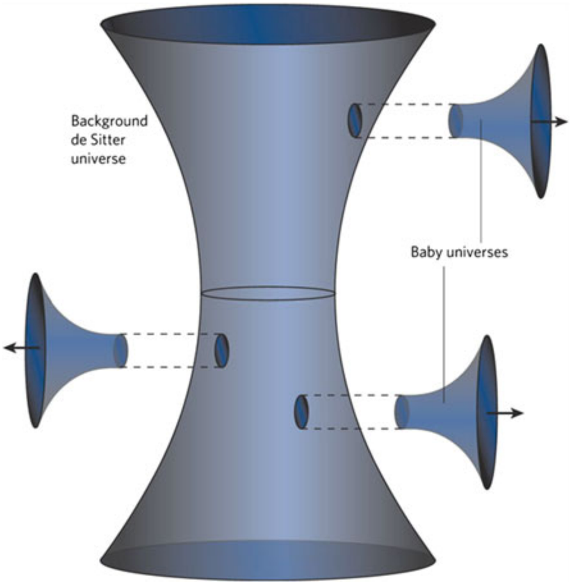


Fig. 22.8 A representation of a de Sitter space, giving birth to baby universes with opposite arrows of time. From *Nature* 440 (April 2006: 1132–1136; by permission of Nature Publishing Group)

(Part II, Chap. 10), which means that the equations allow both time-evolved and time-reversed states as possible solutions. This commitment to symmetry requires that the seemingly asymmetric character of the observable universe must be compatible with the principle of symmetry. One solution is the Gold universe, which postulates symmetric initial and final conditions (low or high entropy at either end) but this scenario is implausible. There is strong empirical evidence that the universe started in a low-entropy condition. Its acceleration means that it is unlikely to return to a low-entropy condition in the future. The Gold universe also requires that the arrow of time reverses when the universe begins to retract but even in a contracting universe the entropy will increase.

If the principle of symmetry is to be adhered to without violating the Second law of thermodynamics, the scenario of baby universes may provide the answer. This model satisfies two desirata. The multiverse displays overall temporal symmetry since baby universes exhibit opposite directions of time (Fig. 22.8). Within each universe, entropy increases from a low initial to a final state: thus creating an arrow of time. Baby universes are pinched off from an underlying de Sitter space-time, which essentially consists of empty space, filled with thermal radiation. Under the right conditions, a baby universe can then evolve to an ‘adult’ universe. Baby universes are the result of quantum fluctuations in the underlying de Sitter space.

> (…) the natural evolution forward in time is for space to expand and empty out, eventually approaching a de Sitter space. But from there, if we wait long enough, we will see occasional production of baby universes via quantum fluctuations. These baby universes will expand and inflate, and their false vacuum energy [non-zero energy] will eventually convert into ordinary matter and radiation, which eventually dilutes away until we achieve a de Sitter space once again. From there, both the original universe and the new universe can give birth to new babies. This process continues forever.

Note that baby universes do not give rise to an eternal return but produce arrows of time. It is the underlying de Sitter space, which exists forever, like the self-creating universe, without displaying an arrow of time.

> In the parts of spacetime that look like de Sitter, the universe is in equilibrium, and there is no arrow of time. But in baby universes, for the time in between the initial birth and the final cooling off, there is a pronounced arrow of time, as the entropy starts near zero and expands to its equilibrium value. (Carroll 2010: 362)

Overall this multiverse model respects temporal symmetry, since baby universes with opposite directions of time can be born out of the background de Sitter space. Each baby universe exhibits a local arrow of time, like the Milky Way but their arrows of time can point in opposite directions (Fig. 22.8).

How realistic are these scenarios when compared with space-time models of the actual universe?

22.2 Space-Time Models and the Universe

We have already come across Minkowski's notion of space-time, which is a four-dimensional representation of Einstein's Special theory of relativity (Part II, Sect. 11.2.1). The notion of four-dimensional space-time is often associated with the Block Universe but, as we found, it is not an unavoidable consequence of it. One important element is missing from Minkowski space-time: gravitation. As discussed in relation with Einstein's elevator thought experiment, Einstein replaced the Newtonian notion of gravitation by the notion of space-time curvature. In 1916 Einstein proposed a more realistic space-time model, which incorporated the curvature of space and time. Does this later model make room for an arrow of time? At first, Einstein's General theory of relativity modelled a static universe, which, like Newton's, was infinite in time but finite in space. But then the American astronomer Edwin Hubble discovered the recession of the galaxies. This discovery occurred independently of Einstein's General theory. It is expressed in Hubble's law, which states (in words) that galaxies in extragalactic space recede from each other, as seen from the Earth, as a function of their distance. The further a galaxy is away from the Earth, the greater its recession velocity. Although Hubble did not make this step, it follows from the expansion of the universe that in the distant past galaxies must have been much closer to each other. Ultimately the universe must have originated in some Big Bang event. The discovery of a dynamic universe was the result of an application of the equations of the General theory to the universe by Alexander

Friedman (1922, 1924) and Georges Lemaître (1927). Although the transition from a static to an evolving universe marks another significant shift in our understanding of the universe, it should be noted that an evolving universe is already part of the Kantian cosmology (1755). Kant explains the current state of the universe—its observable constellation of galaxies as a nested hierarchy—as a result of the work of Newtonian forces on some sort of original chaos. According to Kant's cosmology it took millions of years for the universe to form its visible cluster of galaxies through the effect of Newton's laws. The modern universe looks quite different from Kant's island universe, since Kant argued in terms of Newtonian mechanics, whilst modern cosmology applies the General theory of relativity and Quantum Mechanics to cosmological events. But Kant's island universe harbours as much an arrow of time as the various multiverse scenarios discussed in the last section.

According to the four-dimensional representation of space-time, as introduced by Minkowski's geometric model, the passage of time is a human illusion or at least a human construct. The reason for this conclusion is the time dilation effect of the Special theory of relativity: every reference frame carries its own clock and experiences time dilation effects according to the speed of the frame. There are therefore, in Pauli's words, as many clock times as there are reference frames. Observers attached to these reference frames do not agree on the simultaneity of events and hence there is no global Now, to which these observers could refer. As mentioned above, many physicists inferred from this geometric representation of space-time that the material world must be a Block Universe. The passage of time only exists in the human mind. One difficulty with this view is that it leaves a wide gap between the human experience of the physical world as a dynamic, evolving interrelated system and the static representation of this world in relativistic physics. A further problem is that this view ignores the many signposts in the physical universe, which strongly suggests the existence of the anisotropy of time.

In fact, even the geometric representation of space-time can be understood as harbouring an arrow of time. It is revealed in the light cone structure of space-time. Part II, Chap. 11 only considered one lobe of the representation in order to illustrate the notion of relative simultaneity of events. But a complete representation requires a consideration of both the past and the future light cones, on which events can be arranged in a number of ways.

If Minkowski space-time is modelled in the shape of an hourglass, events in space-time can be connected in three different ways:

When there is a *time-like* connection between them, the events are close enough in space and far enough in time for a connection to be established between them. They lie within the light cones as seen from **Here-Now** (Fig. 22.9). This connection is secured by mechanical means or sounds, which propagate at a speed slower than that of light; but they have the capacity to at least potentially establish a causal link between two events. For instance, if a signal is sent down a wire, it takes time for the signal to reach its recipient because emitter and recipient are separated in space.

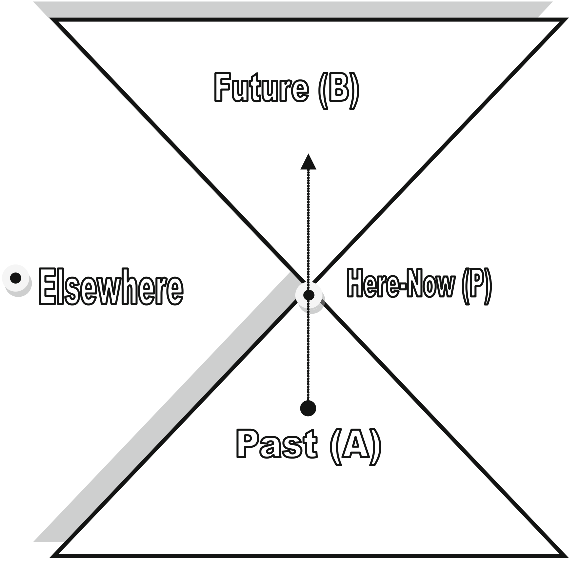


Fig. 22.9 Einstein's consideration of the (local) direction of time in response to Gödel's idealistic interpretation of the Special theory of relativity. A *time-like* world-line runs from event *A* to event *B*, where both points lies within, not outside, the light cone. *A* and *B* are linked by an irreversible signal (Einstein 1949: 687)

Time-like connections typically exist between the simultaneity hyperplanes in Minkowski space time. That is, for one event to causally affect another event time must elapse; these events can only be connected by the propagation of signals, or mechanical means.

When events have a *space-like* separation, they are too far apart in space and too close in time for any finite signal to connect them. Such events do not lie within the light cones that can be reached, by subluminal signals, from the **Here-Now**. They reside in **Elsewhere**. In cosmological terms such events lie so far away from the Earth that their light signals have not yet reached us. But space-like connected events are also all events, which coexist on a simultaneity hyperplane in Minkowski space-time but cannot be linked instantaneously.

Finally, there are the *null-like* connected events, which lie on the photon world-lines. In Minkowski's geometric representation, for these events time does not pass, since their space-time separation is always zero.

This representation makes no room for a world-wide objective Now but there is nevertheless a distinction between the past and future, between cause and effect. For signals emanate from the past and propagate into the future of the light cone. If the right structure is adopted, space-time becomes 'orientable', that is it possesses a global geometric feature, which permits a clear distinction between past or cause and future or effect and therefore a global arrow of time. Such a model of space-time forbids the occurrence of closed time-like curves, which would allow a world-line to curl back to its beginning. The geometrical feature of this hourglass model is such that it displays a continuous global distinction between past and future. It is compatible with the human impression of an 'objective' passage of time. As Eddington pointed out:

> By dividing the world into Absolute Past and Future on the one hand, and Absolute Elsewhere, in the other, our hour-glass has restored a fundamental differentiation between time and space. (Eddington1932: Chap. III, p. 50; cf. Wallace 2012)

Although Einstein's theory is often associated with the Block Universe, he was aware of the temporal orientability of space-time. As pointed out above the lack of absolute simultaneity in Minkowski space-time was taken to mean that there was no universal Now, and hence it was further inferred that there was no objective passage of time. In a characteristic thought experiment, Einstein considers the emission of a signal from a point *A*, whose source is located in the past light cone of an observer at Here-Now, *P*, to a point *B* in the future light cone of the observer (Fig. 22.9).

According to Einstein this process is irreversible. On thermodynamic grounds he asserts that a *time-like* world-line from *A* to *B*, through *P*, takes the form of an arrow, which sees *A* happen before *B*. Note that this order of events would be the same for all *time-like* related observers. According to Einstein this process secures the

> one-sided (asymmetrical) character of time (…), i.e. there is no free choice for the direction of the arrow. (Einstein 1949: 687; Einstein 1920: 139–41; cf. Zeh 1992: Chap. V)

Einstein hints at thermodynamic aspects of signal propagation in space-time, which usually are not included in discussions of the Special theory. The asymmetrical character of time is here described as a fundamental 'before-after' relation between events *A* and *B* but it is based on a physical property of signal propagation between these events, without recourse to an observer. There is an earlier emission event, at *A*, and a later reception event, at *B*, and it takes time for the signal to travel from *A* to *B*. Prompted by Einstein's thought experiment we should investigate what such thermodynamic processes tell us about the passage of time.

The first point to note is that a geometric model of space-time, with a clear time-orientability, must still be shown to be a realistic model of the actual universe. The model may be time-orientable but that does not mean that the world it represents is also time-oriented. As discussed in Part I, a model of 'reality' does not necessarily reflect reality, since it employs abstractions and idealizations. In other words, the temporal orientability of the model is merely a necessary condition for the definition of a global arrow of time. What is still needed is a physical criterion in the real universe to distinguish the two temporal orientations. As Einstein indicated such a physical criterion may be found in the energy flows of the universe, some of which are captured in the Second law of thermodynamics.

The second point to make is that Einstein's field equations describe a number of possible universes, including universes with closed time-like curves and Gold universes. Recall that such universes allow their inhabitants to clearly mark a local one-directional passage of time but not a global arrow of time. But amongst the possible solutions of Einstein's field equations are also so-called FLRW space-times (Friedman-Lemaitre-Robertson-Walker models), which are time-orientable and seem to represent the actual universe better than closed universe models.

The standard FLRW models of the universe are equipped with an upward-pointing time axis (starting from the Big Bang) and each model has a '1-parameter family of non-intersecting homogeneous space-like 3-surfaces T_t giving space at time t' (Penrose 2005: 719). One can imagine a demon, sitting on a co-moving patch (a time-slice), who travels along with the history of our observable universe. Such a demon would observe that the co-moving patch has an entropy value today of approximately 10^{101}, which is much larger than at earlier times (cf. Penrose 2005: 718; Carroll 2010: 334). Such a co-moving patch may help to define a universal Now—a cosmic time—in our universe, the Milky Way. But such a cosmic time is constituted by the average mass-density of the universe. It may not be useful as a clock but it constitutes an arrow of physical time.

Under the rubric of physical time we have identified both local and cosmic arrows of time. They are to be distinguished from the psychological awareness of time (phenomenal time) as well as human conventions of time reckoning (human time). On one interpretation of the philosophical consequences of the Special theory of relativity, the Block Universe is a static stack of simultaneity slices from which physical time is excluded (cf. Figure 25.1). There is no distinction between past and future. Minkowski space-time is not seen as time-orientable. On this interpretation, then, time consists solely of phenomenal time. However, the Block Universe view ignores several criteria, which could serve as indicators of physical time and its arrows. Human awareness of time is not captured in the static interpretation of Minkowski space-time. What is the relationship between physical and phenomenal time?

Chapter 23
Physical and Phenomenal Time

> *(…) experience makes it easy to confuse the egocentric now with something objective in a way experience does not allow that with the egocentric here.*
>
> Callender (2008: 360)

If physical time exists, it must be inferred from a number of physical criteria, which characterize temporal asymmetry in a physical sense, e.g. statistical-mechanical entropy, dynamic changes or the expansion of the universe. It is interesting to note that a debate in the metaphysics of time between Eternalism (Block Universe), Presentism (Moving Now) and Possibilism (fixed past, open future) has tried to muster the results of scientific theories (quantum mechanics, theory of relativity, thermodynamics) in support of these rival conceptions. According to Eternalism, past, present and future equally exist, while Presentism accords existence only to the momentary moving Now. The significance of the present seems to imply that there must be a unique Now, on which all observers can agree. In the language of space-time physics this claim amounts to the demand of a ‘unique hyperplane’ or ‘preferred foliation’. But it is difficult to identify such a unique foliation because of the problem of relative simultaneity. It is possible to identify a ‘cosmic’ Now in FLRW space-time models, which reflects the average distribution of matter in the universe. Possibilism requires the past to be fixed, the present moment to be distinguished and the future to be open (Savitt 2001). The motion of the present Now constitutes temporal becoming. This view is also known as the Evolving Block Universe, because it adheres to the four-dimensional representation of space-time, but bestows a dynamic history on it, which takes its start from the Big Bang and adds new time slices as the universe evolves. The latter interpretation suggests that the passage of time cannot simply be a psychological affair—an impression of flow confined to the minds of individual observers. As space-time does not depend on the existence of observers, the impression of flow must correspond to the occurrence of physical events in the real world, from which the passage of time is inferred.

One of the central experiences of our psychological awareness of time is temporal passage. A key question regarding temporal consciousness is whether the

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_23

phenomenology of the passage of time has a physical counterpart in a mind-independent succession of events. In other words, could there be a mental clock—an awareness of passage in the mind—without a concomitant awareness of a succession of physical events? Could there be phenomenal passage without physical passage?

In order to answer this question the conceptual possibility of a pure mental time must be explored. The investigation shows that a phenomenology of the passage of time—rather than a mere succession of experiences—cannot be understood without due consideration of the physics of time. In other words physical time is a necessary counterpart to phenomenal time.

Defenders of the possibility of mental time assume that humans would remain aware of the passing of time (a) even in the absence of any changes in the physical world (Lucas 1973: Part I, §2; cf. Newton-Smith 1980: 14; Shoemaker 1969: 367); (b) 'even if all our senses were prevented from functioning for a while' (Le Poidevin 2009: 1), i.e. even in the absence of sensory input. In this case humans would be aware of the passage of time, 'through the changing patterns of our thoughts.' As Eddington said, rather poetically:

> When I close my eyes and retreat into my inner mind, I feel myself *enduring*, I do not feel myself *extensive*. (Eddington 1932: 51; italics in original)

This assumption of pure mental time—that time is related solely to successive stages of conscious events—has two implications:

A. Private or personal time has to be granted primacy because, by implication, even in a frozen world or in a mind, deprived of sensory information about a dynamic world, an internal temporal flow continues.
B. Public time or clock time is derivative of private time, since it becomes the extension of the human mind (Lucas 1973: 37, §56).

Although the postulation of mental time, with its two implications, enjoys a certain *prima facie* plausibility, it leads to two problems, which have had reverberations throughout the history of time reckoning.

The first implication invites the question how a train of thought, which presumably constitutes private time, can occur in the absence of external input. The second implication throws up a problem, which is well recorded in the study of psychological time, namely that private time, as an extension of the mind, lacks both the regularity and invariance, which is characteristic of clock time.

Empirical studies have shown that different people judge the length of temporal intervals differentially, where this difference in temporal assessment may depend on such factors as age, mental states, moods, the influence of drugs (see Klapproth 2011; Treisman 1999; Fisher 1966; Hoagland 1966). This differential assessment has two aspects:

A. *Mental time lacks regularity.* Regularity means that the same interval repeats itself over a finite period of time. An internal clock, say a heartbeat, does not tick regularly so that it does not indicate the same temporal length for the same interval of time (as indicated by a clock whose minute- or second hand sweeps out the same area in the same amount of time, according to synchronized, triangulated clock readings). Hence there is no common yardstick by which one person, *A*, could compare the length of a given interval on separate occasions. Thus the assessment of the length of a temporal interval, say between the beginning and the end of an event, *E*, differs from interval to interval. *A*'s judgement lacks regularity.

B. *Mental time lacks invariance*. Two test persons, *A* and *B*, cannot compare their respective assessments of the lengths of an interval given on two separate simultaneous or successive occasions. It is not *invariant* since a switch from person *A* to person *B* is not guaranteed to reproduce the same temporal evaluation. By contrast if *A* and *B* are equipped with synchronized, well-functioning clocks, they will agree, barring accidental mistakes, on how much time has passed for a given event. But the lack of invariance of mental time is not restricted to *A* and *B*. The same person, say *A*, will not be able to judge whether the same temporal interval on two different occasions is the same, even if *A*'s internal clock ticks rather regularly. *A*'s judgement lacks invariance for *A* cannot compare two separated intervals in mental time.

Mental time is therefore severely limited: it seems to indicate temporal succession to individual minds but this assessment of succession lacks both regularity and invariance. Yet, despite the variation in individual assessments of temporal periods, which are empirically confirmed, phenomenal experience tells us that, under normal circumstances, there is some overlap or loose agreement between individual observers, as far as their temporal judgements are concerned. This intersubjective agreement about the simultaneity of events and their linear order seems to be due to temporal integration mechanisms in the brain (see Part IV, Sect. 23.3). In cognitive models of such mechanisms temporal realism is assumed, i.e. that the world has objective temporal relations. It is worth emphasizing, as will be discussed, that a regular sense of succession is presupposed in most models of temporal awareness, even in some idealist models. The question therefore arises how this awareness of regular passage may emerge.

23.1 Temporal Realism and Anti-realism

Some modern thinkers hold that a pure mental time is possible. The Oxford philosopher J.R. Lucas, for instance, holds that both subjective and public time have the order type of real numbers (Lucas 1973: §§2, 6; cf. Penrose 1995: 385). Real numbers are used to compute temporal intervals. Galileo's fall law $-y = v_o t + 1/2at^2$—allows *t* to take the value of, say $\sqrt{\pi}$. Thus the equation allows *t* to take on irrational values, even though no conventional clocks could measure them. Subjective time also has the order type of real numbers. But what is invariant in intersubjective time is of the order-type of rational numbers. Thus people agree about the temporal order of their experiences but, due to the changes in the perception of the length of temporal intervals, people do not agree about the metrical properties of their temporal experience (Lucas 1973: 19). More significant in the present context, however, is that Lucas holds that all order types result in a complete ordering of events such that an event E_1, which occurs earlier than an event E_2, satisfies a clear linear order: $t_1 < t_2$.

Thus Lucas attributes two characteristics to mental time: it has a real number metric and satisfies a complete ordering relation. The question is whether mental time satisfies these criteria. Before this question can be addressed, a distinction between temporal realism and temporal anti-realism should be drawn.

For current purposes, let a temporal realist be someone who holds that the experiential passage of time is grounded, in a way to be discussed, in the succession of physical events in the external world. Let a temporal anti-realist be someone who holds that phenomenal passage is *sui generis* and does not require a basis in the physical world. Hence, for a temporal antirealist, a pure mind can experience the passage of time as long as there are ‘low-level sensory flows’ (Dainton 2011: 384).

Temporal realists are entitled to make two assumptions: (1) that the perception of passage is regular—where this regularity is grounded in the regularity of the external world and preserved in the mental processing of information—and (2) that it leads to a complete ordering or a linear succession of events. But it is not obvious, as will be discussed below, that temporal anti-realists are entitled to these assumptions. Without these assumptions temporal anti-realists must grant that the perception of temporal ‘passage’ may be regular, random or chaotic and still maintain that it is appropriate to call even an irregular succession of mental thoughts ‘the passage of time.’ Such an understanding of passage runs counter to a long-established understanding of the lapse of time. Plato called the planets ‘the instruments of time’ because in Plato’s geocentric worldview, their circular motion is objective, perfectly regular and invariant. Newton’s notion of absolute time states that ‘absolute, true and mathematical time, of itself, and from its own nature flows equably without regard to anything external’ (Newton 1960: 6). The passage may not be measurable but even an irregular passage may mark a flow of time. Even though mental time lacks all these characteristics, the anti-realist may still insist that it marks ‘passage’. Even if flow is neither regular nor invariant it must mark an order of time, a before-after relation, in the mind. But now the question arises whether a complete and regular ordering can be established for ‘private’ time.

Psychological studies seem to confirm that even though each person is conscious of a different amount of elapsing time between events, depending on their psychological states, they all agree on the order of events (even in a relativistic setting for time-like related events). This linear order criterion may not be satisfied in highly mentally disturbed patients who may confuse past and future or live permanently in the present. Such psychological exceptions may not disturb the anti-realist about temporal passage, since it seems still to be true generally that ‘the topological properties of different people’s temporal experiences, are invariant but not the metrical properties’ (Lucas 1973: 19). Furthermore, beyond this agreement on temporal order, if not length, it also seems that, phenomenally, we assume the existence of a common Now. However, such observations do not help the anti-realist about temporal passage, because the possibility has not been excluded that a linear order or a common Now may be due to the work of external physical time. For instance, the impression of a unique, even universal Now, may be explained by the work of temporal integration mechanisms in the brain. It is generally assumed in neuro-science that the brain processes information from the external world, and hence that there is a world-to-mind direction of fit. Such a fit cannot be assumed if passage resides solely in the mind. What an anti-realist account must assume, in order to defend temporal flux, is a well-functioning and reliable memory, memory storage of past events and the ability to distinguish past from present events in a linear order. Saint Augustine held that the mind provides a topological order of time: we remember the past, perceive the fleeting present and anticipate the future. For only in this way can a succession of events—not its metric but its topological properties—be established in an anti-realist model of temporal flow. However, as the next section will show, a consideration of the implications of 'passage' in a pure mind poses serious difficulties for the postulation of pure mental time.

23.2 Memory and Entropy

Let us start with a thought experiment, according to which the existence of a pure, disembodied mind is assumed, preoccupied with the contemplation of its own thought processes, without input from the external world.

If the awareness of temporal flow is to be a 'concomitant of the mind' alone, memory of past *real* events cannot be assumed to exist to help the mind establish a 'before-after' relation between thought events. Memory is usually characterized as self-referential, i.e. it has a world-to-mind direction of causation. We only remember an event if the real experience of the event causes the present memory of the event (see Searle 2010: 38). But could a pure mind, contemplating the changing pattern of thought, use *imagined*, rather than *remembered* past events, to establish the passage of time? One immediate difficulty arises: just as the experience of passage is not the same as the succession of experience (Dainton 2010, 2011) an imagined past is not the same as a real remembered past. Although an imagined past is part of the mind's awareness, it is not a recall of a past event. However, just as the revisit of a place, which was seen before, is by necessity a later visit, could not the mental revisit of a thought or an imagined past, which has occurred to the mind before, establish a rudimentary form of temporal passage? Admittedly such a recurrence of thought would fail to satisfy the criteria of regularity and invariance, which are normally associated with temporal flow, but to which the pure mind is not entitled. The recurrence of a thought would constitute a very primitive topological 'before-after' relation. And thus it would be correct to say that 'minds are necessarily located in time' (Lucas 1973: §2).

In our thought experiment we imagine a mind—a disembodied mind, perhaps in an otherwise empty universe—which revisits an original thought, say *T*, indefinitely—like a runner running eternally around a circular track—and ask whether phenomenal passage can exist without physical passage? The question is posed not in the sense of a neglect of physical passage, but in the sense of an absence of physical passage. The original thought, *T*, is the type and its recurrences are the tokens, t_n. Would a pure mind orbiting a particular thought type, and revisiting it again and again in the form of its tokens, experience a 'passage of time'?

It is tempting to conclude that the train of thought or the successive reflections of the same thought, *T*, constitutes a passing of time in a pure mind. But even a pure mind, in order to establish a passage of time, needs to establish a linear chronology of thought tokens, t_n, as well as a well-functioning memory. The ability to recall a particular thought—in its detail but perhaps in a different guise—requires memory so that the remembered thought becomes a real—albeit imagined—thought that happened prior to its remembrance.

How could our hypothetical pure mind keep the chronology of the past in linear order? A first suggestion may be that the thought tokens, which occur to it each time, must be numbered: t_1, t_2, t_3, …, where the contents of *T* is always identical, even though it occurs in different guises, and only separated by the order-type of, say, real numbers. But this procedure will soon fail since the pure mind will also need to remember the metric, i.e. the numbering of the recurring thought tokens. Unfortunately, memory offers no more reliance than the act of remembering itself—hence it will happen that the pure mind mis-applies the metric to the recurring thoughts, thus messing up the linear chronology. (It is well known how notoriously difficult it is to remember the sequence of past events from memory alone.) So a pure mind faces the problem not only of remembering past tokens of the thought but also of keeping the tokens in a linear order. Linear order is essential for the experience of passage, since otherwise no clear distinction between private and physical time could be made (cf. Norton 2010).

A pure mind will not be able to procure a linear order of the thought tokens because its applications of a postulated metric are defective. More can be said about the proceedings of the pure mind if we further assume, as is customary in the philosophy of mind, that it is not a free-floating entity, but a process, which is related to the brain. Under this assumption a mind-brain system, engaged in its spiralling exercise of revisiting the thought tokens, needs energy, both for the exercise itself, and for the storage of memory items. The thought tokens must be remembered as the particular thought tokens, which were contemplated before and ordered in a linear fashion. But where there is energy, there is entropy. And entropy is commonly regarded as one indicator of the arrow of time. The question arises whether the imaginary mind could take its increase in entropy as an indicator, not of a metric but a topological order of events.

In the popular literature entropy is usually characterized as an increase in disorder or a loss of information. Thus biological organisms are born and decay, hot liquids grow cold and do not reheat on their own. But, as we have seen, it is more precise, following L. Boltzmann's insight, to characterize it in terms of the number of ways in which a (physical) system can spread through an available space (often called a ‘phase space’) or increase the ‘realizability’ of its possible states (Landsberg 1982: 75). Thus a crystal consisting of 8 atoms has only one way of realizing a temperature near absolute zero, namely when none of the atoms vibrate. But if some energy is supplied to one of the 8 atoms, enough to vibrate, there will be eight possible ways of doing this, since the energy can be supplied to each of the eight atoms so that its energy increases. (Whilst in a ‘frozen’ state it is known that each atom is ‘still’, in a minimally vibrating state it is no longer known which of the eight atoms is vibrating.)

From these examples it can be concluded that there is a difference between the (phase) space, which a system currently occupies and the (phase) space, which is available to it for the realizability of its states.

In terms of thinking and memory, the realizability (or spreading) of states means that the mind-brain[1] system is an information gathering and utilizing system (IGUS), which draws information from the external world and processes it (Hartle 2005). But the external world undergoes change, through the use of energy, which is accompanied by an increase in (overall) entropy. This increase in entropy, from past states to future states, necessarily affects the IGUSes in terms of their ability to form records. As it turns out entropy will pose further serious problems for the anti-realist about temporal passage, if the following hypothesis is taken into account:

> It can be argued that memory can only work when entropy increases in the direction from the memorized event to the time of recalling it. (…) In this way, one expects that the observer’s subjective experience of the flow of time always coincides with the direction in which the observer’s entropy increases. (Kupervasser et al. 2012: 1181)

How does this hypothesis affect the imaginary mind? It is a premise of the thought experiment that the imaginary mind, by stipulation, does not process external data but its operation still requires energy both for the act of remembering and the storage of memory data. The longer the imaginary mind contemplates and repeats the hypothetical thought *T*, the more memory is built up but with the unavoidable consequence of an increase in entropy and the loss of information about the ordering of t_n. Could the pure mind not use its own entropy processes as an indicator of the passage of time? Unfortunately, the hypothetical mind could not measure its own entropy. Entropy processes are not measured directly but are inferred from external macroscopic parameters, like temperature, pressure and work. Our hypothetical mind, in order to measure its own entropy gradients, would

---

[1] Above it was argued that mind processes are not to be regarded as identical with brain activity. We regarded it as a reasonable hypothesis that mind processes ‘emerge’ from brain processes but that much work is left to be done to identify a physical correlate of notions like ‘emergence’. What is assumed, in line with most contemporary philosophy of mind, is that the mind is not completely separate from the brain. For more on the mind-brain relationship, see Searle (2004) and Part II, Chap. 12; Part III, Chap. 17.1.

have to refer to external parameters, which, by the premises of the thought experiment, it cannot do.

Without some simple metric it will soon be impossible for the hypothetical mind to know how often thought, T, occurred to it or whether one token, t_n, occurred to it sooner or later than another token. (It will be neglected here that the hypothetical mind can only store a limited amount of information so that by adding new versions of T it will have to erase old ones, which itself requires energy.) That is the hypothetical mind loses its grip on passage. It cannot establish a linear order. As is well known from everyday experience without some external referent—some metric or dating system—it is often impossible to know, from mere memory, whether event E_1 occurred before or after E_2. As the impression of passage is purely internal to the hypothetical mind, losing grip on passage also means that it loses the ability to distinguish between remembering and anticipating an imaginary thought or its tokens. For there is no way of distinguishing a remembered event from an anticipated event, if it is purely imaginary.[2]

It seems that in order to secure the linear order of events some external reference point is needed, against which such processes as remembering, numbering and memory storage can be measured. The reference point may be established in terms of some external events or in terms of some entropic process. But the imaginary mind of our thought experiment does not satisfy the criteria Lucas imposed:

1. The hypothetical mind may be in possession of a metric, even a real number metric, but it needs to remember the order it imposed on the thought tokens, which is impossible from memory alone.
2. If the imaginary mind cannot remember whether token t_n occurred before or after token $t_{\pm\mathrm{n}}$, it cannot establish a linear order, not even of a partial kind, let alone a complete ordering of the tokens. That is, a pure mind cannot satisfy the criteria of regularity and invariance. The introduction of a second disembodied mind, inhabiting the same empty universe, will re-emphasize the difficulties. These two disembodied minds would face the same difficulties as psychological test persons, *A* and *B*, who only experience a private time. They will not be able to agree on the succession of events in their respective minds. If some external events, *E*, were to occur in their universe, these two pure minds would not be able to agree on the succession of *E*, and hence would not be able to establish a linear, invariant order of events, *E*.
3. Brain processes, like remembering and the storage of records of past events, necessarily lead to entropy gradients, which could establish an arrow of time. But such entropy gradients themselves need to be recorded, stored and

---

[2] Experimental investigations of the 'stopped clock illusion' (Yarrow et al. 2001) suggest that perceptual experience of a given event at a certain time, t_1, may be influenced by later events. According to the authors the experimental 'data support ideas of conscious experience as an ongoing, often post hoc reconstruction emerging from multiple cognitive systems' (Yarrow et al. 2001: 304).

remembered, which by (1) and (2) cannot be achieved without some external referent.

If the passage of time implies regularity and invariance, a reference to external events is required. But if these conditions are relaxed and no 'serial order of moments' is associated with the passage of time, the hypothetical mind of our thought experiment is only capable of a random succession of experiences but not of an experience of lapse of time.

According to the thesis cited above, 'the observer's subjective experience of the flow of time always coincides with the direction in which the observer's entropy increases.' This thesis presupposes that the observer is a physical system or at least a mind-brain system, in which entropy increases. The observer grows older and will experience changes in their environment, which clearly establishes that physical time passes by these criteria. But in the thought experiment, introduced above, the mind is not embodied. This extreme idealization means, as argued, that such a being must lose track of the flow of time. This conclusion implies that the idealization must be relaxed. Memory, recording events and remembering are acts, which require energy—real energy—, the entropy of which could be used to gauge the passing of time. If an embodied mind is allowed, again being preoccupied with the recurring contemplation of a particular thought, *T*, there now exists a physical reference point, against which the passing of phenomenal time can be assessed. The increase in entropy is one criterion—but not the only one—of an arrow of time. But physical time is only a short-term expression for the regular succession of events in the external world. The succession of events is registered in the brain by a flow of information, which is processed in the temporal mechanisms. Human minds are information gathering and utilizing systems (IGUSes), which process information from the external world; this process causes an experience of the passage of time.

'The flow of time is the movement of information into the register of conscious focus and out again' (Hartle 2005: §2). Hence phenomenal time—as the experience of passage—presupposes physical time—as the regular succession of events in the external world. There must be temporal mechanisms in the brain, which process information and contribute to experiential flow, i.e. to the impression of the flow of time and of a universal Now.

23.3 The Impression of Flow and a Universal Now

According to the Block Universe interpretation of Minkowski space-time, the passage of time is a human illusion. There is no room for the impression of flow, the march of time, in the static space-time representation of the physical world. And yet the impression of flow is one of the most fundamental of human experiences. If it were a mere illusion, it would be unlikely to be both a common *and* universal feature of human awareness. There is an unwelcome disconnect between theory and experience: On the one hand, according to a dynamic interpretation, there is *physical time*. In Leibniz's words, it is the order of succession of events. This order of succession of events—this local arrow—can be observed by humans in their environment: it may take the form of a general impression of the degradation of useable energy. Physical time also includes the global arrow of time, which can be inferred from the observation of cosmic events. On the other hand there is *phenomenal time*—the time of human consciousness. It distinguishes between past and future, and perceives the flow of time. These two notions of time must be connected: firstly, as argued above, the notion of phenomenal time presupposes the notion of physical time; secondly human beings are information-processing physical systems; thirdly human beings have the ability to reflect on the clues offered by nature and to construct models of time. But our models of time must be realistic—they must take the evidence from the physical world and from science into consideration.

From this point of view, however, the theory of relativity suffers a deficit. It simply does not account for the existence of phenomenal time. But a concentration on phenomenal time would be insufficient too, since it tends to ignore the results of science. We therefore need to consider the origin of the impression of flow and the common Now from the point of view of 'the physics of Now' (Hartle 2005).

Recall that our visit to the land of demons suggested, amongst other things, that the concept of time was an inference from a number of physical criteria, which could serve as signposts for the identification of both local and global arrows of time. The reason for this assumption was the fact that time, although ubiquitous, is unlike other physical parameters. Energy, mass and velocity are both observable and measurable properties of physical systems. But time is not a directly observable or measurable feature of the physical world. The notion of time needs to be inferred from regular and invariant processes in the physical universe. This need to infer leads to agreed clock time but also to different, even contradictory philosophical models of time. Whilst all competent observers agree on notions like mass, energy and other physical parameters (like clock time), they do not agree on philosophical models of time. Regularity and invariance make time measurable. It is therefore important to distinguish *physical* from *phenomenal* time. In addition it is convenient to introduce the notion of *human* time. Physical time is the order of succession of events. Phenomenal time is the direct awareness of the passage of time, which, as argued above, presupposes physical time. Human time is the calendar time of the social world; let it include the philosophical models.

Phenomenal time is also a shared experience: humans agree on the notion of Now and the flow of time. If that is the case it is not unreasonable to expect to find a link between physical and phenomenal time, between physical and human time, despite the fact that physics—or the space-time representation—makes no room for the impression of flow. As we have seen 'past', 'present' and 'future' are not properties of four-dimensional space-time, since it only makes room for the agreed succession of time-like and null-like related events. But humans are information-gathering and utilizing systems. As such humans are part of space-time and observers of the material universe, which they represent as a four-dimensional space-time. If that is the case notions like the 'flow' of time from past to future or

'Now'cannot be completely divorced from physical reality; they must be consistent with it; hence they cannot be completely illusory. The near universality of the impression of 'flow' must be due to the ability of conscious beings to register processes, like the dynamic change of events, the transition from exploitable to non-exploitable energy or the regularity of periodic processes, and to infer from such processes the notions of past, present and future, as well as dating systems for the use in human societies.

The notion of *human time* is a conceptual consequence of the existence of intelligent observers in a law-regulated universe. There is no doubt that many regular physical processes exist in the physical world. Consider for instance the orbit of the Earth around the sun or the daily rotation of the planets on their axes. These are periodic processes, which recur after a certain interval. More precise processes are found in the oscillations of atoms or the regular pulses coming from celestial objects. Such periodic processes constitute the basis for the measurement of time. It is not sufficient for periodic processes to be regular—i.e. divisible into a finite sequence of equal segments—they must also be invariant. The invariance of regular periodic processes means that a change in location or time will return the same value for the measurement of the interval.[3] If your wristwatch is a reliable indicator of time in one part of the world, then a displacement of the watch to another part—a change in location—will not affect its regularity. The Greeks observed the orbits of the then known six planets with astonishing precision, given that their observations were based on the geocentric worldview. The change to the heliocentric worldview, two thousand years later, did not fundamentally change the knowledge of the orbital periods. A translation in time from the Greek to the modern period kept the orbital time of the planets invariant (unchanged).

Let us adopt Leibnizian language: physical time is the order of the succession of events and occurs irrespective of the absence or presence of conscious observers. How can we construct the various notions of time from this position? Physical time is based on natural units, as provided by regular, periodic and invariant processes in the material world. The dinosaurs no doubt lived their lives unregulated by calendars but the Earth still orbited the sun in a periodic fashion. It requires the presence of conscious observers to construct a notion of time from the observations and choice of periodic processes. Thus human time emerges. It is partly based on conventions. Whilst physical time is based on natural units, human time is mostly based on conventional units of time. The 7-day week, introduced by the Romans, the subdivision of the day into 24-h, of the hour into 60 min and of minutes into 60 s, the division of the year into 12 months and the lengths of the months into 30 or 31 days (except February), introduced by the Romans, are all conventional units of time. They are conventional because they respond to human social needs about time reckoning although there may be no physical processes, to which they

---

[3] The passage of time can be observed from irregular processes—the motions of particles in a liquid, and of mass molecules in a container, the movement of pedestrians in a city—because they constitute sequential change but such processes do not give rise to a measurable passage of time.

correspond. To give an example, the beginning of the year (1st January) is purely conventional, since there is no natural event, which would single out this particular date. Equally the beginning of the day at midnight is a convention. Already the Babylonians introduced the 7-day week and named the days of the week, like the Egyptians, according to the sun and the known planets: moon, Mars, Mercury, Jupiter, Venus and Saturn (Wendorff 1985: 118). The division of the year into twelve months (4000 BC) was inspired by the 12 orbits of the moon around the Earth in one tropical year. But this creates a problem of time reckoning because the time between lunar phases is only 29.5 Earth days but the solar year has 12.368 lunar months (Zeilik 1988: 152; Wendorff 1985: 14). As a consequence, the length of the month is now purely conventional and no longer related to the lunar month. The division of the day into 2 × 12 h is explained by geometrical considerations. During the summer only 12 constellations can be seen in the night sky, which led to the twelve hour division of day and night. According to the sexagesimal system, there are 10 h between sunrise and sunset, as indicated by a sundial, to which 2 h are added for morning and evening twilight (see Whitrow 1989: 28–29; Wendorff 1985: 14, 49). When the year and the day are set to start also depends on conventions and social needs. In ancient Egypt, for instance, the year began on July 19 (according to the Gregorian calendar), since this date marked the beginning of the flooding of the Nile (Wendorff 1985: 46). In the late Middle Ages there existed a wide variety of New Year's days: Central Europe (December, 25); France (March, 21; changed to 1st January in 1567); British Isles, certain parts of Germany and France (March, 25) (Wendorff 1985: 185; Elias 1988: 21f). Note, however, that not all such conventions are arbitrary. The equinoxes, the summer and winter solstices are important dates in the calendar, yet they correspond to particular positions of the Earth with respect to the sun.

The division of the globe into 24 time zones (1884) reveals an interesting mixture of both natural and conventional units of time. Each time zone covers 15° of the globe, which corresponds to one hour. Across each time zone the minutes remain the same; the 24 meridians are only separated by one hour (both to the west and east of Greenwich). The sun's apparent journey across the globe constitutes a natural unit of time; but it is a convention to make the minutes the same within and across time zones. Strictly speaking, each location has its own local time, as determined by the position of the sun at this particular location. Prior to the introduction of universal time, clock time differed in each location both according to the hour and the minutes. For instance, when it was noon at Greenwich, it was 12:09 p.m. in Paris, 12:50 p.m. in Rome and 13:35 p.m. in Athens; but it was 11:46 a.m. in Madrid and 11:23 a.m. in Lisbon.

Despite these aspects of conventionality, it must be emphasized that the conventional units of time must keep track of natural units of time, as is the case with equinoxes, solstices and time zones. For otherwise, conventional units of time will fall out of step with the periodicity of the natural units. The measurement of time is inseparably connected with the choice of certain reference frames, like the 'fixed' stars, the solar system, and the expansion of galaxies or atomic vibrations (Clemence 1966: 406–409). It was one of the great discoveries of Greek philosophy to have realized that there exists a link between time and cosmology. The existence of conventional units of time thus presupposes the existence of natural units of time. Whilst one would expect agreement amongst conscious observers about the periodic physical processes, i.e. agreement on physical time, if human time is an inference from observations of these processes, it is natural to expect differences in dating systems and indeed philosophical consequences drawn from the observation of physical processes. Such differences in calendars do exist in different cultures, as do differences in philosophical views about the nature of time. Eternalism, Presentism, Possibilism are different conceptual consequences drawn from the physics of time. If human time is thus linked to natural units of time, it may be expected that phenomenal time is equally based on physical time. Phenomenal time signifies the direct human awareness of flow, passage and a universal Now.

> The flow of time is the movement of information into the register of conscious focus and out again. (Hartle 2005: 3)

A universal Present is not a feature of space-time yet humans as information-gathering and utilizing systems seem to share a universal Now. Is this a contradiction? The physics of space-time tells us that there is no common Now and yet a community of human information users readily accepts a common Present. Users of phenomenal time do not doubt that the observer on the pavement shares with the motorist a sense of simultaneity. What the motorist perceives as Now—the traffic light jumps to ‘red’—the pedestrian equally sees as Now, and crosses the road. There is no contradiction. Recall that relativistic effects only manifest themselves at speeds close to the speed of light. Although strictly speaking a pilot ages less than his non-travelling twin brother, the effect is so minute that the two brothers keep approximately the same age. In the same way an approximate common Present can be characterized. The Special theory of relativity includes the principle that all inertial reference frames are on an equal footing as far as the description of physical events are concerned. For an observer at rest the clocks of a fast-moving observer seem to slow down; but the fast-moving observer can be regarded as being at rest, in which case the clocks of other observers are now seen as slowing down. But observers moving at ordinary speeds, far below the speed of light, can single out particular reference frames, on whose clocks they agree. These clocks define a common Now.

So the physics of Now can explain the common Now without positing ‘being present’ as a phenomenal property. Even though humans have at present various experiences, this experience does not mean that they experience a property called ‘presentness’ (Callender 2008; Dorato/Wittmann 2015). According to Saint Augustine the present does not exist because any moment in time can be subdivided into smaller intervals of time. Uttering the name ‘Saint Augustine’ takes time: strictly speaking the sound has a past, a present and a future. Even the restriction to one syllable can be broken down into a past, present and future segment. By this reasoning Saint Augustine arrives at the counterintuitive suggestion that the Now does not exist. But in observing the ‘present’ human observers ignore time-lags to a certain extent and regard as ‘present’ what strictly speaking is a succession of impressions. The construction of an extended Now is helped by temporal integration mechanisms in the brain, which are similar in each person. What helps this impression is the endurance of macroscopic objects in our environment. Observers operate with numerous local Nows—time spans of enduring objects or events. Even a 100-m sprint is perceived as occurring Now because it is close enough in time and space to be perceived as simultaneous. But even spatially distant events—on the other side of the world—can be perceived as simultaneous, because they are part of the reference frames, on which observers on both sides of the spatial divide have agreed. This agreement no longer holds for very distant locations, like the Andromeda galaxy (2.2 million light years from Earth). Although we may naively assume that the clocks tick the same everywhere in the universe, the theory of relativity has taught us otherwise. Whilst it is easy and reasonable to assume that my local Now is the same as the local Now of someone on the opposite side of the Earth, the confidence is easily lost when we consider events on the other side of our galaxy. What is now for us may not be now for them! Within a certain spatial circumference the intersubjective agreement of local Nows leads to the impression of a global Now, which is perceived as objective (Callender 2008; Butterfield 1984).

In this way the explanatory gap between physical and phenomenal time can be bridged. It confirms that it is indeed important to distinguish between physical and phenomenal time. Phenomenal time is the 'processing of information from the external world' in brain mechanisms. Human time is a conceptual construction from the observation of the succession of physical events in the universe; the latter constitute physical time.

To return to Landsberg's Demon, he is much less concerned with phenomenal or human time than with physical time. His field of operation is cosmology and in particular the question of the arrows of time. If space-time is time-orientable, both the Demon and his human apprentices can draw a clear distinction between past and future. But Landsberg's Demon casts a wider eye. What evolution will the universe undergo: will it be a cyclic, contracting, expanding or oscillating universe—or a de Sitter universe, given birth to baby universes? Is it an Evolving Block Universe, embedded in a larger multiverse?

Chapter 24
The Evolution of the Universe

> *The running away of the galaxies does not mean that they have a kind of aversion from us.*
>
> (Eddington 1935: 210)

According to the Minkowski space-time representation the passage of time does not fit into the objective world. Rather the physical world is represented like a map, which contains all the world-lines of all material particles, including photons, which define the boundaries of the light cones. The light cones then give rise to the distinction between time-like, space-like and null-like connections. Although it is possible to insert a distinction between past and future even in the representation of Minkowski space-time, by making four-dimensional space-time time-orientable, there is no room for the flow of time, which therefore is regarded as a purely subjective impression (see Weinert 2004, 2013a for further discussion). As noted before, space-time becomes time-orientable by making the arrow of time an intrinsic geometric feature of space-time. But such a choice should be grounded in physical criteria. One such criterion is the actual expansion of the universe. With its expansion the universe becomes a sink of radiation so that entropy increases in the direction of the expansion. The expansion itself is represented as the diverging world-lines of galaxies, which correspond to Hubble's law. When this idea was first proposed (Gold 1962, 1966, 1974), it was assumed that a contraction of the universe would result in a decrease of entropy. The arrow of time would therefore reverse. The universe would return again to a low-entropy Big Crunch, apparently lending support to Nietzsche's Demon. Such a cyclic evolution, if theoretically possible, would require the assumption of a future attractor-state, which breaks the trend, according to the Second law, of the ever-increasing entropy of the universe. There is little evidence of a Nietzsche or Loschmidt Demon at work on a cosmic scale. If there is a demon, it must be Landsberg's Demon.

The British astronomer Arthur Eddington introduced the expansion of the universe as an additional criterion (apart from the local increase in entropy) for the cosmic of arrow of time. But Eddington did not assume that the universe would return to its original low-entropy state. He regarded both the 'dissipation of energy' and the 'expansion of the universe' as irreversible processes, and hence as signposts

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_24

for the arrows of time. The expansion of the universe is a global signpost, whilst the increase in entropy is a 'unique local signpost'. But how do we know, despite the lack of evidence, that at the end of the universe no Loschmidt Demon lies in wait with the intention of returning the universe to its earlier low-entropy state? Landsberg's Demon surely knows that a scenario, which is theoretically possible, does not thereby become probable. Human cosmologists think in terms of typical and atypical processes. A return of the universe to an earlier state would be extremely atypical (only a theoretical possibility, although not excluded by the laws of physics, as Loschmidt pointed out), whilst the expansion of the universe is revealed as typical, statistically relevant behaviour in a dynamic universe. As Eddington argued, if an expanding universe is taken into consideration we are no longer forced to conclude 'that every possible configuration of atoms must repeat itself at some distant date'.

> In an expanding space any particular congruence becomes more and more improbable. The expansion of the universe creates new possibilities of distribution faster than the atoms can work through them, and there is no longer any likelihood of a particular distribution being repeated. If we continue shuffling a pack of cards we are bound sometime to bring them into their standard order – but not if the conditions are that every morning one more card is added to the pack. (Eddington 1935: 68; Zeh 1992: Chaps. 5 and 6)

Thus Eddington seems to argue that, as the occupied phase space of a configuration expands into the accessible phase space, the probability of a return to initial conditions becomes less likely. As more of the available phase space is taken up, a Gold universe becomes increasingly improbable. Human scientists may ask Loschmidt's Demon:

> (…) why not reverse all the molecules' velocities so the system will go back to its initial state?

But

> (t)he answer is again probabilistic: among all possible velocity distributions, the ones returning to the initial state have a negligible weight. (Omnès 1999: 239)

A Gold universe scenario not only faces the switch-over problem, it is also confronted with the evidence of an expanding universe. In fact the latest cosmological discoveries show that the universe is expanding at an accelerated pace, which makes a universe with closed time-like curves extremely implausible. The last century has seen a dramatic shift in our views of the universe. At the beginning of the 20th century the universe was seen through the Newtonian prism as static and eternal. Kant's cosmological island view of the universe (1755) derived from an application of Newton's principles to cosmology and Laplace's nebular hypothesis (1796). It contained an arrow of time. At the beginning of the 21st century the established view is that of an accelerating expansion of the universe. The vision of the universe becomes that of an inhomogeneous distribution of galaxies—galaxies are ordered in superclusters—and small anisotropies in the cosmic background radiation (Brandenberger 2013). These observable asymmetries themselves show that the cosmic material displays asymmetries, which could serve as criteria for the anisotropy of time.

If it is assumed that the universe will indeed expand forever, the question becomes what its ultimate fate will be? This is a question which can be addressed to Landsberg's Demon. The contracting universe scenario has lost much of its credibility. A Gold universe is an extremely unlikely scenario. But Landsberg's Demon should be able to confirm what future astronomers will see. Surprisingly, they will see that neighbouring galaxies—like the Andromeda galaxy, which lies approximately 2.5 million light years from our own Milky Way—will have merged with the Milky Way; and other galaxies will have disappeared from their view and 'escaped beyond the event horizon' (Krauss/Scherrer 2008). Future astronomers will even have difficulties to observe the Big Bang because the cosmic background radiation will have become unobservable. Ironically, for

> these future astronomers the observable universe will closely resemble the 'island universe' of 1908: a single enormous collection of stars, static and eternal, surrounded by empty space. (Krauss/Scherrer 2008: 3)

If the expansion of our universe has such dramatic effects, what else does the Demon see, far into the cosmic future? Will the Demon confirm that the universe is an evolving rather than a static Block Universe? Is the Big Bang really the beginning of the universe or only of our universe? And when our universe has evaporated into a 'heat death' will new universes be born out of it? Landsberg's Demon has a vantage point from which he can survey the vast volcanic landscape of the multiverse. But human observers need to rely on the available evidence.

First, at the present moment no available empirical data suggest that a final condition exists, which may have an attractor effect on the state of the current entropic level. According to current cosmologies, the actual universe seems to be characterized by asymmetric boundary conditions, which have an effect on its actual behaviour. There are 'inhomogeneities and anisotropies in the later universe', which are not present in its early phase (Brandenberger 2013: §1; Mukhanov 2015). The universe starts in a low-entropy Big Bang but is expected to collapse into high-entropy black holes, which will eventually evaporate to a state of even higher entropy. The relaxation time, τ, of thermodynamic systems seems to be much shorter than the lifetime of the universe, T, which according to current calculations of the evaporation of black holes seems to amount to a lifetime of approximately 10^{100} years. Furthermore, there exists a large entropy gradient between the earlier entropy (10^{88}) and the final entropy (10^{120}) of the universe such that the arrow of time would not reverse (cf. Carroll 2010: 63).

But the Demon needs to tell us more. He needs to soar above the plane of the multiverse to inform us not just of the ultimate fate of 'our' universe but whether the phoenix of new universes will rise out of the ashes of the dying universes. And here indeed the Demon envisages tantalizing scenarios: future cosmologists will no longer put the Big Bang in by hand; a low entropy past will no longer be just stipulated. The Big Bang as the unexplained ultimate beginning may not have occurred as earlier cosmologists assumed. Rather the Big Bang may be a transition phase from a dying to a new universe. It will explain how our universe and other universes came into being. Every universe, which arises from the multiverse, will expand and eventually die and in this process mark a cosmic arrow of time. The Demon does not share his knowledge with today's cosmologists. They are obliged to consider various scenarios on the basis of the available evidence: cyclic universes, baby universes, oscillating universes. Whatever fate will eventually befall the universe it seems clear that cosmic evolution marks a sequence of events, in which the initial conditions are markedly different from the final conditions. And this difference, brought about by dynamic changes, is the unmistakeable signpost of a cosmic arrow of time. Time and change are therefore intimately linked.

Chapter 25
Time and Change

Things of this World are in so constant a Flux, that nothing remains long in the same State.
John Locke, *Two Treatises of Government* (CUP 1989; ed. P. Laslett: Part II, §157)

The association of time with change goes back to ancient Greece. Plato declared the planets 'the instruments of time' because, according to the Greek geocentric worldview, the planets moved in circular orbits around the central, stationary Earth. Both Aristotle and Saint Augustine rejected the identification of time with change because, as they argued, motion is measured in terms of time but time is not measured in terms of motion. Motion can be slow or fast or a condition of stasis obtains and still time passes. Nevertheless both Aristotle and Saint Augustine recognized in change an important factor for the measurement of time. Time and change imply each other, since time could not be measured without change, and change could not be measured without time. Humans are used to change in their familiar environments. In fact, the whole multiverse, as well as its individual bubble universes, is a stage for dramatic dynamic change, too. For today's cosmologists, the planets are no longer the 'instruments of time'. More reliable instruments have become available. The demon on the co-moving patch can observe the dramatic changes that have taken place since the Big Bang: the formation of solar systems, of galaxies and their clusters, all in motion with respect to each other.

But is change a necessary or sufficient condition of the passage of time? In other words, could time exist in the presence of temporal vacua, i.e. periods of empty time without any change taking place? If change were a necessary condition of temporal passage, then, in the absence of change, time could not exist. This conclusion follows from a relational view of time (Leibniz). But if change were merely a sufficient condition of temporal flow, then it is conceivable that time may pass, even in the absence of external change. The latter option would concede the possibility of temporal vacua, as does the absolute view of time (Newton). Or time may pass purely in the mind, as the train of thoughts, which would make time a concomitant of the mind (mental time).

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_25

It has been argued that external change is a necessary and sufficient condition for the anisotropy of time. Both the construction of human time and the awareness of phenomenal time are linked to physical time, which is understood here in the relational sense of time.

Such a relational view of time has been developed by Gottfried W. Leibniz, the German polymath. According to Leibniz:

> *Time is the order of non-*contemporaneous *things*. It is thus the universal order of change in which we ignore the specific kind of change that has occurred. (Leibniz 1715: 202; cf. Wheeler 1967)

Time, in other words, is the order of succession of material events. As Leibniz believed in a mechanical universe, with a clockwork regularity, he assumed that the order of succession was both regular and invariant. Time does not exist as a physical parameter in its own right. We have no direct knowledge of physical time and its direction. Physical time must be inferred from the criteria, which the physical universe offers to its observers. Time, according to Leibniz, is not absolute; it is a relation between material events in the universe. Some events are more appropriate as criteria for the establishment of time than others; e.g. the increase in entropy or, to some extent, the expansion of the universe. The events in the material universe constitute time: the universe, according to Leibniz, is a clock (Fig. 25.1).

The relational view fits in very well with the conception of physical time. But it seems to rule out *empty* and *possible* moments of time: could there be moments of time when no events occur? In regard to *empty* time, several scenarios can be distinguished:

1. We can imagine a totally frozen universe, in which entropy has reached its maximum value: there is no change, not even atoms oscillate. This universe has

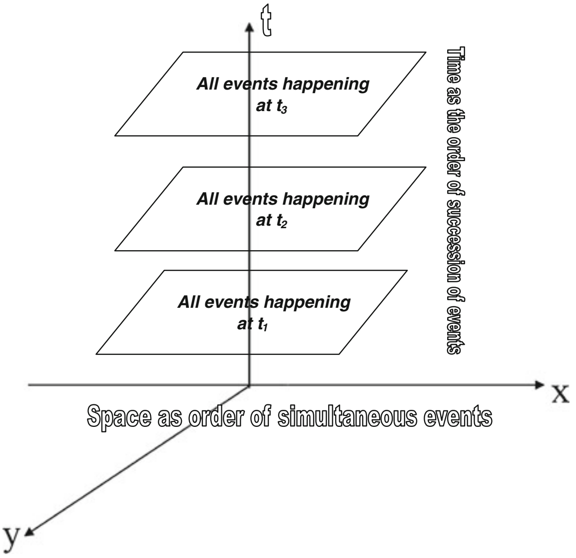


Fig. 25.1 A Leibnizian view of space and time

suffered a 'heat death'. In the absence of all change there is no experience of the march of time, let alone the possibility of measuring change. On the relational view time ceases to exist. In such a frozen universe, not even mental time would exist because human brains would be frozen too.

2. We can imagine a partly frozen universe, which evolves from some event sequence $E_0 \mapsto E_1$ but is then interrupted by a period of temporal vacua when events simply freeze and grind to a halt. After this partial freeze the universe resumes normal service and events $E_3 \mapsto E_4$ continue. In such a scenario time elapses between E_1 and E_3 even though it remains unknown and unmeasurable how long the freeze has taken. In this scenario then there is a passage of time but, similar to chaotic motion, the lapse of time cannot be measured because the link between E_1 and E_3 is unknown and could be irregular.
3. Finally we can imagine frozen time in some locality, against a background of ongoing periodic regularity. Some of Jupiter's moons may be such icy wastelands; nevertheless their temporal existence is measurable because a moon's frozen state can be measured against the ongoing regularity in other parts of the universe.

Even on the relational view, it is logically conceivable to speak of *empty* time, but a period of changelessness is an unlikely event in a dynamic universe, with little explanatory value. Time is intimately related to change, either in the local neighbourhood or the cosmic landscape. When a multiverse is a changeless foam, as the de Sitter universe or the self-reproducing universe, no arrows of time emerge. To accept empty time would either involve the stipulation of mental time, unrelated to physical time or the postulation of Newton's absolute time, according to which the universe has a clock irrespective of material events. Mental time, however, presupposes physical time. Newton's absolute time is notoriously unobservable. Even Newton, who believed that his theory of mechanics needed a notion of absolute time, which is not related to any material events in the universe, was obliged to assume physical approximations to absolute time in the 'fixed' stars and Jupiter's moons (Weinert 2013a: Chap. 2.8.3). The Special theory of relativity introduced the notion that physical time is clock time. But clocks are instruments, whose regularity is ultimately based on some invariant physical regularity in the physical universe. Physical time is therefore much closer to the relational view of time than Newton's absolute time.

We have arrived at the conclusion that models of the arrow and nature of time are inferences from the observation of periodic, regular and invariant physical events in the universe. Physical time is ultimately the order of the succession of events in the universe. The question then arises whether one arrow of time may be more fundamental than other arrows or whether some arrows of time can be derived from other arrows. We have already seen that the causal arrow of time can be related to the increase in global entropy. Does entropy provide the master arrow of time? Arthur Eddington at one point certainly held that the Second law of thermodynamics—the increase in entropy in isolated systems—constituted an irreversible arrow of time. He even considered that the Second law held 'a supreme position among the laws of Nature' (Eddington 1932: 74). Although Eddington later introduced the expansion of the universe as a further signpost of the arrow of time, Eddington's early identification of the increase in entropy with the arrow of time can be read as an attempt to define a master arrow of time. Was Eddington right?

Chapter 26
Is There a Master Arrow of Time?

> *These considerations [about decoherence] also rule out the conceivability of a Laplacean demon (who is assumed to observe the world as an external system without reacting upon it). While classically allowed, a quantum demon would have to decohere the world.*
>
> (H.D. Zeh ^{2}1996: 97)

It has already been observed that the *identification* of the anisotropy of time with physical processes like entropy, dynamic changes or the expansion of the universe leads to mistakes. *De facto* the universe seems to be expanding at an accelerated pace and is therefore unlikely to return to its original state. If the expansion or contraction were to serve as the basis for the identification with the arrow of time, a uni-directional arrow of time would be ruled out. But such a scenario would face the switch-over problem. One would have to find a dynamic reason why the universe, after reaching its maximum extension, would suddenly begin to return to its initial low-entropy state. But even if the universe were to re-contract its entropy would continue to rise: for instance, the sun would still radiate its energy into cold space and all the normal processes, which suffer from a loss of useable energy, would continue as usual. Thus even if the universe re-contracted, the arrow of time, in terms of entropy, would not reverse. If, by contrast, the entropy gradient were to be used to *identify* the arrow of time, the problem is that entropy is only a statistical concept and hence entropy fluctuations are to be expected. But we think of the arrow of time as uni-directional and not subject to short-time reversals or deviations. It is ill-advised to use any physical criterion to *identify* the anisotropy of time.

It was proposed that local and cosmic arrows of time are to be conceived as theoretical constructions from the available criteria. Then a cluster of these criteria can be used as signposts to mark arrows of time. One astonishing phenomenon is the *parallelism* of the arrows of time. Recall that according to the causal arrow the cause precedes the effect; according to the psychological arrow the past is remembered but the future is anticipated; according to the radiative arrow, radiation emanates from a source and dissipates in an outward direction; according to the cosmic arrow, the universe expands and will not return to its initial Big Bang

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_26

condition; and according to the entropic arrow, systems have a tendency to transit from an ordered to an unordered state by way of using energy which cannot be recovered. However different these arrows are, they are uni-directional. In theory some of these arrows may be reversible, in practice they are irreversible. This parallelism of the many arrows raises the question whether they have a common origin or whether one arrow is more fundamental than the others. This is the question of the master arrow of time. We may attempt an answer by way of an analogy. Darwin faced a similar conundrum when he observed surprising degrees of similarity between groups of finches on the Galapagos Islands (Fig. 26.1a).

The birds belonged to the same species but their beaks differed markedly between the different islands. As is well known today Darwin's solution to the problem of diversity was the theory of natural selection. The different morphology of the finches was due to adaptive pressures from the different environmental conditions on each island and the need for survival. The finches on the various islands had adapted to the different environmental conditions, which offered them a variety of food supplies (fruit, leaves, and insects). The differences in beak shape were a response to the differences in the environmental conditions. The principle of natural selection was the mechanism, which explained the differences between the subspecies of finches and generally the differential adaptation of species to their different environments. But if the environment changes, subject to dynamical processes, it is to be expected that organisms, belonging to the same species will eventually diverge if they become separated by some geographic barrier to the point where they will no longer be able to mate. In this way, Darwin argued, new species may evolve by what he called 'descent with modification'. Descent with modification (evolution) means that whole lineages can split and evolve along a separate trajectory. Such evolutionary processes then explained the astonishing diversity of life on Earth. The diversity of species, according to Darwinism, is due to two factors: isotropic variation or the random genetic mutations, which take place on the genetic level *and* natural selection, according to which only those individuals of a species tend to survive, which are fit for a particular environment. Descent with modification also implies that species have a common ancestor, as is the case for instance in the case of hominids, including *homo sapiens*. The similarity between species, say between humans and chimpanzees, can then be traced back to their common ancestor (Fig. 26.1b). Contrary to a widespread opinion, Darwin did not teach that 'man descended from the ape'. The correct Darwinian view is that humans and other hominids have a common ancestor, which explains the similarities between them; the dissimilarities are due to the different environmental pressure which they face.[1]

Returning to the question of the many arrows of time and their parallelism it is tempting to adopt an evolutionary view. The Darwinian analogy suggests that we

---

[1] Descent with modification can also explain the similarity between different languages: the family of Indo-Germanic languages displays similarities because they can be traced back to a common proto-language. On Darwinism see Weinert (2009): Part II.

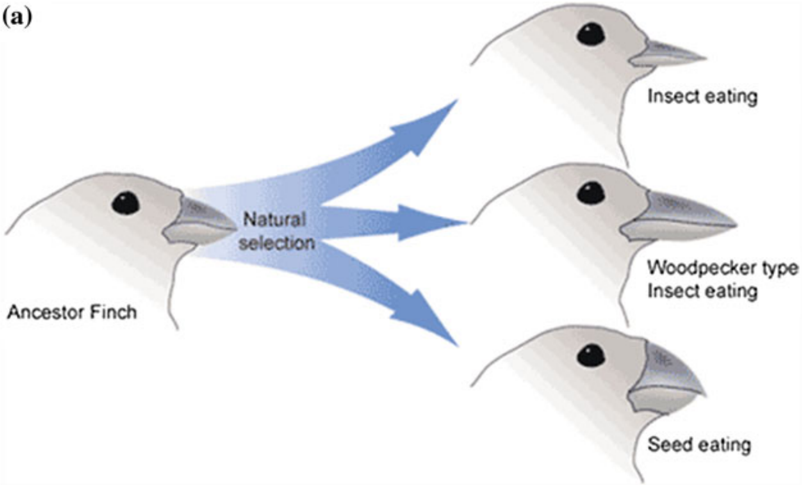


Fig. 26.1 **a** The adaptive features of Darwinian finches. **b** A Darwinian tree of life. *Source*: Wikimedia Commons

should separate the question of the parallelism of the many arrows of time from the question of the master arrow of time. The separation of these questions has echoes in the history of evolutionary thinking. Darwin was *not* the discoverer of evolution. Many thinkers had noticed and described the existence of different species and the splitting of lineages. By the time Darwin published *The Origin of Species* (1859) the evolution of species was a well-established fact. But the cause of this evolution was disputed. Darwin's genius consisted in his discovery of a material process—natural selection—which could explain the appearance of the evolutionary tree.

Similarly, the existence of the many temporal arrows and their uni-directional nature is undoubted. As observed above most of these arrows seem to point in the direction of greater order and information. For instance, establishing the cause of an effect increases our knowledge and information about the causal process involved in bringing about the effect. As we grow older the psychological arrow builds a store of memories but the future remains uncertain. Furthermore, not all these arrows originate at the same time. In this respect they resemble the emergence of species. According to the evolutionary tree, new species evolve—as the sprouting of new branches on the evolutionary tree—far above the ground level. Similarly, causal, cosmic and thermodynamic arrows are far older than, say, the psychological arrow, since humans are a late product of the evolutionary process. *Homo sapiens* had an ancestor in earlier hominoid groups but modern humans are no older than approximately 30,000 years. With the emergence of *homo sapiens* came the emergence of phenomenal time; civilization created human time. But physical time existed well before the appearance of *homo sapiens*. Hence arrows of time can emerge at different stages in the evolution of the universe.

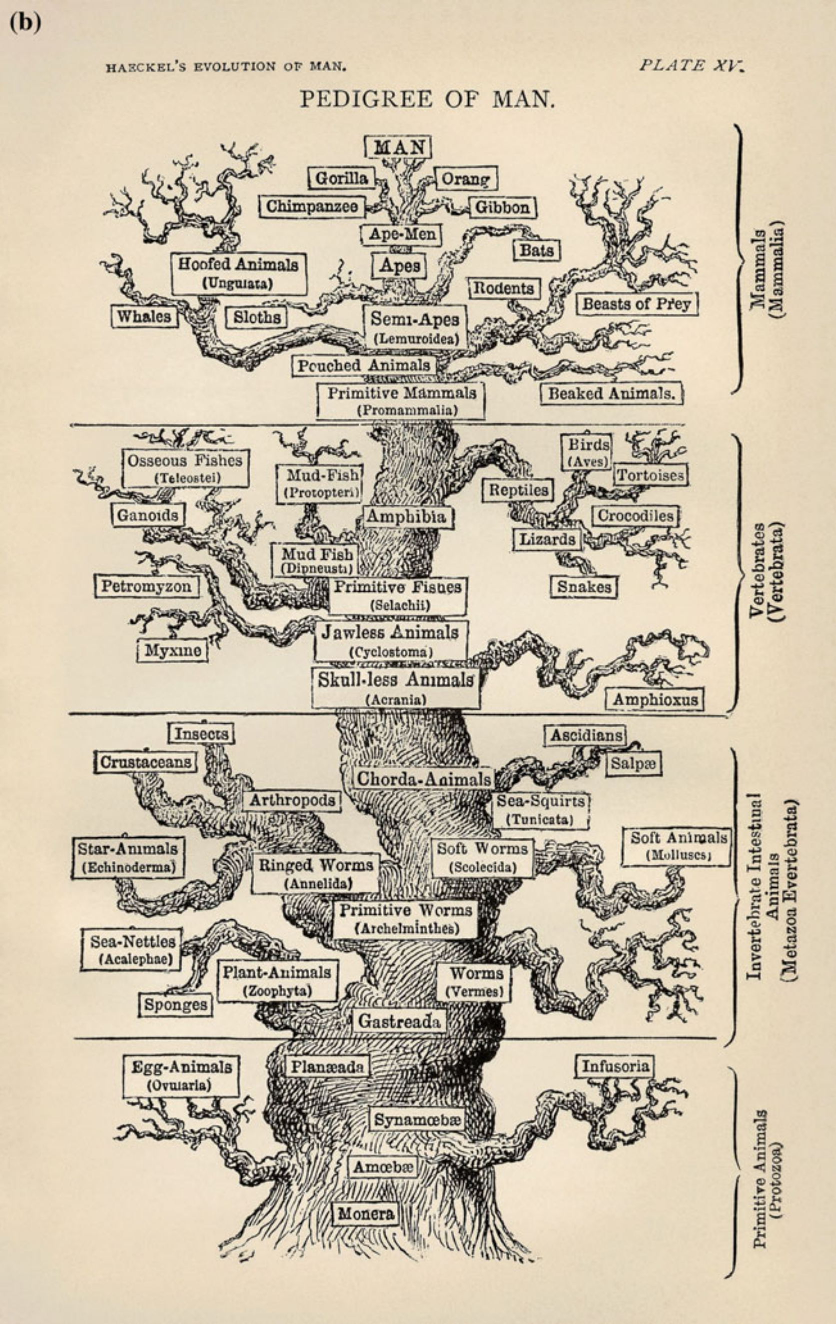


Fig. 26.1 (continued)

The differential emergence of arrows of time strongly suggests that there is no master arrow of time, just as there is no master species. According to Darwin the evolutionary process is contingent and hence dependent on the given conditions (random mutations and environmental pressures). Even if the Big Bang is not a singularity, as in the multiverse scenarios, it lies at the root of the (local) arrows of time, just as the emergence of life on Earth lies at the root of all the diversity of the species. But the original spark of life is not the cause of its diversity.

Similarly, the Big Bang, if cosmologists are to be believed, lies at the origin of a low-entropy beginning of the universe and the resultant thermodynamic arrow of time. Some arrows of time keep a tight fit with the thermodynamic arrow of time—for instance the radiative arrow and the cosmic arrows of time. There may be good reasons to think—as some have argued (Zeh 1992)—that the radiative arrow can be reduced to the thermodynamic arrow. But in the case of the expansion of the universe, there exists only a coincidental link between the cosmic arrow—as based on this accelerated expansion—and the thermodynamic arrow. Certain cosmological models allow the expansion of the universe to come to a halt and switch over to a contraction to form closed time-like curves. But even in a contracting universe, the entropy would continue to increase; for what contracts is the occupied phase space itself but not the individual processes, which take place within it.[2] The causal arrow, as we have seen, is not directly linked to the thermodynamic arrow, for causal processes may lead to both increase and decrease in entropy, at least in the local causal field.

Finally, there is a close link between the passage of phenomenal time and the arrow of physical time. As has been argued phenomenal time is dependent on physical time but the anisotropy of time, as experienced by humans, is not directly deduced from entropic processes. Arthur Eddington held that humans had a direct awareness of phenomenal time and its flow. Certainly Nature provides many processes to enable human observers to experience the lapse of time. They may, in some cases, be due to thermodynamic processes (like the cooling of liquids, the aging process, radiation and emission events). But such processes are experienced as *sui generis*, as indicators of the anisotropy of time in their own right. Humans can develop a sense of the passage of time without knowledge of the Second law of thermodynamics.

Rather than speaking of a master arrow of time, the notion of *multi-fingered time* is more appropriate. Of course some arrows of time will be regarded as more important than others. Yet they are not master arrows of time because such a thesis would imply that other arrows of time can be reduced to them. But this does not fit the evolutionary analogy. In this respect, the evolutionary image adopted here deviates from the Evolving Block Universe. The latter sees in the thermodynamic arrow a master arrow of time, as Eddington did at some point.

---

[2] Consider an analogy: if the universe were a very large billiard table, the borders of which slowly began to retract, the balls would retain their normal trajectories and the players would notice nothing amiss.

> The other arrows derive from the global master arrow of time resulting from the universe's early expansion from an initial singularity in an Evolving Block Universe. The arrow of time at the start is the time direction pointing away from the initial singularity towards the growing boundary of spacetime; this then remains the direction of time at all later times. (Ellis 2013: 248)

According to the Evolving Block Universe model, as developed by cosmologist G. Ellis, the Big Bang lays the foundation to all the other arrows so that the thermodynamic arrow becomes the master arrow of time. But just like the evolution of life, the evolution of the universe is a contingent process, producing not one but several arrows of time. Like species they may be more or less related.

The notion of multi-fingered time is usually associated with the theory of relativity (cf. Zeh 1992: Chap. 5). As we have seen, according to the theory of relativity, clock-time is frame-dependent. As reference frames move at relative velocities with respect to each other their clocks indicate different clock times, as seen from a frame considered at rest. From the point of view of a rest frame, the clock of a moving frame seems to slow down. But it is worth recalling that entropic processes, which could form the basis of a clock, are frame-invariant.

Most of the arrows we have discussed are not affected by relativistic velocities. Most systems, in which arrows of time arise, move at speeds far below any relativistic effects. But even though the physical processes move at non-relativistic velocities, they provide the criteria or signposts for inferring arrows of time. These physical processes are objective and occur irrespective of human awareness. But it depends on human inferential capacities to take these physical criteria as a basis for the inference of arrows of time. Although it is reasonable to assume that inferential capacities are equally universal and objective, nevertheless the capacities are influenced by human knowledge and culture.

Thus it becomes possible that different models of time are claimed to be the conceptual consequences of agreed physical criteria. As observed above, so-called archaic societies possess little scientific knowledge and make a distinction between profane and mythical time. Because of the cyclic nature of the experience of archaic societies—the seasons—it is almost natural for the archaic mind to develop a cyclic notion of time, even though it does not stand up to philosophical scrutiny. Scientific knowledge at the beginning of the Christian age became sophisticated, due to the magnificent efforts of Greek thinkers. The ancient Greeks developed a linear notion of time, at least as far as the sublunary sphere was concerned. This was the sphere of decay and change, whilst the supralunary sphere was characterized by symmetry and permanence, hence it did not give rise to a cosmic arrow of time. The cosmic arrow of time was a realization of the 19th century and intimately related to the discovery of entropy and the Second law of thermodynamics. But in the 19th century, thinkers influenced by Newton and Leibniz agreed that time was universal: clock time was assumed to be the same throughout the universe. As Newton said in his essay 'De Gravitatione' (1680):

> The moment of duration is the same at Rome and London, on the Earth and on the stars, and throughout the heavens.

The dispute between the Leibnizians and Newtonians, between relationism and substantivalism, did not concern the question of the universality of time but that of the materiality of time. Newton argued, in connection with his theory of mechanics, that time was 'absolute', by which he meant that it did not depend on any material processes in the universe. Hence the universe 'has a clock'. Leibniz, by contrast, replied that time was relational, by which he meant that it depended on the succession of material events in the universe. Hence the universe 'was a clock'.

Einstein's Special theory of relativity (1905) then went further and questioned the notion of the universality of time. Time became relative—more precisely, the ticking of clocks became relative to inertial reference frames. In the words of W. Pauli, 'there are as many clocks as there are reference frames', and as these clocks appear to tick at different rates as seen across different reference frames, it seems that there is no universal 'Now'; nor is there agreement on the simultaneity of events for observers who are attached to reference frames which are in relative, relativistic notion with respect to each other. In this sense, then, clock time became relative to different reference frames. According to the theory of relativity, observers in relative motion with respect to each other, do not agree on their respective clock times. Time became multi-fingered. But on the evolutionary view, presented here, time is also multi-fingered because many different arrows of time emerged during cosmic evolution, both in individual universes and, according to some models, in the multiverse itself. There are many physical criteria, which could serve as indicators of the arrows of time. They may serve as independent criteria. Hubble's discovery of the recession of the galaxies from each other, from which the expansion of the universe followed, could serve as a criterion for the cosmic arrow of time. Although the expansion of the universe is related to the Second law of thermodynamics, the expansion itself can be understood without any reference to thermodynamic processes. Equally, many thermodynamic processes in our local neighbourhood can serve as criteria for the passage of time, without reference to cosmic expansion.

The notion of multi-fingered time—that is the multiple criteria provided by the universe to infer arrows of time—also suggests that a concentration on one criterion—like the relativity of simultaneity in the Special theory of relativity—can be too restrictive. The Special theory inspired many physicists, and philosophers alike, to infer that time does not exist in the physical universe and must therefore be a human illusion. But to infer the non-existence of physical time from the notion of relative simultaneity is misleading. First, no distinction is made between physical, phenomenal and human time. The reference systems of the Special theory are equipped with clocks—mechanical devices—which may or may not be read by human observers. These clocks do indeed undergo time dilation as seen from a reference system considered to be at rest. But thermodynamic processes are frame-independent—entropy increases in all systems moving with respect to each other even at relativistic velocities. Second, human observers moving at non-relativistic velocities have a similar experience of the march of time, even though there are variations according to age and psychological state. Third, the expansion of the universe provides a cosmic arrow—defined as its average mass density—and would be experienced the same for all observers, if they could share the co-moving patch, from which the demon observes the evolution of the universe. Fourth, the emission of light signals and other thermodynamic and electro-magnetic processes in Minkowski space-time clearly define an 'earlier-later' distinction, on which all time-like related observers can agree. To infer the 'subjectivity' of time from the notion of relative simultaneity in the Special theory is to neglect many other criteria, which clearly suggest an inference to the anisotropy of physical time. Finally, the relativity of time, as implied by the theory of relativity, also stands in contrast with the ordinary experience of the 'flow' of time of macroscopic observers at familiar (non-relativistic) velocities. Human beings, as information using and gathering systems, infer the notion of time and its properties from the information, which the material world around them furnishes. Although there is no master arrow of time, there are many signposts, which allow humans to infer the anisotropy of time. To multi-fingered time therefore corresponds a multiplicity of physical criteria, which strongly suggests that time is not a human illusion.

Is time a human construct, then? Time is a human construct if this means that the notion of time and its property—anisotropy—need to be inferred from the observation of physical criteria. In this sense, humans become aware of physical time, which clearly exists prior to human consciousness. Awareness of phenomenal time is dependent on the workings of the human mind. Through the efforts of civilization, humans construct human time. But a time sense exists demonstrably in plants and animals, in which we do not presuppose conceptual awareness, and hence no fully developed notion of time. Biological time can be understood as the time within the coordinate system of a living organism (Fisher 1966). A biological study of time involves the study of temporal rhythms in animals and plants as well as human beings. There is no doubt that biological organisms possess biological clocks, for which there exists evidence in three different areas (Hamner 1966; Whitrow 1980: Chap. III; Wright 2006):

- The discovery of photo-periodism by Garner and Allard (1920) revealed the influence of the relative length of day and night on the flowering response of many plants.
- The celestial orientation in birds and insects shows their time sense. Birds use the 'apparent' position of the sun in the sky and compensate for the sun's movement during the day. Migrating birds also find their direction by following patterns of stars in the night sky, thus reminding us of the connection between time and cosmology.
- There are circadian rhythms (approximately 24-h rhythms) in plants and nocturnal animals, which are taken as a direct manifestation of biological clocks.

There are also some interesting biochemical considerations of the time sense in human beings (Hoagland 1966). We judge psycho-physiological time with our brain. Our judgement of time depends on the speed of chemical processes in the brain. If the body temperature is raised chemical reactions increase which leads to more psychological time passing in a given interval of clock time. Two units of ‘private’ time may pass in one minute of clock time with the consequence that time appears to go slowly. If the body temperature is lowered the rate of biochemical changes is reduced; this leads to less physiological time passing in a given interval of clock time: clock time appears to go fast. Such biochemical considerations may hold the key to an explanation of the impression of faster moving time in old age, as the cerebral oxygen consumption slows down with advancing years. Experiments have shown that certain types of tranquilizers decrease the metabolic rate and lead to the perception that time ‘passes like magic’. By contrast certain stimulants like LSD result in an increased metabolic rate and lead to the perception of time passing slowly (see Fisher 1966: 364). Although Saint Augustine knew nothing of the biochemistry of the brain, he nevertheless became the discoverer of psychological time, since he made the mind the metric of time.

That time is a human construct also has a philosophical sense: it means that only humans develop sophisticated models of time on the basis of inferences from chosen criteria. The criteria are objectively given, like say, the expansion of the universe or the increase in entropy in thermodynamic systems. The choice of criteria obviously influences the inferences, which are drawn and the conceptual consequences, which are accepted.

Chapter 27
What Landsberg's Demon Tells Us and Does not Tell Us About the Arrows of Time

> *The facts of physics do not oblige us to accept one philosophy rather than the other.*
>
> (Bell 1987: 77)

If our models of time are conceptual consequences from given criteria—often provided by science—these models of time do not follow with deductive certainty from scientific theories. It does not follow from the theory of relativity that time is a human illusion or that arrows of time do not exist. Landsberg's Demon tells us that Nietzsche's Demon is mistaken. To claim that Minkowski's space-time represents a Block Universe is to restrict the criteria, from which inferences are drawn, to just one: relative simultaneity. But there exists a multiplicity of criteria as the history of time reckoning from Plato to Leibniz, Newton and Eddington testifies. When we face such conflicting inferences, it is advisable to consider all available criteria to make reasonable inferences about the notion of time. This looks like an easy task for the demons of science but should not be beyond the abilities of their all too fallible human counterparts. In fact, even demons may make philosophical mistakes. The art consists in identifying the most reliable criteria.

Science tells us that there are numerous objective linear or periodic processes, on which to base inferences about the notion of time. These processes are regular and invariant enough to make the inferences reliable. Advancement of knowledge may sometimes show that a criterion fails to make the grade. A clear case of a change of criteria with respect to the notion of time is the motion of the Earth. The Greeks believed that the Earth was stationary at the centre of the universe; the planets (including the sun) and the stars performed their daily and annual rotations around the centre. As the celestial objects were supposed to pursue a circular orbit, the criteria of invariance and regularity were satisfied. Thus Plato could claim that the 'planets were instruments of time'. Copernicus upset the geocentric worldview by showing that the sun was actually the 'centre' of the solar system. But Copernicus still believed that the orbits of the planets around the sun were circular. The German astronomer Johannes Kepler finally abandoned the age-old adherence to circular orbits and showed, mathematically, that planets move in elliptical orbits around the sun. These orbits could still have formed reliable criteria for the passage of time.

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_27

But Newton suspected, quite rightly as it turned out, that the planetary orbits suffer from irregularities, which make them poor candidates as reliable criteria. For instance, there are variations in the Earth's orbit around the sun due to seasonal changes; the orbital velocity of the Earth is also slowing down due to the tidal movement of its water masses; finally the geometric and the physical axes do not quite coincide, which also leads to irregularities. Due to these irregularities, Newton chose to adopt an unchanging notion of time. He postulated his conception of absolute time, which was not meant to be dependent on any physical processes. Leibniz rejected the view that time was not based on material processes in the universe. He formulated his relational theory of time, according to which time is the order of succession of events in the universe. The motion of the Earth around the sun was eventually abandoned completely and replaced by more regular (and invariant) criteria, like atomic oscillations or the pulses of neutron stars. Both provide a degree of precision, which is unattainable by the planetary orbits. Modern clocks are based on such highly precise and regular physical processes. In fact, the precision is achieved through a triangulation of such physical processes. Modern clock time is triangulated time, which is constructed from the coordination of several precise physical systems in the material world around us.

Although science furnishes us with increasingly precise regular and invariant physical criteria, surprisingly it does not provide us with a fully developed notion of time. Our investigation therefore departs from Reichenbach's claim:

> If there is a solution to the philosophical problem of time, it is written down in the equations of mathematical physics. (Reichenbach 1956: 17)

Science gives us clock time—according to Galileo's fall experiments and Einstein's theory of relativity, time is what the clocks tell us. As clocks behave differently in gravitational fields and in fast-moving reference frames, it seems to follow that there is no universal time, no universal Now. But clock time does not exhaust the nature of time: its mysterious flow, its passage, its arrows. Cosmology provides us with a cosmic arrow of time, courtesy of Landsberg's Demon. But this Demon does not reveal the nature of time. Hence a fully developed notion of time—a philosophical model of time—is indeed a construct from a cluster of criteria. They form the basis from which the models are inferred. But drawing these philosophical consequences no longer lies within the purview of science proper. Science, then, does not fully answer philosophical questions about the nature of time. There are many notions of time: physical, philosophical, psychological, and sociological. Science and philosophy provide clues as to how they may be related. But the fact that science gives us criteria to construct some of the properties of time, without providing a full picture of its nature, is not an open invitation to freely speculate about the notion of time. In his *Critique of Pure Reason* (1781/1787), Kant argues that even metaphysics had to be confined to well-defined constraints. In a similar way our models of time are not free inventions of the human mind. As far back as Greek geocentrism, there has always existed a close link between time and cosmology. Already the Greeks constructed their models of time on the basis of their scientific knowledge. There exists, then, a close link between the physical criteria, provided by science, and conceptions of time. Over the centuries, as the criteria have changed, different models of time have been developed (absolute, idealist, relational). Landsberg's Demon encourages us to think that science will continue to furnish increasingly precise criteria to serve as signposts for the notions of time and in particular its arrows. After all, Landsberg's Demon is a denizen of the multiverse.

Landsberg's Demon will know what today's cosmologists can only surmise, namely whether the multiverse itself has an arrow of time or whether only individual universes do. If the multiverse is in a state of stasis, only its cosmic offspring—individual universes—will display an arrow of time. If the multiverse is in a state of flux, of dynamic change, it will itself exhibit an arrow of time.

All the demons we have encountered are reluctant to share their knowledge with their human apprentices. As far as the demons are concerned, humans will never graduate from this apprenticeship. But it is good practice to contemplate the power of demons for they challenge our knowledge claims. They remind us, in Newton's metaphor, that we are like children playing on the beach whilst a vast ocean of knowledge remains to be discovered. In today's terms the ocean to be explored concerns questions about the nature of time and its properties. It concerns the long-term evolution of the universe and the multiverse. It is the task of human researchers to find the answers, for the demons, as always, remain silent.

Part V
Conclusion

Chapter 28
Conclusion

> *We may think of reality as a set of concentric spheres, progressively revealed as we detach gradually from the contingencies of the self. Th. Nagel, The View from Nowhere.*
>
> (Oxford UP 1986: 5)

This was a book about the demons of science and how they are used as argument patterns in philosophical and scientific reasoning about the world. The demons, all powerful as they appear to be, advance particular views about the make-up of the surrounding cosmos: that it is deterministic, probabilistic or even cyclic in nature. The demons provide metaphysical constraints on scientific modelling. As the demons become proponents of particular worldviews or (to employ a popular term) paradigms, their claims invariably have philosophical reverberations, for instance with respect to time and the human mind.

The demons, we found, serve as thought experiments and that they have the status of conceptual models. Thought experiments can be inconclusive, misleading or even sources of alternative conclusions. The demons's perorations do not add to the store of empirical knowledge.

Why, then, do scientists and philosophers alike hold them in such high regard? The demons are *agents provocateurs*. They are used to challenge existing knowledge claims, test their coherence and plausibility and propose conjectures of their own. The demons make bold claims, against which our theories can be measured and assessed. Laplace's Demon maintains that the world is deterministic, seemingly depriving us of arrows of time and free will. Maxwell's Demon shows that the Second law of thermodynamics is probabilistic, rather than deterministic, thus seemingly questioning the notion of entropy as a useful measure of the anisotropy of time. Nietzsche's Demon announces that the universe is cyclic, condemning us to an eternal recurrence of events. The demons are like youthful hotheads but a calm consideration of their claims reveals more balanced views with respect to the arrows of time and the human mind.

The demons's provocations are nevertheless useful exercises because—like all good philosophy—they force us to pause and reconsider our philosophical

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3_28

assumptions. Their main job is to add to our *understanding* of the world around us, even if it is in a round-abound way. By engaging with their claims, we get a better sense of the world in which we live and the role of our theories. They inform us, unwittingly, what science can and cannot tell us about the world. The demons are not perfect. They disagree with each other. Their great disadvantage is their presumptuousness. Their great advantage is that they are not hampered by human limitations. If all knowledge is conjectural, as Popper held, then the demons embody another form of testability.

Bibliography

Albert, D. (2000). *Time and chance*. Cambridge, Massachusetts, London: Harvard UP.

Alvarez, W. (1997). *T-rex and the crater of doom*. Princeton, New Jersey: Princeton University Press.

Ambegaokar, V., & Clerk, A. A. (1999). Entropy and time. *American Journal of Physics, 67*(12), 1068–1073.

Aristotle (1952a). *Physics. Great books of the western world*. The works of Aristotle (Vol. I). Chicago, London, Toronto: Encyclopaedia Britannica.

Aristotle (1952b). *On the heavens. Great books of the western world*. Chicago, London, Toronto: Encyclopaedia Britannica.

Bailer-Jones, D. (2003). When models represent. *International Studies in the Philosophy of Science,17*, 59–74.

Bailer-Jones, D. (2009). *Scientific models in philosophy of science*. University of Pittsburgh Press.

Bardon, A. (2013). *A brief history of the philosophy of time*. Oxford: Oxford University Press.

Barrow, J. D. (2007). *New theories of everything: The quest for ultimate explanation*. Oxford: Oxford University Press.

Bayne, T. (2011). Libet and the case for free will scepticism. In R. Swinburne (Ed.), *Free Will and Modern Science* (pp. 25–46).

Bell, J. S. (1987). *Speakable and unspeakable in quantum mechanics*. Cambridge: Cambridge University Press.

Bennett, C. H. (1987). Demons, engines and the second law. *Scientific American, 257*(5), 88–96.

Bernard, C. (1865). *Introduction à l'étude de la médicine expérimentale*. Paris: Garnier-Flammarion.

Bishop, M. (1999). Why thought experiments are not arguments. *Philosophy of Science, 66*, 534–541.

Blackmore, S. (2003). *Consciousness: An introduction*. London: Hodder Education.

Bohm, D. (1957). *Causality and chance in modern physics*. London: Routledge & Kegan Paul.

Bohr, N. (1929). Atomic theory and the fundamental principles underlying the description of nature. In N. Bohr (Ed.), *Atomic theory and the description of nature* (pp. 103–119). Cambridge UP 1934.

Bokulich, A. (2001). Rethinking thought experiments. *Perspectives on Science, 9*, 285–307.

Boltzmann, L. (1886). The second law of thermodynamics. In L. Boltzmann, *Selected Writings* (pp. 13–32) (1974).

Boltzmann, L. (1897). On Zermelo's paper "On the mechanical explanation of irreversible processes". In L. Boltzmann, *Irreversible Processes* (pp. 238–45) (1966).

Boltzmann, L. (1898/^{3}1923). *Vorlesungen über Gastheorie*, II. Leipzig: Johann Ambrosius Barth.

Boltzmann, L. (1966). Kinetic theory. In: S. G. Brush (Ed.), *Irreversible Processes* (Vol. 2). Oxford/London: Pergamon Press.

Boltzmann, L. (1974). Theoretical physics and philosophical problems. In B. McGuinness (Ed.), *Selected writings*. Dordrecht/Holland: D. Reidel Publishing Company.

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3

Boltzmann, L. (1979). *Populäre Schriften*. Braunschweig: Vieweg.

Born, M. (1955). Is classical mechanics in fact deterministic? *Physikalische Blätter 11*(9), 49–54 reprinted in *Physics in My Generation*. London/New York: Pergamon Press (1956), 164–170.

Brandenberger, R. (2013). Do we have a theory of early universe cosmology? *Studies in History and Philosophy of Modern Physics*, *46*, 109–121.

Brown, J. R. (1991). *The laboratory of the mind*. London: Routledge.

Brown, J. R. (2004). Peeking into Plato's heaven. *Philosophy of Science, 71*, 1126–1138.

Brown, J. R. (2014). Thought experiments. *The Stanford Encyclopaedia of Philosophy*.

Brush, S. G. (1976). Irreversibility and indeterminism: Fourier to Heisenberg. *Journal of the History of Ideas, 37*, 603–630.

Bunge, M. (1959). *Causality and modern science*. Harvard UP.

Bunge, M. (1977). Emergence and the mind. *Neuroscience, 2*, 501–509.

Bunzl, M. (1996). The logic of thought experiments. *Synthese, 106*, 227–240.

Butterfield, J. (1984). Seeing the present. *Mind, 93*, 161–176.

Callender, C. (2008). The common now. *Philosophical Issues, 18*(1), 339–361.

Ĉapek, M. (1951). The doctrine of necessity re-examined. *Review of Metaphysics*, *V*, 11–54.

Čápek, V., & Sheehan, D. P. (2005). *Challenges to the second law of thermodynamics*. Dordrecht: Springer.

Carroll, S. (2008). The cosmic origins of time's arrow. *Scientific American*, 48–57 (June).

Carroll, S. (2010). *From eternity to here*. Oxford: OneWorld.

Cartwright, N. (1999). Models and the limits of theory: Quantum hamiltonians and the BCS models of superconductivity. In M. S. Morgan & M. Morrison (Eds.), *Models as Mediators* (pp. 241–281).

Cassirer, E. (1956/1936). *Determinism and indeterminism in modern physics*. Yale University Press.

Caticha, A. (2014). Towards an informational pragmatic realism. *In Minds and Machines, 24*(1), 37–70.

Chalmers, D. (1998). *The conscious mind*. Oxford: Oxford University Press.

Clark, A. (1997). *Being there*. Cambridge, Massachusetts, London: MIT.

Clatterbuck, H. (2013). The epistemology of thought experiments: A non-eliminative, non-platonic account. *European Journal for Philosophy of Science, 3*(3), 309–329.

Clemence, G. M. (1966). Time measurement for scientific use. In J. T. Fraser (Ed.), *The voices of time* (pp. 402–12).

Cooper, R. (2005). Thought experiments. *Metaphilosophy, 36*, 328–347.

Copernicus, N. (1543/1995). *On the revolutions of heavenly spheres*. Amherst, New York: Prometheus Books.

Costa de Beauregard, O. (1966). Time in relativity theory: Arguments for a philosophy of being. In J. T. Fraser (Ed.), *The voices of time* (pp. 417–433).

Dainton, B. (2010). Temporal consciousness. *Stanford Encyclopaedia of Philosophy*.

Dainton, B. (2011). Time, passage and immediate experience. In C. Callender (Ed.), *Oxford handbook of the philosophy of time* (pp. 382–419).

Davies, P. (1974). *The physics of time asymmetry*. Leighton Buzzard/London: Surrey University Press.

Darwin, C. (1871). Introduction by James Moore and Adrian Desmond. In *The descent of man*. Penguin Books (2004).

Daub, E. E. (1970). Maxwell's demon. *Studies in History and Philosophy of Science Part A,1*(3), 213–227.

D'Abramo, G. (2012). The peculiar status of the second law of thermodynamics and the quest for its violation. *Studies in History and Philosophy of Modern Physics, 43*, 226–235.

Denbigh, K. G. (1981). *Three concepts of time*. Berlin, Heidelberg, New York: Springer.

Denbigh, K. G., & Denbigh, J. S. (1985). *Entropy in relation to incomplete knowledge*. Cambridge: Cambridge University Press.

Dennett, D. (1992). *Consciousness explained*. Penguin Science.

D'Holbach, P. T. (1770). *Système de la nature*. (English Translation: *The system of nature*, London 1817; German Translation: System der Natur, Frankfurt a./M.: Suhrkamp 1978).

De Broglie, L. (1941). *Continu et discontinu en physique moderne*. Paris: Albin Michel.

Deutsch, D. (1998). *The fabric of reality*. Penguin Press.

Dieks, D. (1988). Special relativity and the flow of time. *Philosophy of Science*, *55*, 456–460

Dorato, M., & Wittmann, M. (2015). The now and the passage of time. *KronoScope, 15*, 191–213.

Earman, J. (1967). Irreversibility and temporal asymmetry. *The Journal of Philosophy LXIV*(*18*), 543–549.

Earman, J. (1974). An attempt to add a little direction to "the problem of the direction of time". *Philosophy of Science, 41*(1), 15–47.

Earman, J. (1986). *A primer on determinism*. Dordrecht, Holland: D. Reidel.

Earman, J. (2002). What time reversal invariance is and why it matters. *International Studies in the Philosophy of Science, 16*(3), 245–264.

Earman, J. (2004). Determinism: What we have learned and what we still don't know. In J. K. Campbell, M. O'Rourke & D. Shier (Eds.), *Freedom and determinism* (*Topics in contemporary philosophy series*) (pp. 21–46). Bradford Book.

Earman, J., & Norton, J. D. (1998). Exorcist XIV: The wrath of Maxwell's demon. Part I. From Maxwell to Szilard. *Studies in History and Philosophy of Modern Physics, 29*, 435–471.

Earman, J., & Norton, J. D. (1999). Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and Beyond. *Studies in History and Philosophy of Modern Physics, 30*, 1–40.

Eckhardt, W. (2006). Causal time asymmetry. *Studies in History and Philosophy of Modern Physics, 37*(3), 439–466.

Eddington, A. (1932/1928). *The nature of the physical world*. Cambridge: Cambridge University Press.

Eddington, A. S. (1935). *New pathways in science*. Cambridge: Cambridge University Press.

Eddington, A. S. (1936). *Relativity theory of protons and electrons*. Cambridge: Cambridge University Press.

Edelman, G. (1992). *Bright air*. Brilliant Fire: Penguin.

Edelman, G. (2004). *Wider than the sky*. Penguin.

Einstein, A. (1920/^{15}1960). *Relativity*: *The special and the general theory*. London: Methuen.

Einstein, A. (1940). Considerations concerning fundaments of theoretical physics. *Nature,145*, 920–924.

Einstein, A. (1949). Autobiographical notes. In A. Schilpp (Ed.) (Vol. II, pp. 2–94).

Einstein, A. (1950). On the generalized theory of gravitation. *Scientific American, 182*, 341–356.

Eliade, M. (1954). *The myth of eternal return*. London: Routledge & Kegan Paul.

Elias, N. (1988). Über die Zeit. Frankfurt a./M/: Suhrkamp (English Translation: *Time–An Essay*, Blackwell 1992).

Ellis, G. F. R. (2013). The arrow of time and the nature of spacetime. *Studies in History and Philosophy of Modern Physics, 44*, 242–262.

Falkenburg, B. (2012). *Mythos Determinismus*. Heidelberg/Berlin/New York: Springer.

Feyman, R. (1963). *The Feynman lectures on physics* (Vol. I). Addison-Wesley Publishing Company.

Fisher, R. (1966). Biological Time. In J. T. Fraser (Ed.), *1966*, 357–381.

Fogel, R. W. (1964). *Railroads and American economic growth*. Baltimore: John Hopkins Press.

Frank, Ph. (1932). *Das Kausalgesetz und seine Grenzen*. Wien: Julius Springer.

Fraser, J. T. (Ed.). (1966). *The voices of time*. George Braziller/Allen Lane The Penguin Press (1968).

Freud, S. (1916). Introductory lectures on psycho-analysis, Part IIII. In S. Freud, J. Strachley (Eds.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. XVI). London: The Hogarth Press (1963).

Frigg, R. (2008). A field guide to recent work on the foundations of statistical mechanics. In D. Rickles (Ed.), *The Ashgate companion to contemporary philosophy of physics* (pp. 99–196). London: Ashgate.

Frigg, R. (2010). Fiction and scientific representation. In R. Frigg & M. Hunter (Eds.), *Beyond mimesis and nominalism: Representation in art and science* (pp. 97–138). Berlin, New York: Springer.

Frigg, R., Bradley, S., Du, H., & Smith, Leonard A. (2014). Laplace's demon and the adventures of his apprentices. *Philosophy of Science, 81*(1), 31–59.

Galileo, G. (1954/1638). *Dialogue concerning two new sciences.* New York: Dover.

Gell-Mann, M., & Hartle, J. B. (1990). Quantum mechanics in the light of quantum cosmology. In W. Zurek (Ed.), *Complexity, entropy and the physics of information* (pp. 425–465). Addison-Wesley.

Gell-Mann, M., & Hartle, J. B. (1993). Classical equations for quantum systems. *Physical Review D, 47*(8), 3345–3382.

Gendler, T. (1998). Galileo and the Indispensability of Scientific Thought Experiment. *British Journal for the Philosophy of Science, 49*, 397–424.

Gendler, T. (2004). Thought experiments rethought. *Philosophy of Science, 71*, 1152–1163.

Genz, H. (2005). *Gedankenexperimente*. Rowohlt.

Giere, R. (2004). How models are used to represent reality. *Philosophy of Science, 71*(5), 742–752.

Gleick, J. (1998): *Chaos*. Vintage Books.

Gold, T. (1962). The arrow of time. *American Journal of Physics, 30*, 403–410.

Gold, T. (1966). Cosmic processes and the nature of time. In R. G. Colodny (Ed.), *Mind and cosmos* (pp. 311–329). University of Pittsburgh Press.

Gold, T. (1974). The world map and the apparent flow of time. In B. Gal-Or (Ed.), *Modern developments in thermodynamics* (pp. 63–72). New York: Wiley.

Grant, E. (1981): *Much Ado about Nothing*. Cambridge University Press.

Griffiths, R. (1994). Statistical irreversibility: Classical and quantum. In J. J. Halliwell et al. (Eds.) (pp. 147–159).

Haggard, P. (2011). Does brain science change our view of free will? In R. Swinburne (Ed.) (pp. 7–24).

Halliwell, J. J., Pérez-Mercader, J., & Zurek, W. H. (Eds.). (1994). *Physical origin of time asymmetry*. Cambridge: Cambridge University Press.

Hameroff, St, & Penrose, R. (2014). Consciousness in the universe. A review of the 'Orch OR' theory. *Physics of Life Reviews,11*, 39–78.

Hamner, K. C. (1966). Experimental evidence for the biological clock. In J. T. Fraser (Ed.) (pp. 282–295).

Hannam, J. (2009). *God's philosophers*. Icon Books.

Hartle, J. (2005). The physics of now. *American Journal of Physics, 73*(2), 101.

Hartmann, S. (1999). Models and stories in hadron physics. In M. S. Morgan & M. Morrison (Eds.) (pp. 326–346).

Harris, S. (2012). *Free will*. New York: Free Press.

Hawking, S. G. (1988). *A brief history of time*. London, New York: Bantam Press.

Hawking, S. W. (1994). The no boundary condition and the arrow of time. In J. J. Halliwell, J. Pérez-Mercader & W. H. Zurek (Eds.) (pp. 346–357).

Hawking, S., & Mlodinow, L. (2010). *The grand design*. New York: Bantam Books.

Heisenberg, W. (1928). Erkenntnistheoretische Probleme der modernen Physik. In W. Heisenberg (1984–1985), *Band I*: *Physik und Erkenntnis* (1927–1955) (pp. 23–28).

Heisenberg, W. (1931). Die Rolle der Unbestimmtheitsrelationen in der modernen Physik. In W. Heisenberg (1984–1985), *Band I*: *Physik und Erkenntnis* (1927–1955) (pp. 40–47).

Heisenberg, W. (1967). Philosophische Probleme in der Theorie der Elementarteilchen. In W. Heisenberg (1984–1985), *Band II*: *Physik und Erkenntnis* (1956–1968) (pp. 249–255).

Heisenberg, W. (1969). The concept of "understanding" in theoretical physics. In W. Heisenberg (1984–1985), *Band III*: *Physik und Erkenntnis* (1969–1976) (pp. 335–338).

Heisenberg, W. (1973/^{3}1976). *Der Teil und das Ganze*. München: Pieper.

Heisenberg, W. (1984–1985). *Gesammelte Werke/Collected works*. In W. Blum, H. P. Dürr & H. Rechenberg (Eds.), *Abteilung C, Band I-III*. München, Zürich: Pieper.

Hemmo, M., & Shenker, O. (2010). Maxwell's demon. *The Journal of Philosophy, 107*, 389–411.

Hemmo, M., & Shenker, O. (2012). *The road to Maxwell's demon*. Cambridge: Cambridge University Press.

Hempel, C. G. (1952). Typological methods in the natural and the social sciences. In *Aspects of Scientific Explanation* (pp. 155–171). New York: Free Press/London: Collier-Macmillan (1965).

Hoagland, H. (1966). Some biological considerations of time. In J. T. Fraser (Ed.) (pp. 313–329).

Hoefer, C. (2010). Causal determinism. *Stanford Encyclopaedia of Philosophy.*

Horwich, P. (1987). *Asymmetries in time*. Boston: MIT.

Huggett, N. (2010). *Everywhere and everywhen*. Oxford: Oxford University Press.

Humphreys, P. (1993). Seven theses on thought experiments. In J. Earman, et al. (Eds.), *Philosophical problems of the internal and external world* (pp. 205–227). Pittsburgh: University of Pittsburgh Press.

Hutchinson, K. (1993). Is classcial mechanics really time-reversible and deterministic?. *British Journal for the Philosophy of Science, 44*, 307–323.

Hutchison, K. (1995). Temporal asymmetry in classical mechanics. *British Journal for the Philosophy of Science, 46*, 219–234.

Irvine, A. D. (1991). Thought experiments in scientific reasoning. In T. Horowitz & G. Massey (Eds.), *Thought experiments in science and philosophy* (pp. 149–166). Lanham: Rowman & Littlefield.

Jeans, J. (1943). *Physics and philosophy*. Cambridge: Cambridge University Press.

Kant, I. (1781/1787). *Critique of pure reason* (N. Kemp Smith, Trans.). London: The Macmillan Press (1976).

Kepler, J. (1618–1621). *Epitome of Copernican astronomy*. In J. Kepler (pp. 165–245) (1995).

Kepler, J. (1995). *Epitome of Copernican astronomy & harmonies of the world*. Amherst, New York: Prometheus Books.

Kershaw, I. (1998). *Hitler 1889–1936: Hubris*. London, New York: W.W. Norton & Co.

Kim, J. (1993). *Supervenience and mind*. Cambridge: Cambridge University Press.

Kim, J. (2000). *Mind in a physical world*. Cambridge, Massachusetts: MIT Press.

Klapproth, F. (2011). Stable and variable characteristics of the time perspective in humans. *KronoScope, 11*, 41–59.

Koźnjak, B. (2015). Who let the demon out? *Studies in History and Philosophy of Science, 51*, 42–52.

Kragh, H. (2007). *Conceptions of cosmos*. Oxford: OUP.

Krauss, L. M., & Scherrer, R. J. (2008): The end of cosmology. *Scientific American* (March), 47–53.

Kühne, Ulrich. (2005). *Die Methode des Gedankenexperiments*. Frankfurt: Suhrkamp.

Kuhn, T. S. (1964). A function for thought experiments. In I. Hacking (Ed.), *Reprinted in Scientific Revolutions (1981)* (pp. 6–27). Oxford: Oxford University Press.

Kuhn, T. S. (1973). Objectivity, value judgement and theory choice. In T. S. Kuhn (Ed.) (pp. 320–339) (1977).

Kuhn, T. S. (1977). *The essential tension*. Chicago: Chicago University Press.

Kuhn, T. S. (1983). Rationality and theory choice. *The Journal of Philosophy, 80*, 563–570.

Kupervasser, O., Nikolić, H., & Zlatić, V. (2012). The universal arrow of time. *Foundations of Physics,42*, 1165–1185.

Lamarck, J. B. (1809/1963). *Philosophie Zoologique* (H. Elliot, Trans.). New York-London: Hafner.

Landsberg, P. T. (1982a). Introduction. In P. T. Landsberg (Ed.) (pp. 1–30) (1982b).

Landsberg, P. T. (Ed.). (1982b). *The enigma of time*. Bristol: Adam Hilger.

Laplace, P. (1774). La probabilité des causes par les événements. In *Œuvres Complétes de Laplace* (Vol. VIII, pp. 27–62). Paris: Gauthier-Villar et Fils (1841).

Laplace, P. (1820). *Théorie analytique des probabilités*. Paris: Couvier.

Laskar, J. (1989). A numerical experiment on the chaotic behavior of the solar system. *Nature, 338*, 237–238.

Leibniz, G. W. (1715/1971). *The monadology and other philosophical writings* (R. Latta, Trans.). Oxford: Oxford University Press.

Leibniz, G. W. (1715). The metaphysical foundations of mathematics. In P. P. Wiener (Ed.), *Leibniz—Selections* (pp. 201–216). New York: Charles Scribner's Sons.

Leff, H. S., & Rex, A. (Eds.). (1990/^{2}2002). *Maxwell's demon 2: Entropy, information, computing*. Bristol and Philadelphia: Institute of Physics Publishing.

Leff, H. S. (2007). Entropy, its language and interpretation. *Foundations of Physics, 37*(2007), 1744–1766.

LePoidevin, R. (2003). *Travels in four dimensions*. Oxford: Oxford University Press.

LePoidevin, R. (2009). The experience and perception of time. *Stanford Encyclopaedia of Philosophy*.

Levy, O. (Ed.). (1909–1913). *The complete works of Friedrich Nietzsche* (Vol. 18). Edinburgh/London: T. N. Foulis.

Lewis, G. N. (1930). The symmetry of time in physics. In P. T. Landsberg (Ed.) (pp. 37–45) (1982).

Lichtenberg, G. C. (1971). *Sudelbücher*. Darmstadt: Wissenschaftliche Buchgesellschaft.

List, C. (2014). Free will, determinism, and the possibility of doing otherwise. *Noûs, 48*(1), 156–178.

Lockwood, M. (2005). *The labyrinth of time*. Oxford: Oxford University Press.

Loschmidt, J. (1876). Über den Zustand des Wärmegleichgewichtes eines Systems von Körpern mit Rücksicht auf die Schwerkraft. *Sitzb. d. Mathem. Naturwiss*. CL LXXII Bd. II, Abth.

Lucas, J. R. (1973). *A treatise on time and space*. London: Methuen.

Mach, E. (1883/^{9}1976). *Die Mechanik*. Darmstadt: Wissenschaftliche Buchgesellschaft. [English Translation: *The Science of Mechanics*. Open Court Publishing (1915)].

Mackie, J. (1980). *The cement of the universe*. Oxford: Oxford University Press.

Mainzer, K. (1996). *Symmetries in nature*. Berlin: de Gruyter.

Malament, D. (2008). Norton's slippery slope. *Philosophy of Science, 75*(5), 799–816.

Maloney, O. (2009). Information processing and thermodynamic entropy. *Stanford Encyclopaedia of Philosophy*.

Maxwell, J. C. (1871/^{4}1875). *Theory of heat*. London: Longmans, Green & Co.

Maxwell, J. C. (1873). Essay for the Eranus club on science and free will. In L. Campbell & W. Garnett (Eds.), *Life of Maxwell* (pp. 434–444). Chapter XIV (1882).

Maxwell, N. (1985). Are probabilism and special relativity incompatible?. *Philosophy of Science, 52*, 23–43.

Maxwell, N. (1988). Are probabilism and special relativity compatible?. *Philosophy of Science, 55*, 640–645.

Mayr, E. (2001). *What evolution is*. London: Weidenfeld & Nicolson.

McAllister, J. (1996). The evidential significance of thought experiment in science. *Studies in History and Philosophy of Science, 27*, 233–250.

McAllister, J. (2004). Thought experiments and the belief in phenomena. *Philosophy of Science, 71*, 1164–1175.

McCall, S. (1994). *A model of the universe*. Oxford: Clarendon Press.

McKenna, M. (2009). Compatibilism. *Stanford Encyclopaedia of Philosophy*.

McLaughlin, B., & Bennett, K. (2011). Supervenience. *Stanford Encyclopedia of Philosophy*.

Miščevic, N. (1992). Mental models and thought experiments. *International Studies in the Philosophy of Science, 6*, 215–226.

Mittelstaedt, P. (1976). *Philosophical problems of modern physics*. Dordrecht: Reidel.

Montaigne, M. (1962). *Essais* I-III. In A. Thibaudet & M. Rat (Eds.), *Oeuvres completes*. Paris: Pléiade.

Morgan, M.S., & Morrison, M. (Eds.). (1999). *Models as mediators*: *Perspectives on natural and social science.* Cambridge: Cambridge University Press. Ideas in context (52).

Morrison, M. (1999). Models as autonomous agents. In M. S. Morgan & M. Morrison (Eds.) (pp. 38–65).

Mukhanov, V. F. (2015). The quantum universe. *Physik Journal*, 41-46 (Aug/Sept).

Nagel, E. (1961). *The structure of science*. London: Routledge & Kegan Paul.

Nernst, W. (1932). Zum Gültigkeitsbereich der Naturgesetze. *Die Naturwissenschaften, 10*, 489–493.

Newton, I. (1960/1687). *Sir Newton's mathematical principles of natural philosophy and his system of the world.* Florian Cajori (Ed.), Berkeley, California: University of California Press (1960).

Nietzsche, F. (1882/2006). *The gay science.* Mineola, New York: Dover Publications.

Nietzsche, F. (1881/1911). Notes on the eternal recurrence, in Volume 16 of Oscar Levy edition of *Nietzsche's Complete Works* (in English). London: Macmillan.

Norton, J. (1991). Thought experiment in Einstein's work. In T. Horowitz & G. Massey (Eds.), *Thought experiments in science and philosophy* (pp. 129–148). Lanham: Rowman & Littlefield.

Norton, J. (1996). Are thought experiments just what you thought? *Canadian Journal of Philosophy, 26*, 333–366.

Norton, J. (2003). Causation as folk science. *Philosophers' Imprint, 3*(4), 1–22.

Norton, J. (2004). On thought experiments: Is there more to the argument? *Philosophy of science*, *71*, 1139–1151.

Norton, J. (2010). Time Really Passes. *Humana Mente, 13*, 23–34.

Novikov, I. D. (1998). *The river of time*. Cambridge: Cambridge University Press.

O'Connor, T. (2010). Free will. *Stanford Encyclopaedia of Philosophy.*

O'Connor, D. J. (1971). *Free will.* New York: Anchor Books.

Omnès, R. (1999). *Understanding quantum mechanics*. Princeton: Princeton University Press.

Osiander, A. (1543). Preface to *De revolutionibus.* In N. Copernicus (1543/1995) (pp. 1–2).

Pascal, B. (1897). *Pensées*. In L. Brunschvicg (Ed.) Paris.

Penrose, O., & Percival, I. C. (1962). The direction of Time. *Proceedings of the Physical Society*, *79*, 605–16.

Penrose, R. (1989). *The emperor's new mind.* Oxford: Oxford University Press.

Penrose, R. (1995). *Shadows of the mind.* London: Vintage Books.

Penrose, R. (2005). *The road to reality*. London: Vintage Books.

Penrose, R. (2010). *Cycles of time*. London: The Bodley Head.

Peterson, I. (1993). *Newton's clock*. W. H. Freeman and Co.

Petkov, V. (2005). *Relativity and the nature of spacetime*. Berlin, Heidelberg, New York: Springer.

Planck, M. (1948). Die Physik im Kampf um die Weltanschauung. In *Vorträge und Erinnerungen* (pp. 285–300). Darmstadt: Wissenschaftliche Buchgesellschaft (1975).

Poincaré, H. (1952a). *Science and hypothesis*. Mineola, N.Y.: Dover.

Poincare, H. (1952b/1903). *Science and Method*. Mineola, N.Y.: Dover.

Popper, K. R. (1950). Indeterminism in quantum physics and in classical physics. *British Journal for the Philosophy of Science, 1*(3), 173–195.

Popper, K. R. (1956). The arrow of time. *Nature, 177*, 538; *Nature*, *178*, 382.

Popper, K. R. (1957). Irreversible processes in physical theory. *Nature, 179*, 1297.

Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.

Popper, K. R. (1982). *Open universe*. London: Hutchinson.

Price, H. R. (1996). *Time's arrow and Archimedes' point*. Oxford: OUP.

Price, H. (2002). Burbury's last case: The mystery of the entropic arrow. In C. Callender (Ed.), *Time, reality and experience* (pp. 19–56). Cambridge: Cambridge University Press (2002).

Price, H. (2006). Recent work on the arrow of radiation. *Studies in History and Philosophy of Science Part B, 37*(3), 498–527.

Ptolemy, C. (1984). *Ptolemy's Almagest*. In G. J. Toomer (Ed.) London: Duckworth.

Reichenbach, H. (1920). Philosophische Kritik der Wahrscheinlichkeitsrechnung. *Die Naturwissenschaften, 8*, 146–153.

Reichenbach, H. (1931). Das Kausalproblem der Physik. *Die Naturwissenschaften, 19*, 713–722.

Reichenbach, H. (1956). *The direction of time*. Berkeley: University of California Press.

Ridley, M. (2000). *Mendel's demon: Gene justice and the complexity of life*. London: Weidenfeld & Nicolson.

Rosen, E. (1984). *Copernicus and the Scientific Revolution*. Malabar, Florida: Robert E. Krieger Publishing Company.

Rueger, A. (2000). Emergence—diachronic and synchronic. *Synthese, 124*, 297–322.

Sahlins, M. (1977). *The use and abuse of biology*. London: Tavistock Publications.

Savitt, St. F. (2001). Being and becoming in modern physics. *Stanford Encyclopaedia of Philosophy*.

Schlegel, R. (1968). Time and the physical world. Dover Publications.

Schlosshauer, M. (2008). *Decoherence and the quantum-to-classical transition*. Heidelberg, London, New York: Springer.

Schöne, A. (1982). *Aufklärung aus dem Geist der Experimentalphysik*: *Lichtenbergsche Konjunktive*. C. H. Beck.

Schrödinger, E. (1928). Conceptual models in physics and their philosophical value. In E. Schrödinger (Ed.), *Science theory and man* (pp. 148–165). London: George Allen and Unwin.

Schrödinger, E. (1947). Die Besonderheit des Weltbilds der Naturwissenschaften. In E. Schrödinger (Ed.), *Was ist ein Naturgesetz*? (pp. 27–85). München: Oldenbourg (^{4}1987).

Searle, J. (2004). *Mind: A brief introduction*. Oxford: Oxford University Press.

Searle, J. (2010). *Making the social world*. Oxford: Oxford University Press.

Shoemaker, S. (1969). Time without change. *The Journal of Philosophy*, *LXVI*, 363–381.

Sorensen, R. (1992). *Thought experiments*. Oxford: Oxford University Press.

Schulman, L. (1997). *Time's arrow and quantum measurement*. Cambridge: Cambridge University Press.

Steward, H. (2011). Moral responsibility and the concept of agency'. In R. Swinburne (Ed.) (pp. 141–157).

Strawson, G. (2011). The impossibility of ultimate responsibility? In Swinburne (Ed.) (pp. 126–140).

Suárez, M. (1999). The role of models in the application of scientific theories: Epistemological implications. In M. S. Morgan & M. Morrison (Eds.) (pp. 168–195).

Suárez, M. (2003). Scientific Representation: Against similarity and isomorphism. *International Studies in the Philosophy of Science, 17*(3), 225–244.

Suárez, M. (2004). An inferential conception of scientific representation. *Philosophy of Science, 71*, 767–779.

Suárez, M. (2010). Scientific representation. *Philosophy Compass, 5*(1), 91–101.

Swinburne, R. (ed.). (2011a). *Free will and modern science*. British Academy.

Swinburne, R. (2011b). Dualism and the determination of action. In R. Swinburne (Ed.) (pp. 63–83).

Szilard, L. (1929/1983). On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. In: J. A. Wheeler & W. H. Zurek (Eds.), *Quantum theory and measurement* (pp. 539–548). Princeton University Press (1983).

Thorne, K. S. (1995). *Black holes and time warps*. London: Papermac

Tipler, F. J. (1980). General relativity and the eternal return. In F. J. Tipler (Ed.), *Essays in general relativity* (pp. 21–37). New York: Academic Press.

Treisman, M. (1999). The perception of time. In J. Butterfield (Ed.), *The arguments of time* (pp. 217–246). Oxford: Oxford University Press.

Uffink, J. (2001). Bluff your way in the second law of thermodynamics. *Studies in History and Philosophy of Modern Physics, 32*, 305–394.

Van Strien, M. (2014). On the origins and foundations of Laplacian determinism. *Studies in History and Philosophy of Science, 45*, 24–31.

Von Baeyer, C. (1998). *Maxwell's demon*. New York: Random House.

Von Mises, R. (1930). Über kausale und statistische Gesetzmäßigkeit in der Physik. *Die Naturwissenschaft, 18*, 145–153; also in *Erkenntnis* I (1930), 885–893.

Von Mises, R. (1931). Über das naturwissenschaftliche Weltbild der Gegenwart. *Die Naturwissenschaften, 18*, 885–893.

Von Weizsäcker, C. F. (1939). Der zweite Hauptsatz und der Unterschied von der Vergangenheit und Zukunft. *Annalen der Physik* (5. Folge) *36*, 275–283.

Warren, R., & Warren, R. P. (1968). *Helmholtz on perception.* Wiley.

Watson, G. (Ed.). (2003). *Free will*. Oxford: Oxford University Press.

Weber, M. (1904). "Objectivity" in social science and social policy. Reprinted in *Max Weber on the methodology of the social sciences* (E. A. Shils & H. A. Finch, Trans.). Illinois: The Free Press of Glencoe.

Weber, M. (1905). Objective possibility and adequate causation in historical explanation. Reprinted in: *Max Weber on the methodology of the social sciences* (E. A. Shils, H. A. Finch, Ttrans.). Illinois: The Free Press of Glencoe (1949).

Weber, M. (1922/1968). *Economy and society*. In G. Roth & C. Wittich (Eds.) New York: Bedminster Press.

Weinert, F. (1993). Laws of nature: A structural approach. *Philosophia Naturalis, 30*(2), 147–171.

Weinert, F. (1995). Laws of nature—Laws of science. In F. Weinert (Ed.), *Laws of nature* (pp. 3–65). Berlin: DeGruyter.

Weinert, F. (1996). Weber's ideal types as models in the social sciences. In A. O'Hear (Ed.), *Verstehen and humane understanding, Royal Institute of Philosophy* (Vol. 41, pp. 73–93). Cambridge: Cambridge University Press.

Weinert, F. (1999). Theories, models and constraints. *Studies in History and Philosophy of Science, 30*(2), 303–333.

Weinert, F. (2000). The construction of atom models: Eliminative inductivism and its relation to falsificationism. *Foundations of Science, 5*, 491–531.

Weinert, F. (2004). *The scientist as philosopher*. Heidelberg, Berlin, New York: Springer.

Weinert, F. (2006). Einstein and the Representation of Reality. *Facta Philosophica, 8*(1–2), 229–252.

Weinert, F. (2009). *Copernicus, Darwin and Freud*. Wiley.

Weinert, F. (2013a). *The march of time*. Heidelberg/Berlin/New York: Springer.

Weinert, F. (2013b). The past-future asymmetry. In Raji C. Steineck & Claudia Clausius (Eds.), *Origins and future (The study of time XI)* (pp. 139–166). Leiden, Boston: Brill.

Wendorff, R. (^{3}1985). *Zeit und Kultur*. Opladen: Westdeutscher Verlag.

Wheeler, J. (1967/1997). It from Bit. In J. A. Wheeler (Ed.), *At home in the universe.* American Institute of Physics.

Whitrow, G. J. (1980). *The natural philosophy of time*. Oxford: Oxford University Press.

Whitrow,G. J. (1989). *Time in history*. Oxford: Oxford UP

Wright, K. (2006). Times of our lives. In *Scientific American* (Special Edition: *A Matter of Time*) (Vol. 16, no. 1, pp. 26–33) (2006).

Yarrow, K., Haggard, P., Heal, R., Brown, P., & Rothwell, J. C. (2001). Illusory perceptions of space and time preserve cross-saccadic perceptual continuity'. *Nature, 414*, 302–305.

Zeh, H. D. (1992/^{2}1996). *The physical basis of the arrow of time.* Heidelberg, Berlin, New York: Springer.

Zeilik, M. (^{5}1988/^{9}2002). *Astronomy*. New York: Wiley.

Index

F. Weinert, *The Demons of Science*,
DOI 10.1007/978-3-319-31708-3