

Chapter 2

Time and Cosmology

Our theory of clocks has great influence on our understanding of time. (J. R. Lucas, *A Treatise on Time and Space* 1973, Part I, Sect. 10)

Running through the whole history of time reckoning is a deep connection between time and cosmology. It is essential for the measurement of time, since it provides periodic regularity. The early Greeks established the association of time with periodic regularity on a planetary scale. Regularity, however, is not a sufficient condition for the objective measurement of the passage of time. A further condition is needed: invariance, the importance of which came to the fore with the Special theory of relativity. The measurement of time is here understood as the measurement of less regular intervals by more regular intervals of events, and the improvement in accuracy, which can thereby be achieved through the method of triangulation. In this volume time is meant in a physical sense, which is based on some regular, material events in the physical universe; it is not meant in a metaphysical sense, according to which there exists some entity called TIME in the physical universe over and above its matter, motion and energy. Equally, when the discussion switches to the ‘direction’ of time, what is meant is the direction of processes and events *in* time, not a direction of Time itself.

2.1 Greek Astronomy

(...) if we seek to examine Time, we find ourselves examining Reality. (Gunn 1930, 369)

The intellectual labour of the Greeks produced a first coherent cosmology: the Aristotelian-Ptolemaic theory of geocentrism. According to this view, the Earth resides motionless at the ‘centre’ of the universe. This universe was essentially identical with the existence of the then known six planets. The planets, which included the sun, were understood to be carried around the central Earth on circular orbs, in known periods, from west to east. The fixed stars, beyond the planetary spheres, also circled the stationary Earth in a 24 hour-rhythm from east to west.

The planets were the ‘wandering stars’ because their movements could be observed against the background of ‘fixed’ stars, which did not seem to change their positions in the stellar constellations over long periods of time. In the history of astronomy Plato’s cosmology, as presented in the *Timaeus*, is usually neglected, while Aristotle’s cosmology figures more prominently. Aristotle conceived of the cosmos as a two-sphere universe (Aristotle 1952a, b; cf. Weinert 2009, Chap. I). The region between the Earth and the moon was the *sublunary* sphere; the region beyond the moon to the fixed stars was the *supralunary* sphere. The sublunary sphere was the area of asymmetry, change and flux. The supralunary sphere, in which the planets circle the Earth, was characterized by perfection, permanence and symmetry. This cosmic division reflects the fundamental distinction in Greek metaphysics between Parmenidean stasis and Heraclitean flux. As the planets’ orbits were located in the supralunary sphere, their journey around the stationary Earth was supposed to be circular, for the circle was a perfect geometric figure in Greek culture. Whilst the sublunary sphere was characterized by changing events, by decay and renewal, the planets moved around the central Earth with perfect periodic regularity. The Greeks knew the order of the planets, although they placed the sun where in today’s Copernican worldview the Earth is located and they knew the orbital periods of the planets fairly accurately. Unlike Aristotle, who stipulated that the Earth sits motionless at the centre of the universe—a view, which was later refined by Ptolemy—Plato seems to have attributed a diurnal motion to the Earth (Taylor 1926, 449–454; Cornford 1939, 120–134). Plato was not the only Greek who bestowed a diurnal rotation of the Earth on its own axis. In the 4th century B.C. Heraclides of Pontus proposed that the daily rotation of the stars was produced by the daily rotation of the Earth on its own axis. Aristarchus of Samos (310–250 B.C.) went further and placed the sun at the centre of the universe and gave the Earth a circular orbit around it. But this view had little impact on subsequent developments, since it seemed to contradict the perceptual evidence. To a Greek observer it seemed obvious that ‘the heavens’ moved and that the Earth stood still. A powerful argument against the motion of the Earth on its own axis, which prevailed until the work of Copernicus (1543), was that under the force of the rotation—similar to the experience on a spinning wheel—buildings would crumble and violent winds would blow from east to west, stopping birds from flying eastwards.

2.2 Plato and Aristotle

Whilst Plato’s cosmology had little impact on the subsequent history of astronomy, his views on time were more influential. They mark a clear example of the association of time with cosmic regularity, which persists to the present day. As is well-known Plato (427–347 B.C.) drew a distinction between the changing world of our daily experience and an underlying realm of unchanging forms. The visible world is the world of becoming and change; but it is a shadowy image of an eternal world of unchanging, permanent forms. According to Plato there can be no exact

science of the natural world, because it is subject to change and flux. Physics only tells a likely story but mathematics deals with timeless forms—for instance geometric objects, like circles and triangles—and they alone can be the object of rational understanding (Cornford 1937, Prelude). And yet there is time. Time, in Plato's view, is the 'moving likeness of eternity'. But, significantly, Plato adds that this image of eternity moves according to number, by the will of the demiurge:

But he took thought to make, as it were, a moving likeness of eternity; and, at the same time that he ordered the Heaven, he made, of eternity that abides in unity, an everlasting likeness moving according to number – that to which we have given the name Time. (*Timaeus* 37D; quoted in Cornford 1937, 98; see also Taylor 1926, Chap. XVII; Benjamin 1966; Whitrow 1966; Jammer 2007)

But as the sensible world is subject to irregular fluctuations and oscillations, a measurable passage of time cannot be identified in this realm. Time depends on periodic regularity of motion, which is to be located in the movements of the planets.

(...) Time came into being together with the Heaven, in order that, as they were brought into being together, so they may be dissolved together, if ever their dissolution should come to pass; and it is made after the pattern of the ever-enduring nature, in order that it may be as like that pattern as possible (...) (*Timaeus* 38B, C; quoted in Cornford 1937, 99)

Firstly, then, Plato identifies time with regular physical events, like the motion of celestial bodies; secondly, such orbital motions mark off periods of time, due to their regularity. It takes a planet a certain amount of time—from 87 days (Mercury) to 164 years (Neptune), and 284 years (Pluto)—to return to its previous position. These orbital periods can be used to define periods of human time: the calendar year (orbit of the Earth) or the month (lunar cycle). 'Plato's view of time is inseparable from periodic motion', (Cornford 1937, 103), for in this way time can be 'the moving image of eternity'. The changing events on Earth are too irregular, in Plato's view, to constitute physical time; the underlying reality of eternal forms simply exists in a timeless sense of pure being. The planets are therefore an ideal instrument of time (*Timaeus* 38C–39E). Their observable motion is periodic, regular and, according to Greek cosmology, everlasting; hence invariant.

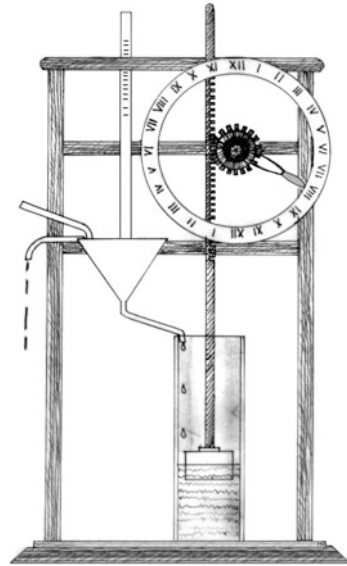
Aristotle (384–322 B.C.) criticized Plato for his theory of forms and his theory of time. Aristotle objected that time cannot be identified with celestial motion: since motion is measured in time, time cannot be measured by motion. Motion can be fast or slow and change, which accompanies motion, has a location in space but time as such cannot be perceived, although we are conscious of the passing of time when we discern change and movement. Aristotle granted that time is dependent on change or motion and awareness of time depends on an awareness of 'before' and 'after' in change. The relation between time and change is reciprocal: Without change, there can be no recognition of time; and without time, there can be no measurement of change. 'The time marks the movement, since it is its number, and the movement the time' (Aristotle 1952a, iv, 220b). But Aristotle did not provide a theory of time; rather he sketched a theory of the measurement of time. In Aristotle's view, time becomes quantifiable change from 'before' to 'after' in the

course of events; hence Aristotle associated regular motion with measurement and magnitude. He argued that regular change was a necessary condition of the measurement of time. Whilst there are many particular movements, which may change and cease, there is one type of regular movement—the eternal circular motion of celestial objects—which provides the perfect measure of time. He regarded time as a numbering process, associated with our perception of a ‘before–after’ relationship between events in motion. Motion is restricted to qualitative (alteration), quantitative (change in size) and local motion (change in place). Time is quantifiable change from ‘before’ to ‘after’, thus implying measurement and magnitude. Aristotle furnishes a definition of the measurable duration of events, rather than a theory of time, since Aristotle remains silent on the question of the nature of time (cf. van Fraassen 1970, 11–17; Whitrow 1989, 42; Benjamin 1996, 12–15; Kronz 1997; Jammer 2007).

But is Aristotle right in his view that time cannot be identical with motion because motion can be slow or fast and it is measured in time but not *vice versa*?¹ It is true that many motions would be unsuitable for an identification of physical time: the motion of particles in a liquid, the motion of gas molecules in a container, the motion of pedestrians in a city, the motion of cars on a motorway are all too irregular to qualify. But if care is taken to distinguish between *physical* and *human* time, some regular physical events will be needed to identify physical time. Experience tells us that less regular motion can be measured by more regular motion. A person’s heartbeat could be used to measure the movement of pedestrians in the street: how long it takes them to cross it, how long it takes them to move from its northern to its southern end. An obvious disadvantage is that heartbeat is irregular so that such measurements will be inaccurate. But heartbeat could be replaced by a more regular clock, say a water clock. So gradually—by a process of triangulation—the precision of clocks would improve until fairly reliable clocks become available. The best clocks at the disposal of Aristotle’s contemporaries would have been sundials and water clocks (Clepsydres, Fig. 2.1), which were already known in Egypt in 1600 B.C. Nevertheless, the Greeks argued that the measurement of time should be based on celestial motion, since it was regular and periodic; it constituted physical time. Aristotle, however, has a point when he refuses to identify circular celestial motion with time. For there are several notions of time: physical time, human time, social time, psychological time and they cannot all be identified with celestial motion. And subsequent centuries replaced the orbit of planets with more robust processes to associate them with physical time. Human time is an abstraction from the observation of diverse physical processes, which give rise to the division of the year and the day, according to our familiar calendars.

¹ As we shall see, John Locke (*Essay* BK XIV, Sect. 21) also recommended a distinction between duration ‘in itself’ and ‘the measures we make use of to judge of its length.’

Fig. 2.1 Clepsydra,
Wikimedia Commons



In order to grasp the distinction between physical and human time, it is important to distinguish *natural* and *conventional units of time*. Natural units of time are based on periodic processes in nature, which recur after a certain interval. They may be quite imprecise, like the periodic flooding of the Nile, on which the ancient Egyptians based their calendar year; or more regular, like celestial phenomena. Some basic units of time, like the day and the year, are based on natural units of time. For instance, the equatorial rotational period of the Earth is 23 h 56 min and 4.1 s; that of Uranus is 17 h (Zeilik 1988, 508). The tropical year—the time that the Earth needs for one revolution around the sun—has a length of 365, 242,199... days or 365 days, 5 h, 48 min and 46 s (see Moyer 1982; Clemence 1966). But the calendar year has 365 days and 366 in leap years, which gives the calendar year an average length of 365.2425 days. As calendar years cannot have fractional lengths, there will always be a discrepancy between the tropical and the calendar year. This difference led to the replacement of the Julian calendar by the Gregorian calendar (1582). The Gregorian calendar will remain accurate to within one solar day for some 2,417 years. One difficulty with the day and the year, as just defined, is that these units of time are not constant, due to slight irregularities in the motion of the Earth. Historically, this discrepancy has led to calendar reforms and redefinitions of the ‘second’ from a fraction of the rotational period of the Earth around the sun to atomic oscillations.

Whilst physical time is based on such natural units, human time is 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), again 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 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. Note, however, that not all such conventions are arbitrary. The equinoxes, the summer and winter solstices correspond to particular positions of the Earth with respect to the sun. 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 12 months (4000 B.C.) 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 (Zeilik 1988, 152; Wendorff 1985, 14), but the solar year has 12.368 lunar months. 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 12 h 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).

Despite these aspects of conventionality, it must be emphasized that the conventional units of time must keep track of natural units of time. 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 inertial 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. The existence of physical time can be justified by a consideration of the consequences of idealist and empiricist notions of time (cf. Schlegel 1968, Preface, Chap. I; Rugh and Zinkernagel 2009).

2.3 The Need for Physical Time

Time and Space are among the fundamental physical facts yielded by our knowledge of the external world. (Gunn 1930, 215)

What happens when physical time is neglected can be gleaned from a consideration of Saint Augustine's famous reflections on time and by the attempts of the British empiricists to grapple with the notion of time.

2.3.1 Saint Augustine (354–430 AD.)

For Saint Augustine time emerges with the creation of the material world by a Deity. Saint Augustine rejects as nonsensical the question whether time existed before God's creation of the universe. God exists in a timeless manner. Time is co-existent with the creation of material events but not co-eternal with God's existence (Augustine 1961, Bk. XI, Sect. 14). At first, Saint Augustine associates time with an awareness of a 'before–after' relationship between physical events. These material events come into being with the creation of the universe. As will be discussed in [Chap. 3](#), Saint Augustine partly embraced what was later to be called a 'relational view' of time (see Box I). This is essentially the view that physical time depends on a 'before–after' relationship, a succession of events in the material universe. Saint Augustine also rejects the Platonic view that physical time is identified with celestial motion because he refers to a statement in the Bible, according to which Joshua made the sun stand still so that a battle could be won (Augustine 1961, BK. XI, Sect. 23). He does not specify which events he has in mind; in particular he does not specify whether the 'before–after' relation between events is of a regular or irregular nature. Saint Augustine's reflections start with the postulation of physical time through the creation of material events. Then his reflections turn to the question of the measurement of time. But when he asks himself how time is measured, surprisingly he does not refer to physical clocks, which were available at his time: water clocks, shadow clocks, sun dials. He slips into psychological language and makes the human mind the metric of time. We remember the past, we are aware of the present, we anticipate the future. There is, however, a certain problem with the awareness of the present. Saint Augustine arrives at the view that 'the present is without duration' because he entertains a mathematical notion of time, in which each interval of time can be subdivided into ever smaller intervals. Consider the time it takes to utter a phrase like 'Saint Augustine'. A speaker needs a certain amount of time to pronounce this name, so that this phrase can be further subdivided into past and future segments. As s/he speaks, the sounds seem to emerge from the future only to recede irretrievably into the past. Perhaps then the present is not the time it takes to utter the phrase 'Saint Augustine' but occurs when a particular syllable is spoken. But even the utterance of a syllable takes time and this interval can be further subdivided into past and future. As long as an event has duration it can be further subdivided into future and past moments. By this kind of mathematical reasoning, Saint Augustine concludes that the present has no duration. It is a metaphorical knife edge. Despite this conundrum time seems to be measurable.

Box I: A brief Introduction to the Philosophy of Time

It is convenient to distinguish three influential philosophies of time, to which the discussion will return throughout.

- The *realist view* (I. Barrow, I. Newton) is the view that time is a physical property of the universe, over and above other properties, and that time would exist even in an empty universe. This view can be expressed in the slogan that the ‘universe has a clock’, since according to this view some master clock exists which measures the ticking of time across the whole universe, even if no physical events occurred in this universe. In modern discussions it reappears as the block universe, often accompanied by an idealist view of time.
- The *relational view* (G. Leibniz, E. Mach, partly Saint Augustine) is the view that time depends on the succession of physical events in the universe, such that time would not exist in an empty universe. This view can be expressed in the slogan that ‘the universe is a clock’ since according to this view there is no master clock and clocks depend on the existence of regular, periodic processes in the universe. This view is a version of the Heraclitean view of flux and becoming.
- The *idealist view* (I. Kant, partly Saint Augustine and, surprisingly, many physicists and philosophers) is the view that time is a property of the human mind. The passage of time in the physical universe is an illusion or a human construction, and hence time depends on the existence of human observers. A complement of this view is the so-called block universe, i.e. the view that physical reality is a timeless, unchanging being. This view is a version of the Parmenidean view of stasis.

We can be aware of time and measure it only while it is passing. Once it has passed it no longer is and therefore cannot be measured. (Augustine 1961, BK. XI, Sect. 16)

The future has not yet arrived and the past has gone. So how do we measure the present time, if it has no duration? Time is coming out of what does not yet exist, passing through what has no duration and moving into what no longer exists (Augustine 1961, BK. XI, Sect. 21). What measurable period could be used as a yardstick to measure temporal intervals? Some of his Greek predecessors offered the periodic regularity of planetary motion as a metric but Saint Augustine is as dissatisfied with this answer as Aristotle was. The movement of bodies is always measured in time, but time is never measured by reference to the movement of bodies, including the sun (Augustine 1961, BK. XI, Sects. 23–24). The measurement of bodies in motion presupposes a notion of time. ‘It is by time that we measure the course of the sun’ (Augustine 1961, BK. XI, Sect. 23). Secondly, bodies (including the sun) move more or less quickly or remain at rest but their motion and rest are still measured by way of time. This answer is of course

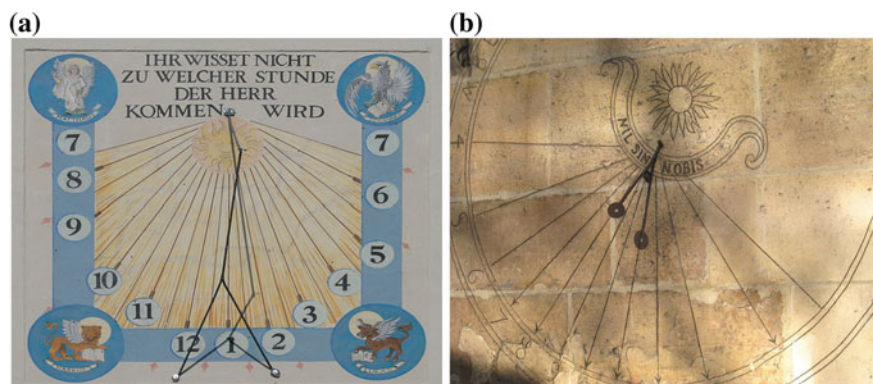


Fig. 2.2 Sun dials, Wikimedia Commons

inaccurate, since the Greeks already employed the relative regularity of physical processes to construct sundials, shadow clocks and water clocks as time pieces to measure less regular intervals; and they regarded the motion of the planets as circular and uniform (Fig. 2.2a, b). But Saint Augustine does not consider these possibilities. This neglect leaves the mind as the metric of time. It is the human mind, which measures time and its ‘flow’. As time is not objective, the human mind measures the impressions, which the passing of external things leave on it. Events flow from future to past, through the present and leave traces on the human mind. It is these impressions, which the mind measures. A long past is a long remembrance of the past; and a long future is a long expectation of the future (Augustine 1961, BK. XI, Sect. 28). The human mind becomes a metric of time: it has an expectation of future events, it is aware of the passing of present events and it remembers past events.

Saint Augustine argues from objective beginnings of physical time to time as the extension of the mind (Augustine 1961, BK. XI, Sect. 26). The conclusion is that the mind can be aware of time and measure it only while it is passing (Augustine 1961, BK. XI, Sect. 16). It is only in individual minds that time is measured, which makes time a subjective experience (cf. Lucas 1973, Pt. I, Sects. 1, 2 for a modern defense of mental time).

Thus Saint Augustine makes an implicit distinction between the ‘before–after’-relation between material events and the ‘past–present–future’ relation, which depends on human awareness.

Consider an analogy with sound. A group of observers sits on a hill above a tennis court in the distance. They can see the players on the court but as sound travels more slowly than light they experience a curious effect. They watch the two players serve the ball before they hear the sound. So on Saint Augustine’s analysis of the measurement of time, when they see player A serve the ball, the sound still lies in the future although its arrival is anticipated. Then the sound arrives and is perceived as the sound waves pass by the observers. The sound then recedes into the distance beyond their earshot; it is in their past and they only remember it.

Events pass from the future into the past but they leave an impression on the mind, and it is this impression, which the mind measures. We do not say ‘past time is long’ because the past no longer exists. We say that a long past is a long remembrance of the past. And a long future is a long expectation of the future (Augustine 1961, BK. XI, Sect. 28).

Saint Augustine’s conception has severe limitations:

1. Although Saint Augustine admits that the awareness of time requires a ‘before–after’ relation between events and physical time is co-existent with the creation of the material world—the measurement of time occurs in the mind and therefore possesses no objectivity. But Saint Augustine cannot evade the need for physical time, since he admits that there is a succession of events in the physical world (Augustine 1961, BK. XI, Sect. 27). This succession of events is objective although Saint Augustine does not consider it worthy as a candidate for a metric to measure the duration of time between two events. Saint Augustine suggests that the mind can serve as a metric thus rejecting several other possible metrics. As mentioned, during Saint Augustine’s time water clocks, shadow clocks and sun dials existed, which could have provided a rough but objective measure of the duration of events. The choice of a metric is conventional but the occurrence of material events is not as they are separated by intervals between an earlier and a later event (see van Fraassen 1970, 77). Saint Augustine treats time as an extension of the mind, which made him the founder of psychological time. But psychological time falls victim to the vagaries of psychological states. Thus, Saint Augustine’s notion becomes idealist and subjective at the same time, since time is measured in individual minds as a result of impressions left on them through passing events. But his notion of time also contains relational elements since he begins his reflections by the observation that we are only aware of time where there is change in the physical world. This change affects the human mind and the past-present-future relation as categories of the mind; it gives rise to a subjective measurement of time.
2. Human time only exists in individual minds, which constitute metrics of time. This thesis reduces time to psychological time. But psychological time suffers from two defects: first, it lacks *regularity* since the way an individual perceives the duration of a particular event depends on their psychological states; second, it lacks *invariance* since the same individual cannot compare the assessment of duration on two different occasions, nor can two individuals compare their respective evaluation of elapsed time on the same occasion. Psychological time is not invariant across different perspectives. But both regularity and invariance are important features of a measurable passage of time and have important philosophical consequences.
3. According to Saint Augustine the present time is mathematically divisible into infinity so that any present moment, however, short, can always be subdivided into smaller units. But this is not a statement about physical time. Saint Augustine conflates the present as a physical interval—a period of some duration—and as a mathematical instant—a duration-less point (Lucas 1973,

Pt. I, Sect. 4; cf. Newton-Smith 1980, Chap. VI; Denbigh 1981, Chap. 3, Sect. 3). There could exist a shortest possible time interval: for some time physicists have speculated that there might exist a ‘chronon’, a unit of time which is determined by the time it takes a light signal to cross the diameter of an electron; more recently the suggestion has been made in the area of quantum gravity that the shortest possible moment of time is defined by Planck time. This is a much shorter interval but it would still make time discrete (cf. Smolin 2006).

2.3.2 *David Hume (1711–1776) and John Locke (1632–1704)*

In Saint Augustine’s reflections we are faced with a curious mixture of a relational view and an idealist view of time. The relational view of time was developed by G. W. Leibniz in the 17th century and it is based on the need for physical time. The idealist view of time was developed by I. Kant in the 18th century in opposition to both Leibniz’s relational view and Newton’s realist view of time. As we shall see Kant’s idealist view of time is beset by the same difficulty as Saint Augustine’s view of the measurement of time. Kant, despite his avowal that time exists only in the mind, implicitly acknowledges the need for physical time in his employment of the notion of causality. But these difficulties are not restricted to idealism; they also afflict the views of the British empiricists. As is well-known the British empiricists take the mind to be a *tabula rasa*, on which the external world leaves impressions, which by cognitive processes are transformed into ideas. But the basic tenet of British Empiricism is that there can be no succession of impressions without a succession of perceptions, which themselves are caused by the succession of real, perceived events. British Empiricism presupposes a succession of events in the physical world, without which no mental impression of objects in the physical world could be formed. Both David Hume and John Locke base their reflections on time on this basic tenet. Thus Hume opens his *Treatise of Human Nature* (1739) with the statement that ‘all the perceptions of the human mind resolve themselves into two distinct kinds’: impressions and ideas. Their difference lies in the degree of liveliness and force, with which they are perceived; their difference is one of ‘feeling’ (sensation) and ‘thinking’ (reflection). For Hume ‘our impressions are the causes of our ideas’ (Hume 1739, Bk. I, Section I). The order of events is then that impressions strike upon our senses, producing perceptions of hunger, heat, pain, pleasure etc.; and these impressions then give rise to ideas (copies of impressions) in our minds (Hume 1739, Bk. I, Section III).

Consequently, both space and time are abstract ideas—the first being characterized by extension, the second by duration. Hume defines time as a succession of perceptions in the human mind. Although Hume accepts that without a succession

of impressions and ideas there can be no notion of time, he does not accept a notion of physical time:

Whenever we have no successive perceptions, we have no notion of time, even though there is a real succession in the objects. (Hume 1739, 35)

Hume seems to fall prey to his own empirical presuppositions. He sees impressions as the causes of ideas, and from this basic tenet of Empiricism he concludes that time 'is always discovered by some perceptible succession of changeable objects' (Hume 1739, 35; italics in original). By tying the notion of time to *perceivable* successions, Hume 'refuses to time any reality save that of a succession of ideas'. But Hume does not seem to realize that this 'mental synthesis' presupposes an awareness of objective change (Gunn 1930, Sect. III.8; Kronz 1997, Sect. 5). Hume implicitly acknowledges the need for physical time, irrespective of human perceptions, because 'perceivable successions of changeable objects' are simply a subset of changeable objects. A perceivable succession presupposes a succession of changeable events. The events and objects affect our senses, and thus lie at the root of our impressions.

It is to John Locke's credit that in his *Essay Concerning Human Understanding* (1689) he shows awareness of this basic difficulty in British Empiricism. This awareness makes Locke's reflections on time more sophisticated than Hume's but ultimately he fails to overcome the difficulty. Locke, like Hume, locates the notions of duration and succession in the flow of ideas in human minds. Ideas themselves have their sources in sensations or reflections. The ultimate source of our knowledge is experience and the experience of 'external sensible objects' (sensation) is one source of our knowledge. But the mind can reflect on these ideas, originating from sensations, and make them 'the object of mental operations' (Locke 1964/1690, Bk II, Chap. I). After this empiricist introduction Locke proceeds to apply his ideas to the notion of time. He makes a distinction between *duration* and *time* (Locke 1964/1690, Bk. II, Chap. XIV; cf. Gunn 1930, Sect. III.5). The perception of duration derives from a succession of ideas in the mind. The perceiver is only aware of the train of ideas when s/he is awake but not during hours of sleep. When 'that succession of ideas ceases' in moments of unconsciousness, 'our perception of duration ceases with it'. Of duration, then, Locke furnishes a very psychologistic account, for 'we cannot perceive that succession (of ideas) without constant succession of varying ideas during our waking hours'. Yet Locke is aware, like Saint Augustine, to whom he refers, that the idea of duration, arising from reflection on the succession of ideas in the mind of the perceiver, gives rise to a consideration of measure. The mind needs to

get some *measure* of this common duration, whereby it might judge of its different lengths, and consider the distinct order wherein several things exist, without which a great part of our knowledge would be confused and a great part of history rendered useless. This consideration of duration, as set out by certain periods, and marked by certain measures or epochs, is that, I think, which most properly we call *time*. (Locke 1964/1690, Bk.II, Chap. XIV, Sect. 17)

Thus Locke distinguishes between the passage of time (duration) and the measurement of time, and reserves the notion of time to measurable duration of external events. Locke recognizes the need for both human time, which for him resides in the flow of ideas, and physical time, which resides in objective physical change. But Locke grants no priority to physical time, and bestows more significance on human time (duration) than on physical time. Just like Aristotle before him, he denies that the idea of succession arises from the observation of motion. We have to distinguish carefully ‘betwixt duration itself’ and the measure of duration for our measures of duration are a matter of choice.

For the freezing of water, or the blowing of a plant, returning at equidistant periods in all parts of the Earth, would as well serve men to reckon their years by, as the motions of the sun; and in effect we see that some people in America counted their years by the coming of certain birds amongst them at their certain seasons, and leaving them at others. (Locke 1964/1690, Bk.II, Chap. XIV, Sect. 20)

Our choice of measure is a matter of convention. More importantly, we have no guarantee of the equality of any pairs of duration, ‘for two successive lengths of duration, however, measured, can never be demonstrated to be equal’ (Locke 1964/1690, Bk. II, Chap. XIV, Sect. 21). Locke points out, as Newton did at the same time, that the motion of the sun suffers from irregularities. But even the ‘two successive swings of a pendulum’ cannot be known to be equal.

Since then no two portions of succession can be brought together it is impossible ever certainly to know their equality. (Locke 1964/1690, Bk. II, Chap. XIV, Sect. 21)

In view of such apparent uncertainty, Locke gives preference to human time. The ideas of succession and duration are formed ‘by reflection on the train of our own ideas’ (Locke 1964/1690, Bk. II, Chap. XIV, Sect. 27). However, this lapse into psychologism and focus on mental time, at the neglect of physical time, suffers from the same defects as Saint Augustine’s psychological time. Mental time is essentially private, which makes it impossible to compare two temporal intervals, which succeed each other. Individual observers cannot know, in principle, whether their ‘measurements’ of the duration of external events is regular and invariant. Thus Locke’s objection to the measurement of physical time also applies to his mental time. Compare this situation to the use of a mental yardstick to measure some spatial length. A joiner is repairing a piece of furniture and needs a short piece of wood. No meter is to hand and he can only rely on his intuition. He has some suitable pieces of wood in his van but he must choose their lengths from memory. His task is impossible. To remember the duration of events, in the absence of any clocks, is equally difficult. There is no regularity in memories and they are dependent on particular perspectives or psychological states; hence they are not invariant. To secure regularity and invariance, there is a need for physical time. But how can the difficulties with the measurable passage of time, emphasized by Aristotle and Locke, be overcome? In the history of time reckoning this problem has been solved by the method of triangulation. We measure less regular processes by increasingly more regular processes, and eventually arrive at fairly accurate time pieces. Locke

emphasizes himself that the ‘revolutions of the sun’ are not the only measure of duration. In his own time his contemporaries began to use the pendulum ‘as a more steady and regular motion than that of the sun or (to speak more truly) of the Earth’ (Locke 1964/1690, Bk. XIV, Sect. 21). Although two successive swings of a pendulum cannot be compared in order to see whether they are equal what can be compared are the oscillations of a pendulum against, say, the rotations of a spinning sphere. Locke acknowledges as much when he speaks of the relations of time.

Thus, when anyone says that Queen Elizabeth lived sixty-nine and reigned forty-five years, these words import only the relation of that duration to some other, and mean no more but this, that the duration of her existence was equal to sixty-nine, and the duration of her government to forty-five annual revolutions of the sun; and so are all words, answering, *How long?* (Locke 1964/1690, Bk. II, Chap. XXVI, Sect. 3)

In the end, then, Locke refers the measurement of time to cosmological processes, like the ancients. Since the Copernican Revolution the dual mobility of the Earth has served humanity as a yardstick for the measurement of the passage of time. We have to turn to Kant’s cosmology to witness the first appearance of an arrow of time.

2.4 Kant’s Cosmology

Before Kant (1724–1804) developed his idealist view of time, he proposed an evolutionary view of the cosmos. His idealist view of time is, as we shall see later, objective rather than subjective. Kant had inherited the Copernican worldview, first proposed in the modern age by Nicholas Copernicus (1543) and later refined by Kepler, Galileo, Descartes and Newton, according to which the sun was the centre of the solar system and the planets orbited the sun in elliptical orbits. But when Kant published his *General History of Nature and Theory of the Heavens* (1755) the extent of the universe was no longer confined, as Copernicus had still assumed, to the solar system and the fixed stars on the horizon. Kant’s treatise is the first systematic attempt to give an evolutionary account of cosmic history (Whitrow 1989, 153). Kant’s evolutionary history is limited to the physical universe, since he explicitly excludes living organisms from his evolutionary considerations. Kant depicts the order of nature as an unfolding, ongoing process and distances his view from the Biblical story of a six-day creation process.

The Creation is never finished or complete. It did indeed once have a beginning but it will never cease. (Kant 1755, Seventh Chapter, 145; cf. Toulmin and Goodfield 1965, 130)

According to Kant the cosmos reveals a dynamic history and is no longer split into two spheres: the realm of eternity, perfection, and symmetry as against the realm of temporality, imperfection and asymmetry. Kant conceives of the whole cosmos as an ordered structure, in analogy with the solar system. The solar system is only part of a larger structure, the Milky Way, which is our home galaxy. But

the universe plays host to other galaxies, which may only appear as 'small luminous patches' in the telescope (Toulmin and Goodfield 1965, 129–131). This whole hierarchical structure is governed by dynamic laws, similar to the ones, which Kepler had formulated for the solar system. Kant argues that the whole cosmos consists of ordered systems, like our solar system, and that the smaller systems can be understood as embedded in larger systems within galaxies. And the galaxies themselves form larger systems, which are known as clusters.

Furthermore this cosmic order was the result of the action of mechanical laws, which moulded the original chaos into the observable order of nature. The mechanical laws work in a regular, predictable fashion. The work of blind mechanisms, which drove the mechanical evolution of cosmic history, does not, in Kant's view, throw serious doubts on philosophical proofs of a divine creator (Kant 1755, Preface). Kant assumes that the Deity is the first cause, who sets the whole machinery in motion, thus leaving an 'agreement between my system and religion'. The implication is, however, that creation is not the act of a single moment (Kant 1755, Seventh Chapter, 145). Kant proposes the idea of a successive extension of 'creation' through the infinite universe. Kant thought that Nature underwent formations of order, out of chaos. The importance of this theory of cosmic evolution for the notion of time is that it throws serious doubt on the Biblical chronology of 6,000 years. Bishop James Ussher (1581–1656), for instance, had calculated the creation of the world as occurring on October 23, 4004 B.C. But Kant's position implies that the gradual establishment of order out of chaos had taken a vast period of time.

There had perhaps flown past a series of millions of years and centuries, before the sphere of ordered Nature, in which we find ourselves, attained the perfection which is now embodied in it; and perhaps a long period will pass before Nature will take another step as far in chaos. But the sphere of developed Nature is incessantly engaged in extending itself. Creation is not the work of a moment. (Kant 1755, Seventh Chapter, 145)

Working within a Newtonian paradigm, Kant relates the creation of order out of chaos to an infinite time scale:

This infinity and the future succession of time, by which Eternity is unexhausted, will entirely animate the whole range of Space to which God is present, and will gradually put it into that regular order which is conformable to the excellence of His plan. (...) The creation is never finished or complete. It has indeed once begun, but it will never cease. It is always busy producing new scenes of nature, new objects, and new Worlds. (...) It needs nothing less than an Eternity to animate the whole boundless range of the infinite extension of Space with Worlds, without number and without end. (Kant 1755, Seventh Chapter, 145)

Although Kant paints a picture of an evolutionary universe, it did not include organic nature. Kant speculated that the whole mechanical world would be understood by reference to mechanical laws before 'a single weed or caterpillar' could be explained from mechanical causes (Kant 1755, Preface, 29). Kant's evolutionary account of cosmic history does not include the evolution of species. This inclusion had to wait 104 years until Darwin published his *Origin of Species* (1859) (Weinert 2009, Sect. II.1).

According to some commentators, Kant seems to have believed in a cyclic cosmology, each cycle ranging from a Big Bang to a Big Crunch (Toulmin and Goodfield 1965, 134), an idea, which has been resurrected in modern cosmology (Penrose 2010). He certainly postulated a ‘Big Bang’ scenario, from which the present order of Nature evolved according to mechanical laws. It seems that our world has attained ‘perfection’—a common assumption amongst Enlightenment philosophers—but other worlds are still evolving towards perfection. According to Kant’s cosmic expansion hypothesis, the formation of the universe is marked by a basic *asymmetry*, since mechanical laws work on the original chaos to produce the current, ordered state of the world. But it is not clear from Kant’s statement whether he subscribed to a cycle of cosmic evolution and cosmic destruction. Rather, Kant seems to recognize that decay and destruction are built into the fabric of nature. He envisages that ‘whole worlds’ and ‘whole world orders’ will decline but he also holds that the phoenix of nature will create new worlds and new world orders.

Millions and whole myriads of centuries will flow on, during which always new Worlds and systems of Worlds will be formed, one after another, in the distant regions away from the Centre of Nature, and will attain perfection. (Kant 1755, Seventh Chapter, 145; cf. Kragh 2007, 78–83)

Thus Kant does not seem to envisage that the whole universe will expand, grind to a halt, and re-contract to collapse into a Big Crunch. Rather, regions of the universe will fall into decline to be replaced by other regions elsewhere in the universe, where ever productive nature will install new orders. As we shall see later, these speculations are surprisingly close to the views of Ludwig Boltzmann and modern cosmology. This pattern of decline and renewal, this destruction of old worlds and the ever present operation of mechanical laws, invests the cosmic evolution with an arrow of time. But Kant did not only provide an evolutionary account of cosmic history, he also developed his own theory of space and time. His views on time became very influential and were taken up by proponents of the Special theory of relativity. Kant develops this view in his *Critique of Pure Reason* (1781). It stands in stark contrast with his view on an evolutionary cosmos, which clearly implies a physical notion of time. But in his later work, Kant embraces an idealist view of time, which, in contrast to Saint Augustine’s view, is objective in nature. Still, Kant does not escape the difficulties, which accrue, when the need for physical time is neglected in one’s view of time, as will now be discussed.

2.5 Time and Causality

Space and time expand with the universe because they are the universe. (Shallis 1983, 98)

If the universe expands, there must be a succession of events, and this succession may serve as a criterion for the direction of time. Both Kant and Leibniz establish a link between the anisotropy of time and the causal succession of events, a topic,

which will be investigated in this section (cf. van Fraassen 1970, Chap. VI; Whitrow 1980, Sect. 7.2). At least Leibniz situates this link on the global scale.

2.5.1 Immanuel Kant (1724–1804)

In Saint Augustine and the British Empiricists, we have already come across the basic idea behind an idealist view of time. Saint Augustine, as shown, endorses a physical notion of time but declares individual minds its metric when he asks himself how the passage of time is to be grasped. Such a metric of time lacks regularity and invariance, since the perception of temporal intervals and durations is subject to the fluctuations of individual mental states. Saint Augustine therefore became the pioneer of *internal* time. The British Empiricists also accept a notion of physical time, as a succession of material events, but their major emphasis is on mental time. By contrast Kant proposes an objective idealist view of time but he too has to admit the importance of physical time. He holds that in the absence of human beings who perceive objects of the external world, there is no time. ‘Time is a purely subjective condition of our (human) intuition (which is always sensible, that is, so far as we are affected by objects)’. But Kant does not make individual minds the metrics of time. He holds that time and space are ‘pure forms of intuition’ (Kant 1787, B51/A35). He means that time and space belong to the *a priori* conditions of *the* human mind, amongst others, under which humans perceive the external world. All humans possess this form of intuition. Therefore, insofar as external objects appear to us and enter our experience, time ‘is necessarily objective’ (Kant 1787, B51/A35). It is objective, because all objects of the external world appear to the human mind in the same temporal order. But

it is no longer objective, if we abstract from the sensibility of our intuition, that is, from that mode of representation, which is peculiar to us, and speak of *things in general*. (Kant 1787, B51/A35)

Kant does not claim that the external, unperceived world is a timeless block universe. He simply states that the order of the ‘noumenal world’—the world of objects in itself—is unknowable to us. But in the ‘phenomenal world’ of our perception humans experience all perceivable objects as temporally ordered. Time and space may be *a priori* forms of intuition, still this inner and outer sense, respectively, must operate on the appearance of external objects. A similar situation arose for Locke and Hume. Time may be merely a ‘succession of ideas’ but the ideas are formed through mental operations on impressions and these impressions follow a sequential order. Although Locke accepts that external things appear in some regular, temporal order, which constitute time, he places more importance on the mental idea of duration (passage of time). Hume simply reduces time to the succession of perceptions.

Although these thinkers belittle the importance of physical time, they nevertheless introduce an implicit distinction between physical and human time. Kant

argues that time is a 'subjective condition of human intuition'; he nevertheless makes a concession to physical time. In his chapters on the Analogies of Experience (Kant 1787, B233–B256), he argues that a temporal order is produced through the application of the category of relation (subdivided into substance, causality, reciprocity) to perception. Kant's achievement is to have shown how objective knowledge can be obtained from the perception of external objects. In an effort to obtain an objective notion of time, the Principle of Succession in Time, in accordance with the Law of Causality is of particular importance, since it is a rule, by which humans connect cause and effect. It states: 'All alternations take place in conformity with the law of the connection of cause and effect' (Kant 1787, B232; cf. Gardner 1999, Chap. 6, esp. pp. 174–175; Brittan 1978, Chap. 7). All changes occur according to the law of causality. It is this law of causality by which succession in time is determined (Kant 1787, B246).

Kant explains its application in the following terms:

- All appearances of succession in time are only changes in the determinations of substances, which themselves abide (A189/B233). The substances are thus the invariants, whose properties undergo change. For change to be regular, as Heraclitus already suggested, there must be some invariance through change.
- All changes occur through the law of causality, that of the connection of cause and effect (Kant 1787, A189/B233). The law of causality is a rule which states that the conditions, under which an event invariably and necessarily follows is to be found in what precedes the event (Kant 1787, B246). Hence the event incorporates the antecedent conditions (cause), which determine the consequent conditions (effect). The law of causality has a temporal dimension, since for an effect to be caused, the antecedent conditions must be prior to the consequent conditions. This situation even obtains when cause and effect occur simultaneously, as in Kant's example of the ball, which presses a dent in a cushion.

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)

It is this law of causality, according to which we determine perceptions according to the succession in time, which must make a reference to external events, in order to avoid a vicious circle. The events, which succeed each other in a causal manner, do so irrespective of human awareness. The human mind, however, links them as 'cause' and 'effect'.

- Finally, objects can appear to humans as existing simultaneously, only insofar as they interact with each other. Again this perception presupposes an interaction between physical objects, before they can appear as simultaneous to the human mind.

Hence if time is to be an objective feature of the human mind, the rule of causality, by which succession in time is established, must be objective.² That is, in order to avoid a vicious circle the succession of objects must be located in the external world. Kant illustrates this contrast between subjective and objective succession through a judicious example: he compares the perception of a house to the perception of a moving ship. The order, in which the perceptions appear to the observer, is freely chosen in the first case (the house) but necessary in the second case (the ship). In the case of the house

my perceptions could begin with the apprehension of the roof and end with the basement, or could begin from below and end above; and I could similarly apprehend the manifold of the empirical intuition either from right to left or from left to right. (Kant 1787, B238)

But the case of the ship, moving downstream, is quite different:

My perception of its lower position follows upon the perception of its position higher up in the stream, and it is impossible that in the apprehension of this appearance the ship should first be perceived lower down in the stream and afterwards higher up. The order in which the perceptions succeed one another in apprehension is in this instance determined, and to this order apprehension is bound down.' (Kant 1787, B237)

In the latter case then,

(t)he objective succession will therefore consist in that order of the manifold of appearance according to which, in conformity with a rule, the apprehension of that which happens follows upon the apprehension of that which precedes. (Kant 1787, B238)

This rule is the rule of causality, according to which reason carries 'the time-order over into the appearances and their existence' (Kant 1787, B245). Kant never explicitly states that events follow each other according to a 'before-after' relationship and thus constitute physical time. Kant is preoccupied with the necessity and universality of knowledge, which cannot be obtained from sensual experience alone. Kant locates the necessity and the objectivity of succession in apprehension in the rule of causality. Still, Kant is not interested in succession as a 'merely subjective play of my fancy' (Kant 1787, B247), but in its objectivity. Hence Kant implicitly acknowledges the need for physical time. Objective change can only be perceived and experienced if the change takes place objectively in the external world. However, his idealistic tendencies prevent him from calling this 'objective succession' simply (physical) time. But there is no escape. An objective idealist view of time must presuppose a succession of events in the physical world in order to secure the regularity and invariance of succession, which Kant requires for his theory of how appearances are turned into objective knowledge. 'The function of causality is to order objectively the sequence of perceptions given by the sensitivity according to a rule' (Kant 1787, B238–B239).

Although the Kantian view was to find support in the Special theory of relativity, it does not imply that time is a human illusion. As we shall see Kant

² Note that Kant identifies causality with determinism, as was customary prior to the arrival of quantum mechanics, see Weinert (2004), Chap. 5.

perceives time as an objective feature of the structure of the human mind, which produces the same succession of events in every human observer.

According to Kant, the outer world causes only the matter of sensation, but our own mental apparatus orders this matter in space and time, and supplies the concepts by means of which we understand experience. (Russell 2000/1946, 680)

When Kant talks about time as a ‘subjective condition of our (human) intuition’, whilst claiming that there is no time in the absence of human observers, he is really talking about the structure of human time (Locke’s duration). There is indeed no human time—no calendars, no dating systems—in the absence of humans. Kant’s emphasis on human time is the result of his Copernican turn (Kant 1787, BXVII), according to which our experience of given objects, in so far as they are knowable, must conform to our concepts. As Kant is a transcendental idealist and empirical realist, he does not really deny the existence of an objective succession of events. This impression is confirmed by his evolutionary cosmology and by a ‘causal theory of time, which he embraces in his second analogy of experience’.³

2.5.2 *Gottfried W. Leibniz (1646–1716)*

The rule of causality guarantees the objective order of events. It was in fact G. W. Leibniz who grasped ‘the importance of the subject of order to the theory of time and space’ (cf. van Fraassen 1970, 35; Jammer 2007). As discussed in [Chap. 3](#), Leibniz is famous for his ‘relational theory of space and time’, which he contrasted with Newton’s view of ‘absolute time’. What does Leibniz mean by ‘order’? Leibniz was a believer in the mechanical worldview, which adopted, amongst other principles, the axiom of the universal causal concatenation of all events.

My earlier state involves a reason for the existence of my later state. And since my prior state, by reason of the connection between all things, involves the prior state of other things as well, it also involves a reason for the later state of these other things and is thus prior to them. Therefore, whatever exists is either simultaneous with other existences or prior or posterior. (Leibniz 1715, 666)

From this viewpoint of determinism, which was shared by many 17th century thinkers, and culminated in Laplace’s famous demon (Weinert 2004, Sect. 5.1), Leibniz derives a causal order of events in the world, and hence a causal order of time.

³ Brittan (1978), Chap. 7 defends the view that Kant holds a causal theory of time, according to which the causal order provides an objective, empirical criterion of temporal order. While this thesis emphasizes that Kant must concede the importance of physical time, it omits certain aspects of Kant’s view: (a) it neglects that Kant explicitly sets out an objective idealist view of time in his *Transcendental Aesthetic* (Kant 1787, B46); (b) it underemphasizes that Kant is preoccupied with an a priori rule of causality and that this rule belongs to the synthetic ‘principles of pure understanding’.

Time is the order of non-contemporaneous things. It is thus the universal order of change in which we ignore the specific kind of changes that have occurred. (Leibniz 1715, 202; italics in original; cf. van Fraassen 1970, 35–44)

Thus Leibniz, too, embraces a causal theory of time. But it is a cause-effect relationship of events in the world, rather than an a priori rule. Leibniz postulates a universal chain of causality amongst events, thus ignoring ‘the specific kinds of changes that have occurred’. With his insistence that time is the order of the succession of events Leibniz proposes a topology of time. In his essay, written after 1714, he claims that this order is a causal order, not of particular events, but of a universal kind. The principle of universal causality is also a feature of classical physics. But in the framework of classical physics, this metaphysical notion of causality is substantiated by the use of deterministic laws, which govern the material events. Newton’s classical mechanics operates with a notion of absolute of time, which Leibniz opposes with his relational view of time. Yet, despite these opposing views, there is the task of specifying the topological notion of the order of succession within the framework of classical physics. Leibniz himself hints at a possible solution when he points out that this order has a magnitude:

(...) there is that which precedes and that which follows, there is distance and interval. (...). Thus although time and space consist in relations, they have their quantity none the less. (Leibniz 1715–1716, Fifth Paper, Sect. 54, 234–235)

Thus if the spatial and temporal relations have quantity or magnitude, they are subject to regularities. For our discussion of the Newtonian and Leibnizian notions of time (Chap. 3), it is of great importance to note that, in accordance with his mechanical worldview, Leibniz holds that nature is subject to laws of nature:

The natural forces of bodies are all subject to *mechanical* laws. (Leibniz 1715–1716, Fifth Paper, Sect. 124, 237–238)

For the regularity, which is associated with ‘mechanical’ laws, gives rise to a metric of time. As will be discussed, Newton makes his notion of absolute time a prerequisite for the formulation of his laws of motion. In the present context, however, it suffices to remember that Leibniz’s view is concerned with the topology of time (it is a ‘before–after’ order of successive events), rather than a metric of time (a quantification of this order).

2.6 The Topology of Time

Any adequate inquiry into ‘time’ is necessarily partly scientific and partly philosophical. (Denbigh 1981, Preface)

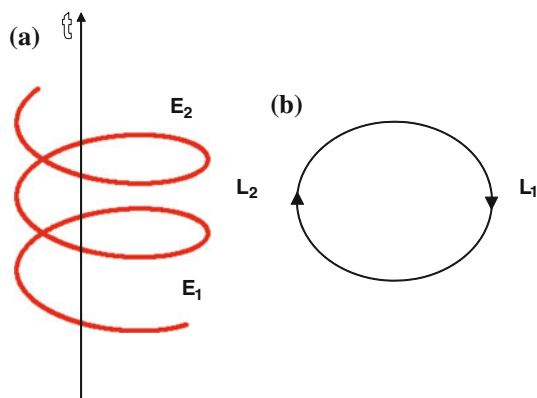
Kant, with his evolutionary cosmic hypothesis, assumes a linear notion of time (an arrow of time). But is a forward-moving arrow of time the only direction conceivable for time? In his view of the evolution of the cosmos from an original

chaos to the observed order of galactic structures through the operation of mechanical laws, Kant underwrites two assumptions: one concerns the *metric* of time—or some device by which the passage of time can be measured and quantified; the other concerns the *topology* of time—or some device by which the shape of time—whether it is linear, circular or cyclic—can be ascertained. As has just been shown, Kant infers from the operation of mechanical laws that the evolution of the cosmos has taken millions of years to reach its present state of order and will take many more millions of years to mature to perfection in other cosmic worlds. This metric is not precise, since Kant did not know that the Earth is approximately 4.5 billion years old and that our universe originated approximately 13.7 billion years ago. Kant's cosmic evolution also assumes a linear topology of time. According to Kant's evolutionary hypothesis the cosmos seems to follow a uni-directional arrow of events from chaos to order. The topology of time concerns the geometric shape of time rather than its quantitative measure. It is possible to have a topology without a metric. Galileo discovered the topology of the moon—its mountains and valleys indicate the geometric shape of the surface of the moon—but he did not calculate the height of the mountains, or the depth of its valleys. Kant and Leibniz had a clear notion of the topology of time but a crude metric. But considerations of the topology of time, even without metric considerations, are important in modern cosmology, where the question of the arrow of time has gained new significance (Chap. 4).

Considerations of the topology of time run through the history of human time reckoning. They may be based on pure speculation or on a theoretical conjecture before evidence favours one or the other topology. An example of speculation occurs in archaic societies, which entertain two notions of time. One is profane time, in which people spend most of their ordinary lives; it is devoid of meaning. People are able to leave profane time and enter mythical time through participation in rituals and ceremonies (Eliade 1954, 34–46). Through the imitation and repetition of archetypes, Man is projected to a mythical epoch, in which the archetypes were first revealed. The Platonic structure of this archaic ontology is noteworthy. Objects or actions only become real if they imitate or repeat an archetype. Objects only acquire meaning and reality if they participate in an archetype. As participation in the archetype involves repetition, archaic and pre-modern societies often entertained a *cyclic* conception of time. They regarded time as a succession of recurring events. Time performs a cycle, as illustrated in the quote by Nemesius, Bishop of Emesa, in the 4th century A.D.:

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)

Fig. 2.3 a A spiral view of time events E_1 , E_2 etc. recur but at a progressively later time. *Source* Wikimedia Commons. **b** A cyclic view of space the same locations, L_1 and L_2 , can be re-visited an indefinite number of times; the spatial sense is truly cyclic



The notion of cyclic time, however, is incoherent from a philosophical point of view, for it presupposes a linear conception of time.⁴ If each state of the universe recurs infinitely many times, then two possibilities need to be distinguished:

- Each individual state recurs repeatedly but separated by a temporal interval, Δt ; hence each state, S , is identical to any other state, S' , (indistinguishable in their properties), except for their location in time. But then these states are distinguishable with respect to their temporal properties. A cyclic or recurring theory of time presupposes a linear progression of time: 'identical states of the universe recur but at a progressively later stage' (Newton-Smith 1980, 66; cf. Fig. 2.3a). This spiral view implies that the same moment cannot be visited twice in time. The spiral view has no spatial analogue since it is possible to visit the same location an indefinite number of times (Fig. 2.3b).
- A truly cyclic theory of time must affirm that all states, S_1 , S_2 , S_3 are identical to states S_1' , S_2' , S_3' , in all, including their temporal properties. But this stipulation is inconsistent: (a) If S occurs before S' , but otherwise both are identical in their properties, then linearity obtains; (b) if $S = S'$ then S and S' are indistinguishable in all their properties, including their temporal occurrence; they are exactly the same states and no recurrence has occurred. Case (b) employs Leibniz's Principle of the Identity of Indiscernibles, which states that if entities A and B have all properties in common, then they are identical: A and B are the same entity.

Even though, then, cyclic time is inconsistent, its impossibility does not exclude other topologies of time, based on theoretical conjectures. If time does not possess a cyclic pattern two geometric possibilities remain, in the absence of any empirical evidence.

⁴ A modern version of cyclic time is Nietzsche's theory of eternal recurrence. On the problematic nature of cyclic time, see van Fraassen (1970), 62ff; Zwart (1976), 244–246; Newton-Smith (1980), 57f; Whitrow (1980), 39f (Sect. 1.9); cf. Capek (1962), 343ff; Lucas (1973), Pt. I, Sect. 9.

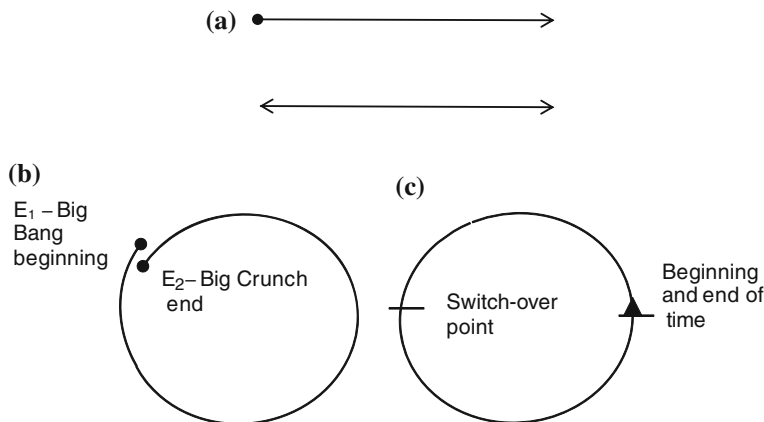


Fig. 2.4 **a** Linear time. In this scenario, in which the topology of the temporal universe is modelled in a linear fashion, the universe may have a beginning (Big Bang) and no end ('heat death') or it may have neither beginning nor end (chaotic inflationary universe). **b** Time as an open circle, in which the physical conditions at the beginning of the temporal universe would not coincide with those at the end, but the universe recollapses after a moment of maximum extension. This scenario, as envisaged by Roger Penrose, means that the physical conditions at the Big Bang would differ significantly from those of the Big Crunch. Penrose calls this scenario the Weyl hypothesis—it has considerable implications for the asymmetry of time. **c** Closed time. In this scenario, in which the topology of the temporal universe is modelled as a closed circle, the Big Bang and the Big Crunch would become physically identical, after the period of recollapse. Does this mean that the arrow of time would flip over at the switch over point, when the universe begins to recontract? Two-time boundary problems are much discussed in modern cosmology (Chap. 4)

- Does time have the topology of an open real line or that of a closed circle? Both scenarios lead to further questions.
- If topological time is to be modelled as an open real line, does time have a beginning but no end? Or neither beginning nor end? (Fig. 2.4a)
- If topological time is to be modelled as a circle, do the beginning and the end of time coincide? Or do beginning and end differ significantly, in which case the end of time would be physically distinct from its beginning (Fig. 2.4b, c).

These topological questions can be raised in the absence of a metric of time and of empirical data, which would permit a clear answer. When cosmologists investigate these issues they will find that the data, if they are reliable, do constrain the answers to these questions. The topological models envisage geometric possibilities of the topology of time and physical investigations will have to make a choice between them. Let us consider the different scenarios.

2.6.1 Linearity of Time

There is a temporal relation, often regarded as the most fundamental relation (Zwart 1976, 34–37), which is ideally suited for the topological structure of temporal

linearity. As Saint Augustine already observed, it is the ‘before–after’ relation between events. There is also a philosophical theory of time—the relational theory—which is suited to this linear topology (van Fraassen 1970, 61). For the relational theory, in Leibniz’s words, time is the order of succession of events. Time unfolds, as events happen one after the other, in a one-dimensional line. But for the relational view, at least according to its classic statement, time has a beginning. It coincides with the creation of the universe and it will last for as long as change happens in the universe. But the order of events cannot be changed: if *A* happens at the same time as *B*, then *A* and *B* happen simultaneously for all observers in the universe; and if *A* happens before *B*, it is logically impossible for *B* to happen before *A*. Whilst this order relation sounds commonsensical, it has been cast into doubt in the Special theory of relativity. If the relational theory is to survive the findings of modern physics, it will have to be reformulated in its language.

The standard topology adopts the order type of linearity, which is represented by a line of real numbers (see Newton-Smith 1980, 51; van Fraassen 1970, 58ff; Lucas 1973). Real numbers, which include both rational and irrational numbers, are needed to express the properties of this order type. The linear view of time ascribes to time the topological properties of linearity (time is one-dimensional), density (a new instant can be inserted between any pair of distinct instants), and either finitude or infinity. It is a mathematical model of linear topology, which may not correspond, in every aspect, to physical time. For instance, if according to the theory of quantum gravity there exists a shortest unit of time—Planck time—then physical time would not be dense in a strict sense of the word; equally if the universe started with a Big Bang, then the universe may expand forever but it would have a definite beginning. There are several cosmological models, which approximate this linear topology.

1. The standard cosmological model is the Big Bang model, which makes the universe spatially homogeneous and isotropic (Weinberg 1978; Penrose 2005, 718; Kragh 2007, Chaps. 4, 5; Carroll 2010). The Big Bang model assumes that the Universe had a definite beginning, and started from a singularity (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 3 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, Sects. 27.7, 27.13). If the universe had a definite beginning, what will be its ultimate fate? Will it expand forever and end in a Heat Death or will it grind to a halt and recollapse to a Big Crunch?

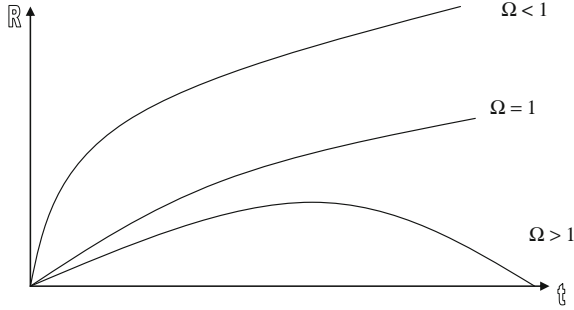


Fig. 2.5 Critical values for Ω and the corresponding evolution of the universe. $\Omega > 1$ indicates a closed universe, which will expand to a maximum extension, and then re-collapse into a Big Crunch. $\Omega = 1$ indicates a flat universe, according to which the universe will expand forever but the rate of expansion will slow down as time progresses, leading to the heat death. $\Omega < 1$ indicates an open universe, according to which the universe will expand at the same rate. As the universe seems to expand at an accelerated pace, the universe will still end up in a heat death, nowadays sometimes called a Big Chill, but faster than in the open universe

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:

$$\Omega = \frac{\rho_o}{\rho_c} = \frac{\text{actual} - \text{mass} - \text{density}}{\text{critical} - \text{mass} - \text{density}} \quad (2.1)$$

The value of $\Omega = < 1, 1 \text{ or } > 1$ and this value indicates the future fate of the universe (Fig. 2.5).

The value $\Omega = 1$ signifies the famous Heat Death, which was predicted by many cosmologists of the 19th century—the total dissipation of energy, such that no energy differentials would be left to perform useful work and sustain life. Is there any evidence in support of the Big Bang theory?

- In 1929 the American astronomer Edwin Hubble (1889–1953) discovered Hubble's law, which states a relation between distance (or redshift) and radial velocity. The farther a galaxy is away, the faster it moves away from the observer: $v = Hd$ (where v = radial velocity of the galaxy, d = distance to galaxy, H = Hubble's constant). Hubble established, on the basis of observations of the redshift of specific absorption lines in the spectra of galaxies that galaxies were moving away from each other at a velocity proportional to their distance. The inference from Hubble's law is that the galaxies must have been closer in the past; and ultimately that the universe must have had a beginning some 13.7 billion years ago, at which point it was in a low entropy state.
- In 1965 A. Penzias and R. Wilson stumbled across the cosmic background radiation, at a uniform temperature of 2.7 °C, which was later understood to be

the afterglow of the hot, young universe, shortly after the Big Bang. The microwave background radiation is a relic of the early stages of the universe. Some 380,000 years after the Big Bang the universe would have cooled down to temperatures, which allowed the formation of atoms like hydrogen and helium. From the abundance of hydrogen in the universe today, it can be inferred that prior to the formation of stable atoms, the early universe must have been filled with an enormous amount of radiation. As the universe expanded this radiation redshifted and cooled to approximately 3 K and this is the microwave radiation, which Penzias and Wilson detected in 1965.

But the Big Bang model provides no indication as to whether the universe will expand forever—leading to a Big Chill (or what the 19th century called the Heat Death)—or will eventually recontract, leading to a Big Crunch. As indicated in Fig. 2.4 an important question from the point of view of the asymmetry of time, to be discussed later, is whether the Big Crunch will resemble the Big Bang. The question of expansion or contraction essentially depends on the amount of dark matter and dark energy in the universe, which cannot be directly observed, but whose 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. A further question, which the standard Big Bang model cannot answer, is how the Big Bang itself was caused and how the special (low-entropy) conditions for space–time inherent in it could have arisen from a singularity (Hawking 1988, 43f; Freedman 1998, 97).

The standard Big Bang model also faced a number of anomalies,⁵ as further work revealed. Three of these problems are the *horizon problem*, the *flatness problem* and the *smoothness problem* (cf. Guth 1997, 183; Atkins 2003, 258ff; Penrose 2005, Sect. 28.3; Kragh 2007, 226; Carroll 2010, Chap. 14). The horizon problem relates to the extraordinary uniformity of the observed universe. This uniformity is most clearly revealed in the cosmic background radiation, which is at a uniform temperature of 2.7 °C in all directions across the sky. Yet the universe is so vast that very distant parts of it cannot communicate with each other because of the finiteness of the speed of light. The cosmic background radiation was released at about 380,000 years after the Big Bang, by which time two opposite points in space were already so widely separated that they could not have communicated through light signals. The only way to solve this problem is to stipulate that the universe began in a state of great uniformity.

The second problem relates to the puzzle that the universe is remarkably close to being spatially flat. The fact that the curvature of space is very close to zero, and hence Euclidean, is referred to as the *flatness problem*. In terms of Ω (the ratio of actual mass density to critical mass density of the universe) this means that $\Omega \cong 1$

⁵ An anomaly is understood here in the Kuhnian sense as a problem, which a theory should be able to solve because it falls within its domain, but which it cannot solve with the principles and techniques it has at its disposal.

today, so that the universe is expected to expand forever but the rate of expansion will tend to zero with time. But this gives Ω a very special value, already present at the beginning of the universe, for which the standard Big Bang model offers no explanation. It also conflicts with recent observations, which suggest that the expansion of the universe proceeds at an accelerated pace.

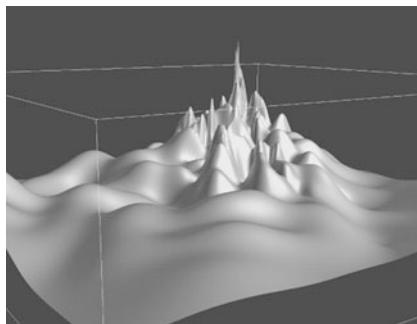
The third problem is the *smoothness problem*, which refers to the even distribution of matter in the universe and the uniformity of space–time geometry. Kant already took note of this uniformity with his nebular hypothesis. Just as Kant’s reliance on Newtonian physics cannot account for this large-scale structure of the universe, so the standard Big Bang model cannot explain the distribution of matter throughout the universe. The standard model implies that the universe exploded so quickly that there was no time for information to be exchanged between its parts (Guth 1997, 213).

2. Cosmologists responded to these problems by developing so-called inflationary models, which can both account for the uniformities and the irregularities (lumpiness) in the distribution of matter-energy in the universe. The cosmic background radiation shows similar irregularities. As J. Gribbon sums up the situation in cosmology:

All these problems would be resolved if something gave the Universe a violent outward push (in effect, acting like antigravity) when it was still about a Planck length in size. Such a small region of space would be too tiny, initially, to contain irregularities, so it would start off homogeneous and isotropic. There would have been plenty of time for signals traveling at the speed of light to have criss-crossed the ridiculously small volume, so there is no horizon problem – both sides of the embryonic universe are ‘aware’ of each other. And spacetime itself gets flattened by the expansion, in much the same way that the wrinkly surface of a prune becomes a smooth, flat surface when the prune is placed in water and wells up. As in the standard Big Bang model, we can still think of the universe as like the skin of an expanding balloon, but now we have to think of it as an absolutely enormous balloon that was hugely inflated during the first split-second of its existence. (Gribbon 1996, 220, italics and bold type have been removed from this quote)

As just indicated, in order to deal with these anomalies cosmologists developed so-called inflationary models, which correspond to different topologies of time. Although there are now a number of inflationary models in circulation, in terms of the topology of time it is convenient to focus on the initial inflationary Big Bang model, due to the American particle physicist Alan Guth. It later gave way to the Chaotic Inflationary model (due to the Russian cosmologist Andrei Linde, as well as the American cosmologists Paul Steinhardt and Andreas Albrecht), which is now the standard version of inflation. Although the first inflationary model was developed by A. Starobinsky in Moscow (1979), the term ‘inflation’ was coined by Guth (in 1981), who developed his own version of an inflationary model. Inflation models deal with the earliest phase of the Universe, inspired by particle physics, and assume a very rapid expansion from a very small region (i.e., the Planck length

Fig. 2.6 Inflationary universe, <http://www.jrank.org/space/pages/2287/cosmological-inflation.html>



10^{-35} m, which is much smaller than the region in the standard cosmology, cf. Guth 1997, 184). The exponential expansion then began at 10^{-35} s after the initial ‘blast’ and lasted until 10^{-32} s. During this time the universe expanded exponentially in size: the universe doubled in size approximately every 10^{-35} s and went through phase transitions. The inflationary period then switched off and the expansion of the Universe settled down to its normal expansion rate, as envisaged in the original Big Bang scenario. In terms of the topology of time, this old inflationary scenario assumed a definite beginning in time but does not try to answer any questions about the ultimate fate of the universe. The old inflationary model also faced problems of its own, for instance to explain why the inflationary periods end suddenly. In order to deal with these problems, Linde proposed a model of ‘chaotic inflation’ (in 1982) (Fig. 2.6). Unlike the Big Bang model the chaotic inflationary model of the universe assumes an eternally existing, self-reproducing inflationary universe. The eternal inflation rolls on like a chain reaction, ‘producing a fractal-like pattern of universes’.

In this scenario the universe as a whole is immortal. Each particular part of the universe may stem from a singularity somewhere in the past, and it may end up in a singularity somewhere in the future. There is, however, no end for the evolution of the entire universe. (Linde 1998, 103–104; cf. Hawking 1988, 115–133; Weinberg 1978; Barrow and Tipler 1986, Sect. 6.2; Penrose 2005, Sect. 27.7; Carroll 2010, Chap. 14)

In this particular scenario the universe as a whole is without beginning and without end, thus avoiding the conundrum of what happened before the Big Bang and which forces brought it about. Each particular part of the universe, each galaxy, may originate from an inflationary bubble in the past, similar to the Big Bang, and end in a Big Crunch in the future. But the evolution of the entire cosmos is without end, as Kant already envisaged. The self-reproducing cosmos appears as an extended branching of inflationary bubbles. This process is called ‘chaotic inflation’ because the present state of the universe could have arisen from a large number of different initial configurations. But not all initial configurations would have given rise to the type of universe observed today. The universe expanded rapidly in places, represented by peaks in a model with valleys and hills.

In such a model, our galaxy would exist in a valley, where space is no longer expanding. This picture gives rise to the notion of ‘multiverse’:

The evolution of inflationary theory has given rise to a completely new cosmological paradigm, which differs considerably from the old big bang theory and even from the first versions of the inflationary scenario. In it the universe appears to be both chaotic and homogeneous, expanding and stationary. Our cosmic home grows, fluctuates and eternally reproduces itself in all possible forms, as if adjusting itself for all possible types of life. (Linde 1998, 104)

As the chaotic inflationary model indicates, a problem in modern cosmology is the very existence of the Big Bang, because it requires the postulation of a low-entropy initial condition. It is commonly assumed that the cosmic arrow of time requires the assumption of a Past Hypothesis, which is the postulated low-entropy state of the initial conditions of the universe. The problem is not the empirical observation that our universe must have started in a low-entropy Big Bang but why it started in such a very special smooth condition. Much of modern cosmology is concerned with developing models, which may provide a tentative explanation of the very special conditions of the Big Bang: baby universes, cyclic universes, bouncing universes, oscillating universes. In many of these models it is no longer assumed that the observable universe started in a Big Bang singularity (see Gott 2001; Penrose 2010; Carroll 2010).

This new cosmology will have interesting consequences for the completion of the Copernican Revolution. The Copernican heliocentric worldview is supposed to have delivered a first blow to human arrogance by having displaced the Earth from the centre of the universe. The multiverse, then, appears as

the final humiliation, this one of hypercosmic proportions: this universe may be but one of countless others. Not only are we the inhabitants of an insignificant (but wonderful) planet near an insignificant (but wonderful) star at an insignificant location in an insignificant (but wonderful) galaxy in an insignificant location in the visible (and wonderful) universe, but that universe is perhaps an insignificant universe among innumerable others, a ‘multiverse’, each of them infinite in extent. (Atkins 2003, 265)

Whilst the Copernican view certainly put an end to any hope of physical centrality, it reaffirmed the humanistic focus on rational centrality. Contrary to Atkins’s negative assessment, apparently inspired by the multiverse idea, modern theories of cosmology continue to demonstrate humanity’s ability to rationally comprehend the surrounding cosmos, irrespective of our humble place in this ‘wonderful’ universe (Weinert 2009, Sect. 6.1). Thinking about the topology of time is one of the outcomes of this exercise in rational centrality.

Despite the considerable progress of modern cosmology, the available empirical data do not (yet) definitely constrain the topology of time to one particular model. Modern cosmology entertains both symmetric and asymmetric models of the universe (Chap. 4). Hence it is still possible for the topology of time to be a circle, growing from a Big Bang to a maximum expansion after which it

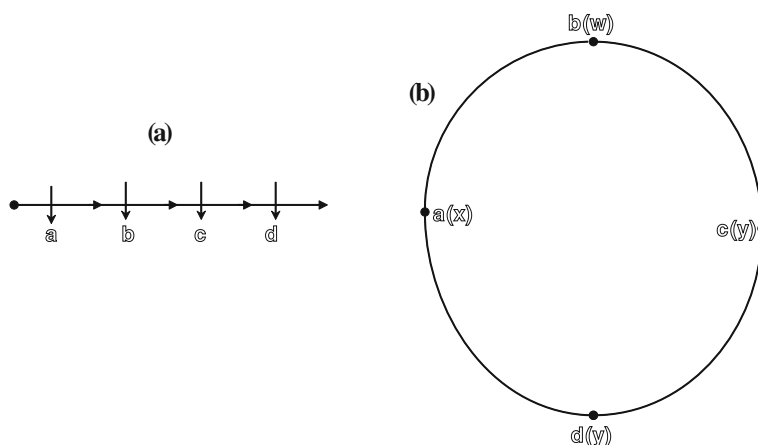


Fig. 2.7 a Linear time. b Closed time

recontracts to a Big Crunch.⁶ Such a scenario would have considerable implications for the arrow of time.

2.6.2 Closed Time

The idea of *closed* time is very different from the previously dismissed view of *cyclic* time—the idea that it is possible for the same events to occur repeatedly an infinite number of times; that it is possible to revisit past events in the same way as places, for instance by walking around a circular arena. Cyclic time is incoherent because each time, each temporal moment, on Leibniz's Principle, occurs only once and cannot reoccur at a different time. However, closed time allows each temporal moment to occur just once but temporal items are related to each other like points on a circle. To describe a circle the irrational number, π ($C = 2 \pi r$), is needed. Hence the order type for circular time must be the real numbers. But the 'before–after' relation, which served well for linear time, is no longer suited for a characterization of closed time.

The 'before–after' relation in linear time is both asymmetric and transitive: It is *asymmetric*, because if 'a' is before 'b', then 'b' is after 'a'; it is *transitive* because if 'a' is before 'b' and 'b' is before 'c', then 'a' is before 'c' (Fig. 2.7a). These characteristics correspond well to our daily experience: If the Egyptian pyramids

⁶ Cf. Newton-Smith (1980), 57ff; Davies (1974), Sect. 7.5; van Fraassen (1970), 62ff. According to certain solutions to Einstein's field equations, found by Gödel, there exist closed time-like curves (CTCs), which would carry an observer, whose existence lasted long enough, back to the beginning of the curve.

were built before the publication of Darwin's *Origin of Species* (1859), then Darwin's book appeared after the construction of the Egyptian pyramids. No metric is required to affirm this asymmetry. Furthermore, if the Egyptians built their pyramids before the publication of the *Origin of Species*, and the Eiffel Tower in Paris was constructed after the publication of the *Origin of Species*, then the Egyptians built their pyramids before the Parisians erected the Eiffel Tower (1889). Again no metric is needed to affirm this transitivity. But the 'before-after' relation is unsuitable for a characterisation of closed time for both these conditions are violated on a closed circle. Without specifying a direction, or assigning coordinate values to the positions, it is found that if 'a' is before 'b', 'b' is also before 'a' and 'a' is before itself; hence this relation is symmetric. Also 'd' lies between 'a' and 'c' (if you go from 'a' to 'c' through 'd') but also (in the opposite direction) 'a' lies between 'd' and 'c'. The relation is reflexive (Fig. 2.7b). However, there is a way to characterize order in closed time by the basic temporal relation of pair separation. Write $S(x, y/z, w)$ for 'the pair of instants, x and y, separate the pair of instants z and w'. That is, there is a route from z to w, which passes through x but not y or a route between z and w, which passes through y but not x' (Newton-Smith 1980, 59; cf. van Fraassen 1970, Sect. III.1; Kroes 1985, 13). This picture singles out a direction around the circle. What this purely formal exercise shows is that specifying a route on a closed circle requires specifying a direction around the circle. But specifying a route amounts to imposing asymmetry on closed time: to reach 'w' from 'z' we are either allowed to travel through 'x' or through 'y'. But how is this asymmetry to be justified? Why do we not assume symmetry? These questions must ultimately be decided on empirical grounds. The closed time scenario represents the universe as returning to its initial state, as in some cosmological models, $\Omega > 1$.

What significance do these logical considerations regarding the topology of time have for cosmology? They suggest possible scenarios for the ultimate fate of the universe. As has just been shown, inflationary cosmology locates the beginning of a bubble universe in the quantum conditions of the very early phase of the universe (10^{-33} cm). These early phases of the universe cannot be described by the General theory of relativity. They fall within the domain of quantum cosmology. As will be discussed later, in quantum cosmology a very basic distinction exists between time-symmetric and time-asymmetric universes (Chap. 4). Time-asymmetric models of the universe assume that the initial and final conditions of the evolution of the universe differ physically, which bestows an 'arrow' of time on cosmic evolution. Proponents of time-symmetric models of the universe postulate a symmetry between the initial and the final conditions of the universe. The universe finds itself in similar states at the beginning and the end of its evolution. If these suggestions are taken seriously, they will have important implications for the question of the arrow of time. If both the initial and final conditions of the universe are assumed to exist in particular low-entropy states—on account of the symmetry principle—then the midway point, the point of greatest expansion with higher entropy, may be explained from both temporal endpoints. In terms of the model of closed time this assumption means that the midway point of greatest expansion can be explained

without preference for a temporal direction. Traditionally, it is postulated that the universe starts in a low-entropy state—the Big Bang. (For present purposes entropy is understood as a measure for the degree of order in a system.) With the expansion of the universe, the entropy increases, in accordance with the Second law of thermodynamics, giving the universe a temporal ‘direction’. But proponents of time-symmetric models object that asymmetric models presuppose an unjustified preference of initial over final conditions. According to this symmetric standpoint, both the initial and final conditions of the universe are equal, or approximately equal, with the result that the thermodynamic arrow tracks the cosmological arrow. The symmetric standpoint does not endorse an asymmetric arrow of time as being ‘typical’ and the time-reversed view as being ‘atypical’.

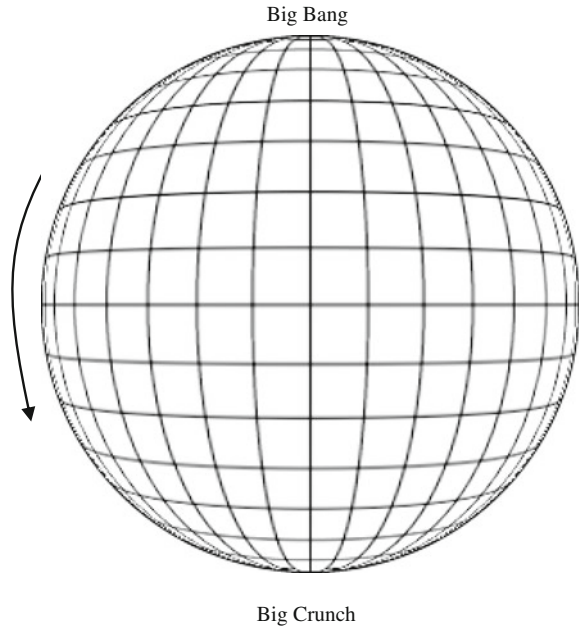
The issues of asymmetry and symmetry for the notion of time will be discussed in greater detail in [Chap. 4](#). For the question of the circular topology of time, and its relation to cosmology, Hawking’s ‘no-boundary proposal’ may be considered. It is a variation of the chaotic inflationary models. One of the problems of the standard cosmology was that the universe began (and possibly ends) in a singularity, a point of infinite density, infinite pressure and infinite temperature, at which the equations of the General theory of relativity break down. Cosmologists therefore realized that at these extreme points quantum mechanics had to be taken into consideration. Hawking made the proposal that space–time could be finite but without boundaries, if it was described in imaginary time.⁷ Space–time would be like the surface of the Earth—finite in extent, without boundaries, and time is measured in imaginary units. The introduction of imaginary time serves the purpose of avoiding singularities, which plagued traditional cosmologies. Hawking made this proposal in order to account for the problems in the standard Big Bang scenarios. For Hawking no real distinction exists between ‘real’ and ‘imaginary’ time. He recommends, in line with his instrumentalist philosophy, that we should adopt a notion of time, which leads to the best description of physical reality. In imaginary time the universe would have zero size at the beginning and the end, and no singularities would occur ([Fig. 2.8](#)). But in real time things would look very different.

At about 10 or 20 million years ago, it would have a minimum size, which was equal to the maximum radius of the history in imaginary time. At later real times, the universe would expand like the chaotic inflationary model proposed by Linde (but one would not now have to assume that the universe was created somehow in the right sort of state). The universe would expand to a very large size and eventually it would collapse again into what looks like a singularity in real time. (Hawking 1988, 138)

Under the no-boundary proposal the universe would be in a quantum state at the beginning and the end of time. The singularities of General Theory of Relativity

⁷ Note that Hawking (1988) sees scientific theories, in an instrumentalist fashion, as mere calculating devices; they propose mathematical models to describe our observations and make predictions but it is not assumed that the structure of the models reflect, in any way, the structure of reality.

Fig. 2.8 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. Nevertheless, this cosmological model is asymmetric since the beginning is characterized by smooth conditions, whilst at the end the universe collapses into black holes (Hawking 1988, 138; see Penrose 2005, Sect. 25.8)



would be avoided: the universe would not start from a lawless Big Bang and crush into a lawless Big Crunch.

In that case, the beginning of time would be a regular, smooth point of space–time and the universe would have begun its expansion in a very smooth and ordered state. (Hawking 1988, 148)

Hence Hawking's cosmological model does not adopt a time-symmetric evolution of the universe since beginning and end are not physically identical. The Big Bang starts in a smooth condition but the Big 'Crunch' corresponds to a collapse into black holes. The universe begins in a Big Bang, grows to maximum expansion, and then collapses into a Big 'Crunch', in real time. The black holes themselves will eventually evaporate, suggesting a Big Chill scenario.

Closed time gives rise to important investigations in cosmology: (a) If closed time is adopted as a topology for the history of the universe, the question arises whether the initial and the final conditions of the universe are indeed approximately identical, thus leading to a 'nice symmetry between the expanding and contracting phases' (Hawking 1988, 150) or whether the two phases are physically different, leading to a linear topology (Fig. 2.4a) or an open circle (Fig. 2.4b). On the topology of closed time a further question arises, namely what happens at the moment of reversal? This is known as the 'switch-over' problem (Fig. 2.4c). Does the arrow of time flip if the universe recontracts and returns to the original state? (b) According to many cosmologists even in a recontracting universe our temporal experience would not begin to run in reverse. But what importance does the

Second law of thermodynamics (increase in entropy) have for the arrow of time? (c) We experience the passage of time as a forward moving, irreversible march of events but are there any physical phenomena, which suggest the ‘flow’ of time? Or is our asymmetric experience of the passage of time an ‘illusion’, which is not grounded in physical phenomena? (cf. Savitt 2002; Norton 2010)

Kant referred to the fundamental laws of classical physics to explain the ordered evolution of the universe. But the fundamental laws of physics are time-symmetric (or *t*-invariant), whilst human experience of the world is characterized by the anisotropy of time. *T*-invariance can be characterized in various ways. An ideal pendulum may be considered, which suffers no friction and swings back and forth with constant amplitude. If a film were shown of this pendulum the observer would not be able to tell whether the film was running forward or backward. The *t*-invariance of the basic equations of physics indicates that the parameter *t* may take on both positive (+) and negative (−) values, so that the solutions to the equations are regarded as physically possible states of a system. Time-symmetric laws do not forbid processes which run counter to what is regarded as the normal progression of physical events. Another way of characterizing *t*-invariance is to say that ‘elementary processes, such as collisions, quantum transitions etc., have precisely the same frequency of occurrence whether they occur from a micro-state 1 to a micro-state 2 or from a ‘time-inverted’ state of 2 to the time-inverted state of 1 in the same temporal interval’ (Denbigh and Denbigh 1985, 49). Hence there exists a fundamental discrepancy between the atemporal fundamental equations of science and our asymmetric experience of a temporal world. We possess records of the past but not of the future; we are unable to change the past but we can influence the future; for many systems a universal tendency towards disorder and decay seems to be at work, if no effort is made to keep order.

But then how is this asymmetric experience of temporal phenomena to be explained, if the fundamental laws are time-reversal invariant? Is our asymmetric experience ultimately reducible to our mental states or is it supported by asymmetries in the physical world?

2.7 The Metric of Time

The way we think about time is conditioned by the way we measure it... (Lucas 1973, 308)

When Saint Augustine speaks of the ‘before–after’ relation in the ‘succession of events’, he does not specify which events he has in mind and whether the succession needs to follow a particular order. Leibniz characterizes time as the ‘order of the succession of events’ in the physical world but he does not specify either which events he has in mind. Are they actual events or possible events? Although both Kant and Leibniz were more concerned with the topology of time—in each case they assumed a linear topology—their insistence that the whole of material nature is governed by ‘mechanical laws’ immediately gives rise to the possibility

of a metric of time. The Leibnizian notion of ‘order’ has particular significance in both classical and modern space–time physics, precisely because such space–times are governed by laws, even though they may not be mechanical laws. The notion of order also signifies the difference between the mere *passage* of time and the *measurement* of time. Saint Augustine and later Leibniz claim that our awareness of time presupposes a ‘before–after’ relationship between succeeding events; these are events with which we are familiar from the physical environment around us. Events can follow each other in a regular or an irregular fashion. Does this fact have any effect on the notion of time? Consider the movement of street crowds past a certain place—perhaps a monument—in a major city. Very early in the morning there are few people and they pass by without any discernible regularity or any periodic pattern, which would allow us to establish some regularity in the movement of the crowd. As the day passes and the crowds swell the flow of people increases. But the flow will move in any direction and at any pace. Some people hurry past the monument, others amble along, yet others stop to look at the monument. This motion of the crowd is reminiscent of an irregular bus service. On the way to work you expect a bus, and in fact a number of buses, but without any idea of their arrival. The flow of buses may simply be out of sync with the timetable. Yet you would be able to tell that time passes. We are all familiar with this annoying phenomenon. You are waiting for, say, a 433 bus to take you to work but although there is no 433 for a long time several 431 s call at your stop. In order to gauge the passage of time all you need to do is count the number of buses, which pass this particular stop. Or you could count the number of cars and the number of people on the pavement. Hence you may be aware of the passage of time without being able to measure it. Almost any occurring process in nature could raise our awareness of the passage of time. *Prima facie* any quasi-periodic process in nature can serve as a basis for the appreciation of the passage of time. For instance, the expansion of an ensemble of gas molecules into an empty container or the chaotic Brownian motion of molecules in a liquid are sufficient to observe some change, which may be associated with the passage of time. Heartbeats, pulses, lunar cycles, the dance of the seasons or the flooding of the Nile in ancient Egypt may serve as a yardstick for the identification of physical time. Arguably mere change in some physical process is sufficient for the observation of the passage of time by a cognizant observer.

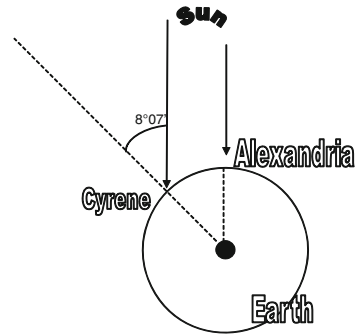
Saint Augustine agreed with Aristotle that the awareness of time depended on a ‘before–after’-relation of events, which undergo physical changes. But Saint Augustine made the human mind the metric of time. He discovered ‘psychological time’. But psychological time does not give rise to a reliable, standardized metric of time. To repeat, the problem is twofold: (A) Psychological time depends on psychological states and these can change within one individual. We judge the passage of time depending on our moods. As our moods change, so does our impression of the passage of time. Hence there is no guarantee that the mind will provide a regular metric. The mind sets its own standard and lacks an external referent. Hence the individual is unable to compare the different impressions of intervals. As Saint Augustine said, we are only aware of the fleeting present moment. But how do we compare the current moment with the preceding one? In

order to *measure* the passage of time we need some regularity. By ‘regularity’ is meant here a periodic process of finite length, which can be subdivided into equal intervals, where the equality of the intervals is measurable by some other measure, for instance spatial distances. If that procedure is inconvenient, a linear process, like the propagation of light signals between two points, may serve a similar purpose. (B) But let it be assumed, for the sake of argument, that the mind, as a metric of time, provides sufficient regularity to measure the passage of time. It has been argued that the passage of time may be related to streams of consciousness: ‘We are aware of the passage of time even when we are not observing any changes in the external world’ (Lucas 1973, Pt I., Sect. 2, cf. Pt. IV, Sect. 56; cf. Shoemaker 1969). But recall that even the mind, in the view of the British empiricists, presupposes physical time. Consider, then, two individuals in a psychological experiment who are asked to ‘measure’ the interval between two events by the use of their psychological clocks. Their impressions will differ, depending on their psychological states. Let us idealize and stipulate that each individual’s personal time is rather regular. If asked repeatedly individual A will say, on several occasions, that the interval lasted, say, 2 s and individual B will say that the interval lasted, say, 3 s. From their respective perspectives the assessment will be regular, under this idealization; but as we switch perspectives between A and B the assessment changes. In other words, the ‘measurement’ is not invariant, since A and B disagree about the length of the interval. Hence, mere regularity is not sufficient for the objective measurement of the passage of time.

For a community of observers to have a reliable clock, the clock must be both regular—it must measure the same interval for the same events on successive occasions—and it must be invariant—it must measure the same interval for different events of the same duration. Invariance expresses the fact that under certain specifiable transformations the value of specific parameters remains unchanged; i.e. as perspectival viewpoints change (from A to B) certain parameters remain invariant. Typically such parameters remain invariant with respect to spatial and temporal changes. For instance, the physics of tennis remains invariant with respect to changes in spatial and temporal location. A game of tennis remains the same, from a physical point of view, whether it is played in Andalusia or Zimbabwe. Neither of these conditions is satisfied by Saint Augustine’s psychological time. There is therefore a need for physical time, which must be based on some periodic external event. Although Plato and Aristotle disagreed about the definition of time they agreed that the measurement of the passage of time requires the satisfaction of these conditions. For the Greeks associated the measurement of the passage of time with cosmic regularity.

To see the need for regularity and invariance for the metric of time, consider the following thought experiment, involving two famous Greeks, Archimedes (287–212 B.C.) and his contemporary Eratosthenes (284–202 B.C.) positioned at two different locations. Archimedes is famous for his discovery of the specific weight of metals and other achievements, whilst Eratosthenes is known for his determination of the circumference of the Earth. Let Archimedes be in Alexandria (Egypt) and Eratosthenes in his birthplace Cyrene (Shahhat, Libya). Today, these

Fig. 2.9 Midday in different locations. When it is midday in Alexandria, in solar terms, it is not yet midday in Cyrene (diagram adapted from Fäbller and Jönsson 2005, 14)



two locations are in neighbouring time zones, each time zone being separated by 15° longitude or 1 h. Their locations are approximately 590 km apart. Both were schooled in Greek astronomy and cosmology and thus believed in the geocentric worldview. Let us imagine that both arrange to hold lunchtime meetings at their respective locations. But how do they determine lunchtime at their respective places? Unlike Saint Augustine, they will not rely on the metric of the mind. They can use sundials as clocks, which will determine for them the highest point of the sun, which signals midday in their respective locations. When the sun is at its highest point in the sky in Alexandria, a sundial in Cyrene, to the west of Alexandria, will throw a shadow, with respect to the vertical, of around $8^{\circ}07'$ (Fig. 2.9). When it is midday in Alexandria, it will not yet be midday in Cyrene. Archimedes will hold his lunchtime meeting earlier, by a fraction of the rotational period of the Earth on its own axis. Do we face the same problem of perspectivalism with respect to the sundials as with Saint Augustine's mental metric? As Archimedes and Eratosthenes 'disagree' about when it is lunchtime, do we encounter the same relativity in the determination of the passage of time, as Saint Augustine's imaginary observers? Not, if the two conditions—regularity and invariance—are fulfilled. Regularity is satisfied because the sun will return to the same location in the sky after approximately 365 days, both in Alexandria and Cyrene. Every observer in Alexandria and Cyrene will experience the same regularity—they will all agree, in their respective locations, that it is midday, as the sundial tells them so. Invariance is also satisfied. Invariance requires that some parameters will not change, even though our perspectives do. If Archimedes were to travel to Cyrene, and thus change his location, he would observe the same regularity about the sun's apparent motion across the sky as in his home town. And if Eratosthenes were to visit Archimedes in Alexandria, he would agree with Archimedes on the occurrence of lunchtime in this location, as a result of the solar regularity. But Archimedes and Eratosthenes have no need to travel. They can use their knowledge of astronomy to calculate the moment of midday in their respective locations. Archimedes can calculate when it is midday in Cyrene, and Eratosthenes can

calculate when it is midday in Alexandria. Their calculations will be invariant with respect to their perspectival positions, since their respective calculations will return the same values. In fact, they will be able to predict when it is lunchtime in their respective towns. For the measurement of the passage of time, then, regularity and invariance are essential conditions.

Strictly speaking, Archimedes and Eratosthenes do not measure time, or its passage. They measure events, periodic events, like the apparent passage of the sun across the sky, from which they derive human notions of time. Based on the periodic nature of cosmic events, like the regular orbit of the planets, they may speak of the passage or the ‘flow’ of time. But the ‘flow’ of time is only a metaphorical expression to capture the ‘before–after’-relationship between events in the physical world. From the ‘flow’ of time—the passage of time—they may be tempted to infer an arrow of time. But the passage of time is not to be conflated with the arrow of time. The arrow of time refers to global properties of the universe, whilst the passage of time is based on the experience of events in their immediate vicinity. Even though local observers in a particular corner of the universe experience an overwhelming forward thrust of events—organisms are born, grow and die—they cannot infer from this experience that the global march of time moves unilinearly, like an arrow, in a forward direction. As can be gleaned from the consideration of the topology of time, our human experience of the ‘flow’ of time is compatible with both linear and closed topologies of time.

It would not even have occurred to the Greeks that there may exist an arrow of time. Aristotle (384–322 B.C.) developed his two-sphere cosmology, which placed the Earth at the centre of a closed universe and made planets circle the centre (Sect. 2.1). The universe was divided into the sublunary and the superlunary spheres. The planets moved in eternal circles, beyond the sublunary sphere of decay but below the ‘fixed’ stars whose canopy provided the boundary and closure of the universe. Even though Ptolemy (100–75 A.D.) would later refine this cosmology into a mathematical theory of geocentrism, the Greeks had no notion of an expanding or contracting universe. The decay in the sublunary sphere may have suggested to them the ‘flow’ of time but the eternal perfection of the supralunary sphere could not give rise to the notion of an arrow of time. In fact, the notion of the arrow of time requires insights, which physics only developed in the 19th and 20th centuries.

The focus of this chapter is on aspects of the measurement of time. A consideration of the measurement of time, in an objective sense, requires us to go beyond psychological and mental notions. Of physical time it seems to be true to say that ‘the way we think about time is conditioned by the way we measure it’ (Lucas 1973, 308). As has been argued, regularity and invariance are essential features of the measurement of time, whose importance we should therefore consider in classical and modern physics.

Later chapters will investigate whether the measurement of time will allow us to infer philosophical consequences about the temporality of the physical world.

2.8 Some Advances in the Theory of Time in Classical Physics

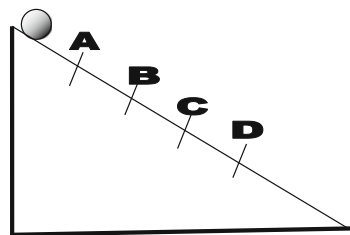
If only Newton had set his absolute space as the space of the “fixed” stars, there would have been no Leibniz-Clarke controversy. (Erichson 1967, 95)

Classical physics made several important contributions to the theory of time, each of which had significant philosophical impact. Galileo worked with the notion of *physical* time, by which he basically meant clock time, as used in his mechanical experiments. Newton introduced the notion of *mathematical* time, which he used to define his laws of motion. And Boltzmann arrived at the notion of the *arrow* of time. The notion of time in classical and modern physics is invariably the notion of measurable time.

2.8.1 Galileo’s Physical Time

In Galileo’s famous experiments with inclined planes a new notion of time emerges in physics. It is the notion of *physical* time, which presupposes that time is measurable. Although Galileo was a supporter of the Copernican heliocentric model of the universe, his notion of physical time is not derived from cosmological considerations. Galileo needed a concrete measure of time for his fall experiments. Galileo’s notion of physical time is an early version of clock time. Clock time indicates a scale of continuous, linear, measurable and abstract units by which the duration of physical events can be measured (Elias 1992, Sects. 21–23; Wendorff 1985, 205–206; Burt 1980, 91ff). As Galileo’s experiments demonstrate, clock time does not necessarily mean the availability of mechanical clocks. Galileo in fact employed water clocks; and Einstein later used light clocks. Einstein showed that the classic assumption that clocks behave invariantly in all inertial reference frames was mistaken. Galileo was not concerned with the Aristotelian question of ‘why’ bodies moved the way they did on the surface of the Earth. His objective was to describe ‘how’ bodies moved and how this motion could be described mathematically. Although the mechanical clock had been invented at the end of the 13th century, Galileo could not use mechanical clocks to measure the motion of falling bodies, because the time pieces at the beginning of the 17th century were not precise enough to measure the short intervals of falling bodies. What is essential is a regular and invariant process by which the duration of an event can be measured. How did Galileo measure the elapsed times in his experiments? He reports that he used a form of water clock—a bucket with a hole in its bottom, from which a thin steady stream of water flowed. The water was collected, in each experiment, in a cup under the bucket. Galileo weighed the amount of water, which was collected when a ball rolled down the incline from A to B, from B to C, and from C to D (Fig. 2.10). The intervals A–B, B–C, C–D were equally long but the amount of water collected did not remain

Fig. 2.10 Galileo's fall experiment



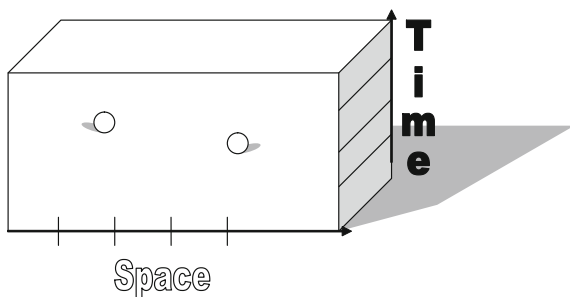
the same. In fact, he collected less water for the interval C–D than for the interval A–B, because of the acceleration of the balls down the plane. From the differences in the weights Galileo obtained the proportions of the weights and calculated the proportions of the times. In order to obtain measurements for the fall times of bodies, Galileo experimented with balls rolling down *inclined* planes.

As a good scientist he carried out the experiments repeatedly on different planes with different inclinations. But in the many repetitions of the experiments, he found no time differences, ‘not even $\frac{1}{10}$ of a pulse beat’ in the measured fall times of the balls in the respective planes. Galileo’s result was that the lengths of the planes were related to the squares of the times: $L \sim t^2$. Galileo concludes that

in such experiments, repeated a full hundred times, we always found that the spaces traversed were to each other as the squares of the times, and this was true for all inclinations of the plane, i.e., of the channel, along which we rolled the ball. (Galilei 1954, 179; cf. Galilei 1953)

In modern terms, Galileo found that $L = v_0 t + \frac{1}{2} a t^2$, where the parameter t indicates clock time. Note that t can take on any numerical value, even if a mechanical clock cannot measure it. Imagine that t has the value of the square root of $\pi - \sqrt{\pi}$; this value can be plugged into the equation to obtain a hypothetical length for L , even though such an experiment may not be realizable in a laboratory, certainly not in Galileo’s lifetime. Galileo’s notion of clock time, as realized in his water clocks, satisfies both the criteria of regularity and invariance. First, the water flows regularly, if the same amount of water is taken each time and the same type of bucket is used. Secondly, as Galileo reports himself, no time differences were found in the many repetitions of the experiments. Galileo’s fall experiments were thus at least invariant with respect to changes in time; but they would also have been invariant with respect to changes in locality. The regularity of the water flow is insufficient by modern standards and it is of a linear rather than a periodic type. Nevertheless the water flowed sufficiently regularly from the bottom of the bucket to constitute a clock for the purpose of the experiment.

Fig. 2.11 Newton's absolute space and time



2.8.2 Newton's Mathematical Time

While in Galileo's experiments the symbol t expresses clock time—or the time interval, which elapses between two events, like the beginning and the end of an inclined-plane experiment—it remains one of the physical parameters, which enter into relations like $L \sim t^2$. In Galileo's work time has no further philosophical significance. But Newton incorporates his notion of absolute time in his theory of motion in a systematic way. In Newton's theory the notion of time therefore takes on philosophical significance. Newton defines his notion of time, to be used in his theory of motion, before he develops his equations of motion. For Newton mathematical time included two aspects: (a) At first Newton considers, like the Ancients, that time makes a reference to observable physical events, like the orbit of the moons around Jupiter. Although he is a child of the Copernican age he does not believe that the planets orbit the sun with perfect regularity. In fact Newton goes further and conjectures, correctly, that even though the motions of such bodies around their respective centres satisfied the requirement of periodic regularity, this periodicity suffered nevertheless from too many irregularities for the formulation of the laws of motions. For instance, the speed of the Earth's rotation varies due to tidal frictions, due to seasonal changes and due to a small difference between the geometric and the physical (rotational) axis of the Earth; (b) Newton therefore postulates an absolute space and an absolute time, which in theory make no reference to material events in the universe. Newton defines 'immovable' space as a space with no relation to any external events, objects or processes; equally for time. These definitions stand in stark contrast to, say, the views of Saint Augustine, which make time dependent on physical events.

By a suitable analogy, absolute space can be envisaged as a cosmic container, which exists irrespective of whether it contains material objects, like planets and galaxies. And absolute time can be imagined as a constantly flowing metaphorical river whose regularity constitutes the basis of a clock. All motions in the material world are measured against the constant tick of the clock. One can consider Newtonian space as an empty container, on whose walls are etched a spatial and a temporal axis (Fig. 2.11). If objects are placed inside the container, their locations and motions can then be traced against the fixed spatial and temporal axes. To illustrate, image that this container is populated with two 'intelligent' molecules,

which spend their days, like fish in a tank, swimming around aimlessly. To make their lives more interesting, the molecules decide, on their next encounter, to determine their respective positions inside the container. Molecule₁ asks molecule₂, 'where are you?' but the answer 'near you' is hardly enlightening. Such relative determinations are useless because the molecules change their positions constantly and supply no absolute determination of their location. But the two molecules can determine their positions inside the container if they can refer their respective locations to the same temporal and spatial axes. (The time axis may not consist of mechanical clocks, simple ticking devices would suffice.) The laws of motion are formulated with respect to these absolute notions. As absolute space and time cannot be observed, these notions provoked considerable debate. What function these notions played in Newton's mechanics will be considered later.

Whatever the philosophical merits of these notions, Newton emphasizes both the importance of regularity and invariance in his absolute notions. Absolute, immaterial space and the 'equable flow' of time are regular with respect to all inertial and accelerated observers (like the 'intelligent' molecules). Absolute space and time are also invariant, since the units of absolute space and time do not change as the molecules change their positions or under a switch of perspectives from one molecule to the other. Newton took the need for regularity and invariance to an extreme when he postulated absolute space and time to ground his laws of motion. If absolute space and time could be observed, no observer would have difficulties in measuring all relative motions against its invariant structure. By analogy, the speed of an aeroplane is measured with reference to the surface of the Earth rather than drifting clouds in the sky. Newton even invented some thought experiments, as we shall see, in an effort to demonstrate that the notions of absolute space and time could be inferred from observations of relative motions.

For present purposes it is important to note, especially in relation to Leibniz, that Newton introduces approximations to 'materialize' his absolute notions. His absolute notions are idealizations from observable physical systems. Thus Newton suggests that 'Jupiter's satellites' may serve as an approximation to absolute time, no doubt because of the perceived regularity of their orbits. If Newton lived today he would certainly change his approximations of absolute time to the recent discovery of neutron stars (1967), which rotate very fast and emit very regular pulses of radiation. They have a much greater regularity than planetary motions. As far as absolute space is concerned, Newton regarded the 'fixed' stars as a good approximation to the image of 'immoveable' space because the constellations were regarded as 'fixed' in the sky; their apparent rotation through the night sky was due to the daily rotation of the Earth.

But because the parts of space cannot be seen, or distinguished from one another by our senses, therefore in their stead we use sensible measures of them. For from the positions and distances of things from any body, considered as immovable, we define all places; and then with respect to such places, we estimate all motions, considering bodies as transferred from some of those places into others. And so, instead of absolute places and motions, we use relative ones; and that without any inconvenience in common affairs; but in philosophical disquisitions, we ought to abstract from our senses, and consider things

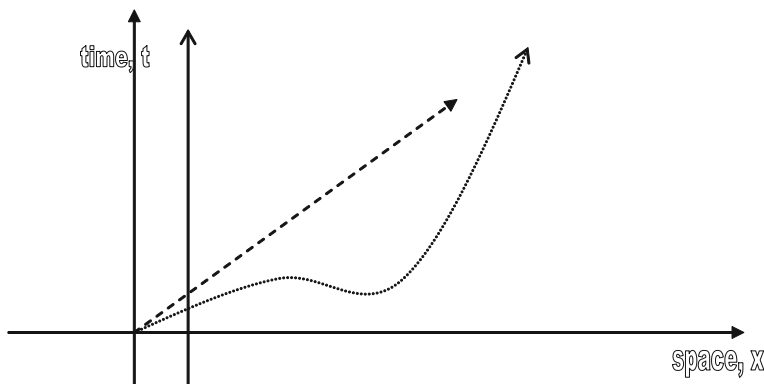


Fig. 2.12 Newtonian space and time axes. Representation of a frame at rest (*vertical line*), an inertially moving frame (*inclined line*), and an accelerated frame (*curved line*) in a Newtonian space and time diagram

themselves, distinct from what are only sensible measures of them. For it may be that there is no body really at rest, to which the places and motions of others may be referred. (Newton 1687, Scholium, 8; cf. Rovelli 2009, 3)

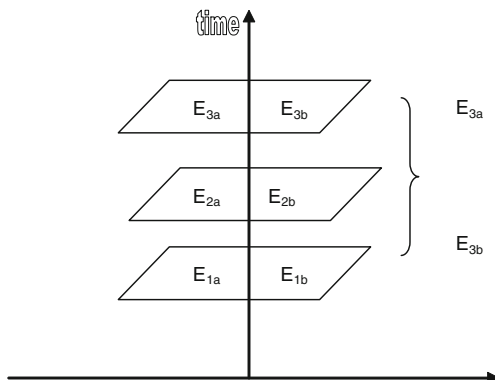
Newton conceives of space and time as existing independently of each other. This separation of the temporal and spatial axes would finally be abandoned in Minkowski space–time. Newton’s view that the spatial and temporal axes are absolute reference points for all observers means that these axes remain perpendicular to each other, irrespective of the motions involved (Fig. 2.12).

In order to appreciate this point, consider again the measurement of the circumference of the Earth, which today is known to be approximately 40,000 km. Depending on the definition of the units used by Eratosthenes he measured the circumference of the Earth to range between 35,000 and 45,000 km. Now consider an astronaut who flies past the Earth at a speed close to the speed of light. Does the astronaut measure the same value as Eratosthenes? The commonsensical answer is to say ‘yes’, since the measurement of a distance does not seem to depend on the velocity of the frame from which it is measured. And yet, as Einstein’s Special theory of relativity was to show in 1905, this commonsensical answer is mistaken. But the commonsensical answer illustrates well one sense of Newton’s absolute notions: the measurement of temporal and spatial distances depends neither on the location of the measuring device, nor on the velocity with which it is moving. On this commonsensical view, if an inhabitant of a distant galaxy could take measurements of the Earth, this galactic creature would measure the same dimensions as Earthlings.

The absoluteness of time and space can also be understood, less metaphorically, in terms of spatio-temporal structure. (Later it will be shown that Newtonian mechanics can be re-interpreted in the language of space–time.) In this sense,

...an absolute element of spatio-temporal structure is one that is well-defined independently of reference frame or coordinate system. For example, in the space–time $E^3 \times R$ of

Fig. 2.13 Planes of absolute simultaneity as the temporal distance between events E_{1a} , E_{2a} , E_{3a} is judged to be the same from the point of view of all reference frames, this means that all sister events E_{1b} , E_{2b} , E_{3b} on their respective planes are judged to happen at the same time. They are taken to be at the ‘same temporal position’, even though they may be widely apart



common sense and some version of Newtonian mechanics there are well-defined relations of being-at-the-same-spatial-position and being-at-the-same-temporal-position (simultaneity) that hold between points of space-time independently of reference frame: both space and time are absolute in this sense in such a space-time. (Friedman 1983, 63)

These relations can be illustrated in terms of Newton’s techniques. In the approximation he uses for the abstract notion of absolute space, he considers that all reference frames, which are in constant uniform motion with respect to each other, may use the ‘fixed stars’ as spatial markers to position themselves with respect to others. It is as if local observers shunned local landmarks and used celestial landmarks to position their respective locations. In terms of ‘absolute space’ or the markings on the cosmic container, all observers refer to the same place. It follows from the notion of absolute time that simultaneity must also be an absolute notion: the relation of ‘being-at-the-same-temporal-position’ means that all reference frames use the ‘astronomical clock’, which in approximation is provided by the orbits of the Jupiter moons (or today the emitted radiation from pulsars). It follows from the use of the same astronomical clock, or the temporal markings on the cosmic container walls that events, which are simultaneous for some observers, must be simultaneous for all observers (Fig. 2.13).

It may not come as a great surprise that Newton’s notions of absolute time and space became the subject of an intense debate and generated philosophical consequences. Leibniz rejected Newton’s notions and defended a relational view. Later Einstein abandoned Newton’s separation of time and space altogether.

2.8.3 Newton and Leibniz, Compared

It has already been stressed that Leibniz (1646–1716) accepted the notion of physical time and that he saw the order of events as subject to a universal principle of causality. Leibniz rejected Newton’s postulates of absolute space and time as merely

metaphysical speculations. He ignored Newton's need for approximations. Leibniz defined space as the 'order of coexisting events' and time as the 'order of succession of coexisting events' (Leibniz 1715–1716, Third Paper, 210–215). He thus made an essential reference to the 'before–after'–relationship in physical events. Leibniz does not specify what he means by 'events'. Even if his notion of 'events' refers to events in the physical world does he mean actual or also possible events? Leibniz had a predecessor to his relational view in the person of Saint Augustine. Saint Augustine, too, insisted on the 'before–after' relation between events but he also failed to specify, which events in the physical world he had in mind. It is often claimed in the literature that Leibniz must restrict the admissible events to 'the set of actual physical events' (Friedman 1983, 63; Earman 1989, Chap. I). And then Leibniz would face the difficulties, which Newton avoided by resorting to the idealization of absolute space and time. For the 'set of actual physical events' may not offer sufficient regularity to ground a stable order in the succession of events. What is not often realized, as we shall discuss further in [Chap. 3](#), is that Leibniz admits of 'possible' events. Thus Leibniz moves in exactly the opposite direction to Newton. Whereas Newton introduces his idealization of absolute space and time and then suggests approximations, Leibniz begins with 'actual' events but then moves to the idealization of possible events and 'fixed' existents.

And fixed existents are those in which there has been no cause for a change of the order of coexistence with others or (which is the same thing), in which there has been no motion. (Leibniz 1715–1716 Fifth Paper, Sect. 47, 231; cf. Manders 1982)

These 'fixed existents' therefore remain invariant in the order and succession of coexisting events. Despite the often-claimed deep gulf between Newton's realism and Leibniz's relationism about time, they share a meeting point in this interaction between idealization and approximation. Leibniz and Newton could therefore agree on the same approximations for the purpose of measuring time and space in the material world. The Leibnizian move to idealizations—'fixed existents' and possible events—can be gleaned from a consideration of the order which must hold between the events. Let us consider the Leibnizian notion of *order* from the perspectives of universality, regularity and invariance. Order may just mean 'a before–after'–relationship between events, governed by some universal principle of causality. But in his Correspondence with Clarke Leibniz goes one step further when he considers that the 'before–after' relationship between events must, in some sense, be law-like. That is, it must be governed by the laws of classical physics. It is the recognition that the cosmic machine is driven by law-like regularities, which makes Leibniz introduce the idealizations. For the laws of classical mechanics require inertial reference frames and these are provided by the invariants in the physical world.

So there are a number of disagreements and agreements between Leibniz and Newton, whatever other differences may separate them (see [Chap. 3](#)). Newton and Leibniz disagreed about the ultimate nature of time. In his philosophical discourse about time, Newton postulated that time was absolute, since it existed irrespectively of material happenings in the universe. Even an empty universe

has a 'clock' for Newton. For Leibniz the universe must be filled with events; the order of succession of events constitutes a clock. For Leibniz physical time is not absolute but relational since it depends on the succession of material events in a particular order. Leibniz agreed with Saint Augustine that before creation time did not exist. Time is coextensive with the creation of events. Despite these disagreements we can recover some common ground. Leibniz agreed with Newton that the order of coexisting physical events constituted a universal simultaneity relation between such events. That is, all observers throughout the universe agree that two events, which occur simultaneously for one observer, occur simultaneously for all observers whatever their location in the universe. Leibniz also agreed with Newton that events, which occur in succession, display the same order of succession for all observers. Thus, whether the 'before-after' relation refers to actual or possible events, there was no doubt in the minds of any observer, from which ever location they may observe the succession of two events— E_1 and E_2 —that E_1 would occur before E_2 . The aspects of regularity and invariance can also be recovered. Both agreed that the succession of events must be regular and invariant, to a certain degree, across all perspectives for measurable time to be possible. For although Newton and Leibniz did not agree on whether material happenings were needed for time to unfold in a philosophical sense, they were in agreement that time was universal. All observers throughout the universe, who are in motion with respect to each other, use the same clock, i.e., they refer to the same temporal axis. Newton secured regularity of temporal succession and spatial coexistence by the stipulation of an 'immovable space' and an 'equable flow of time'. But absolute space and time are beyond the realm of observation so that Newton introduces appropriate approximations, like the regularity of celestial systems (Jupiter moons, fixed stars). Newton secured invariance by making 'absolute' space and time irrespective of particular perspectives. Even in approximation, the regularity of celestial motion is the same for all observers, irrespective of their particular point of observation. Newton's approximations illustrate the regularity of the clockwork universe, since it is based on the law-like succession of physical events. Leibniz guarantees regularity of temporal and spatial intervals by presenting the universe as a cosmic machine, governed by mechanical laws (Leibniz 1715–1716, Second paper, 208; Fourth Paper, Sect. 32, 219). As Leibniz agrees with Newton about the universality of the order of the succession of events—time is universal for all observers, irrespective of their perspective on the world—the invariance of temporal and spatial relations is also guaranteed. As will be discussed later, Newton's mechanics has more affinities with Leibnizian relationism than his philosophy.

We have focussed on the agreement between Newton and Leibniz for the sake of highlighting in this chapter the close connection between time, regularity and invariance in classical physics. The same connection exists in modern physics—in relativity and quantum theory—where it comes to even greater prominence. This is essentially due to the fact that the classical notion of order of events, as defined in classical physics, undergoes a conceptual revision. In the Special theory of relativity both the universality and invariance of temporal relations, in the way

Newton and Leibniz understood them, are abandoned. We shall need to see how at least invariance survives in modern physics. With his notion of time as the order of the succession of coexisting events, Leibniz, like Saint Augustine before him, defines a direction of time; in our terminological convention, the ‘before–after’ relation of local events specifies the *passage* of time. Physicists and philosophers of the classical age did not consider the further question whether time displays a global *arrow* or whether the universe follows a one-directional evolution. In order to consider such questions, a different physical relationship—entropy—had to be discovered.

2.8.4 The Arrow of Time

The discussion of the topology of time showed that a consideration of the anisotropy of time suggests a distinction between the *passage* of time and the *arrow* of time (Denbigh 1981; Grünbaum 1967). The passage of time expresses our daily experience of the one-directional ‘flow’ of time, often from order to disorder, from new to old, from warm to cold. But our impression of the passage of time is compatible with different topologies of time. Time may have the topology of an open line or a closed circle but observers would not be able to tell from their limited experience of the ‘flow’ of time. The arrow of time expresses a global, one-way direction of the universe, from initial to final conditions, where these conditions are asymmetric. But the arrow of time cannot directly be observed; it needs to be inferred from physical criteria.

The arrow of time seems to be quite different from other arrows in our experience, which seem to generate order, information and complexity (cf. Savitt 1995, 4–5; Savitt 1996; Hoyle 1982)

- The *psychological* arrow. As Saint Augustine emphasized, we remember the past but we anticipate the future; we have the feeling that time ‘flows’ relentlessly forward into the future towards a period when we will no longer exist.
- The *historical* arrow: We have records of the past but not of the future; there is evidence of the geological evolution of the Earth (discovery of deep time) and the biological evolution of species towards greater diversity and complexity. Darwin represented this arrow by his image of the tree of life.
- The *cosmological* arrow: As Kant speculated, the universe has expanded from an initial, chaotic beginning, which is today located in the Big Bang, towards the formation of solar systems, galaxies and clusters of galaxies. Observations suggest that the cosmological arrow will take the universe towards a Big Chill in the distant future rather than a re-contraction to a Big Crunch. It is at present not clear whether our observable universe is only one amongst an infinity of universes (multiverse), which are born and die.
- But the *thermodynamic* arrow of time seems to point in the opposite direction, towards disorder and the destruction of information (Layzer 1975, 56; Layzer

1970; Landsberg 1996, 253; Whitrow 1980, Sect. 7.4; Hawking 1988, 145ff; Ladyman et al. 2008). Could such an increase of disorder be used to characterize the arrow of time? And how are the other arrows related to the thermodynamic arrow?

In the 19th century physics discovered the laws of thermodynamics, which were considered to shed light on the question of the arrow of time. Of particular importance were the Second law of thermodynamics and the discovery of entropy. Entropy is often characterized as an increase in disorder in a closed system.⁸ Entropy has also been associated with the loss of information, which is available about a system undergoing entropic change. In Chap. 4 it will be shown that these characterizations are imprecise and that the apparently irreversible increase in entropy is best expressed by a spreading function; entropy itself is best defined in terms of the number of micro-states, which correspond to a given macro-state in phase space. For the time being our discussion is confined to the discovery of entropy and how entropy was used to define the arrow of time.⁹ There are many everyday illustrations of the concept of entropy: food rots, hot coffee cools, if milk and coffee are mixed they cannot be unmixed, and a sugar lump dissolves in hot coffee and cannot be retrieved. But the total amount of energy always remains the same. According to the first law of thermodynamics— $Q = \Delta E_k + \Delta E_p$ —the sum of all energies remains constant in a closed system. But while all mechanical energy (work, W) can be transformed into thermal energy (heat, Q) it is not the case that all thermal energy can be transformed into work: $dE = dQ - dW$. A machine can be employed to perform useful work but some of its energy is lost in frictional heat that is no longer available to do work. To illustrate the Second law, consider a child's playroom (Fig. 4.1a, b). The father puts all the toys into a corner so that the child can play in the morning. When the child enters the room there is a great amount of order in the playroom. There is also precise information as to the whereabouts of the toys. But as the child gets to play, toys are beginning to be tossed about all over the room and when the child takes a lunchtime nap the morning order has been reduced to a lunchtime disorder. The neat ordering of the toys has disappeared and has led to a loss of information as to the whereabouts of the toys. Whilst they were stacked neatly in the corner of the room in the morning, they are now scattered randomly across the room. So the total amount of energy remains the same but entropy has increased.

⁸ A closed system has constant energy, a constant number of particles, constant volume, and constant values of all external parameters that may influence the system (Kittel and Kroemer 1980, 29).

⁹ Thermodynamics (TD) is concerned with macroscopic parameters, like temperature, pressure, and volume. It turned into statistical mechanics (ST) as a result of various reversibility and recurrence objections (see Chap. 4). Statistical mechanics is concerned with the micro-states, which make up the macro-states. The relationship between TD and ST is complex; see Sklar (1993).

Fig. 2.14 **a** A sealed container. Gas molecules are confined to compartment 1 of the container. **b** Gas molecules are no longer confined to compartment 1 and can freely roam in the available phase space

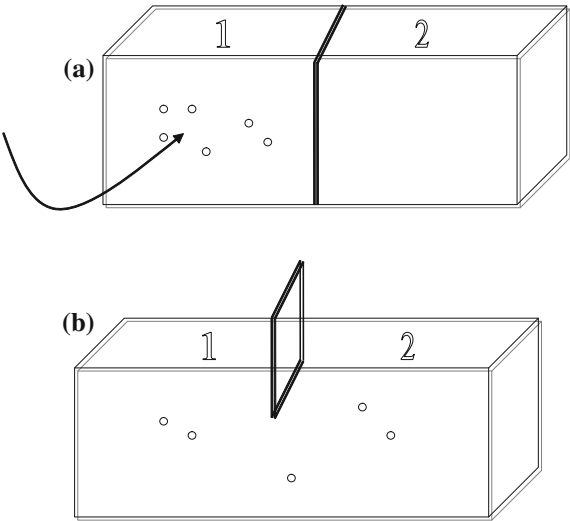
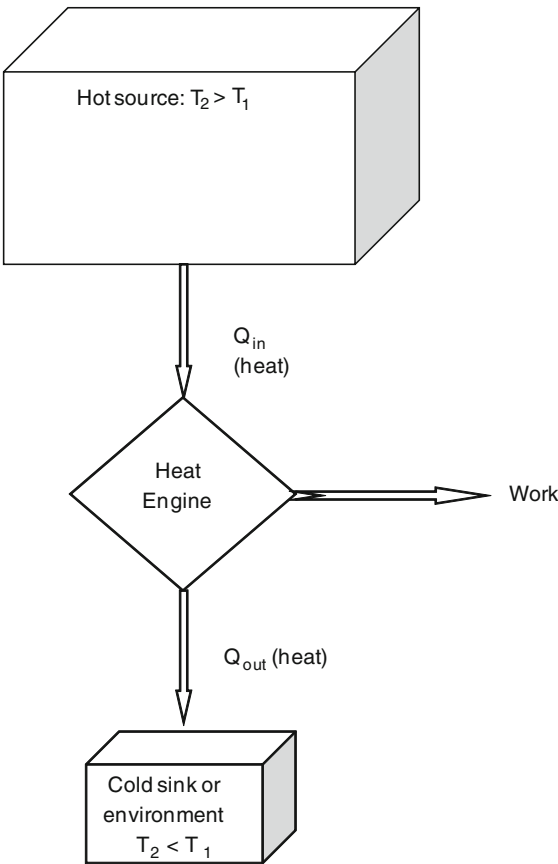


Fig. 2.15 Kelvin’s version of the second law



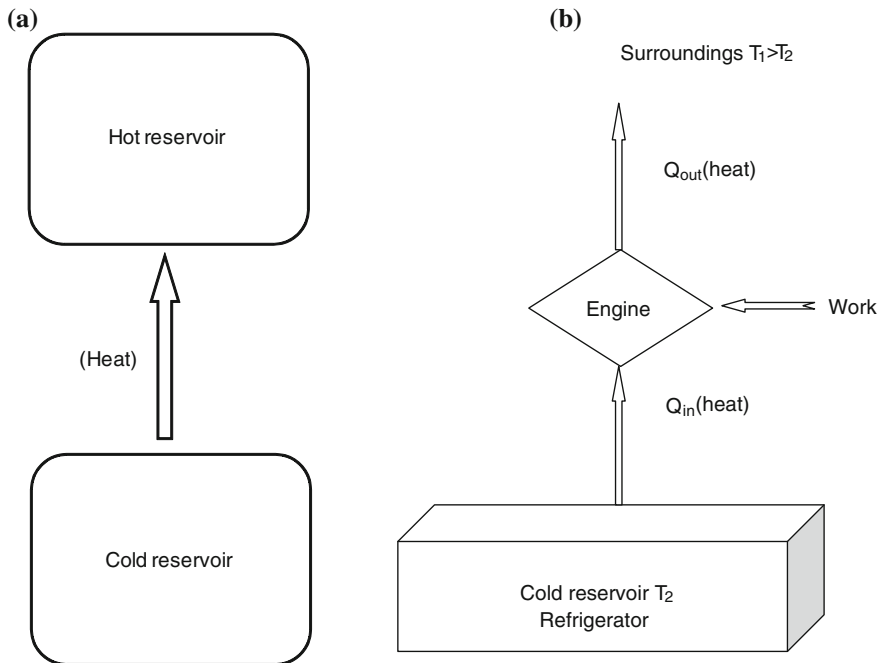


Fig. 2.16 **a** Clausius's statement of the second law. Such processes are not observed to occur in nature **b** Clausius's statement of the second law. Such processes are observed in nature

In a slightly more technical sense it is customary to illustrate the increase in entropy by modelling a closed system, like a sealed container, separated by a removable partition, such that compartment 1 contains gas molecules and compartment 2 is empty (Fig. 2.14a). At first all the gas molecules are confined to compartment 1, so that the probability of finding a gas molecule in compartment 2 is zero. Furthermore, if an atom is placed in, say, compartment 1, there is a good chance that it will be hit by gas molecules, so that the molecules can do work on the atom ($W = Fd$). If the partition wall is removed (Fig. 2.14b), the molecules will spread over the larger volume (phase space) but information about their whereabouts is lost. The 'disorder' of the system is increased, and hence the ability of the system to do useful work is reduced, since, for instance, its pressure decreases.

The notion of entropy can be made more precise, in quantitative terms, by relating it to the Second law of thermodynamics. W. Thomson (Lord Kelvin, 1824–1907) expressed it in the form that every viable engine must have a cold sink (see Atkins 2003, Chap. IV, 113; Kuchling 2001, 292; Fig. 2.15). In other words, heat can only be transformed into work when some of the heat transfers from a warmer to a colder body.

Rudolf Clausius (1822–1888) stated the Second law in the form that ‘heat never flows from a cooler to a hotter body’ (Atkins 2003, 115; Kuchling 2001, 262; Fig. 2.16a). In other words: Heat can only be transferred from a colder to a hotter body by means of mechanical work (Fig. 2.16b).

The Clausius statement means that all work can be transformed into heat but not all heat into work: $Q = W + \Delta U$. Consider a car engine, for illustration. The car engine burns fuel to move the car but not all of this fuel can be used to transform it into kinetic energy. Some of the energy is needed to run the engine, other energy is lost through friction in the engine parts and in the tyres (etc.). Hence whilst we get work out of the fuel, some of the heat from the fuel irretrievably escapes into the environment; it is not available for further work. By contrast all the work the car performs transforms into various kinds of heat (energy); it moves the car forward but also warms up the interior, it creates an air flow, it wears the tyres. While it is always possible to transform kinetic energy and work into heat, it is never possible to transform all heat into work. Clausius then defined a change in entropy for a reversible process as follows:

$$\text{Change in Entropy (dS)} = \frac{\text{Energy supplied as heat (dQ}_{\text{rev}})}{\text{Temperature at which the heat transfer occurs (T)}} = 0 \quad (2.2)$$

And for an irreversible process he defined the change in entropy as $dS \geq dQ/T > 0$. Clausius further proposed that the amount of entropy, in an irreversible process, never decreases. This leads to the Second law of thermodynamics:

$$\Delta S \geq 0 \quad (3), \quad (2.3)$$

(whilst the first law states that the total amount of energy remains constant). In all natural processes the amount of entropy (of disorder) increases, as we are well aware from our everyday experience. Clausius expressed the first law of thermodynamics in the pithy statement:

Die Energie der Welt ist constant. The energy of the world is constant.

And then applies the concept of entropy to the whole universe:

Die Entropie der Welt strebt einem Maximum zu. The entropy of the world strives towards a maximum. (Atkins 2003, 119; Torretti 2007, Sect. 6)

The increase in entropy in natural systems means that the energy available to do useful work decreases. A log of wood keeps the fire burning, which makes the water boil; but once the log has turned into ashes, it can no longer do useful work. Ashes don’t keep the fire burning.

As we shall discuss, a generation later Ludwig Boltzmann (1844–1906) felt compelled to redefine the notion of entropy for statistical mechanics. But he echoed Clausius’s generalization when he declared that the

Second law proclaims a steady degradation of energy until all tensions that might still perform energy and all visible motions in the universe would have to cease. (Boltzmann 1886, 19)

If it is true that there is a general degradation of energy, then all forms of energy will eventually be converted into 'heat' and the universe will reach thermal equilibrium. This tendency towards equilibrium was called the 'heat death' at the end of the 19th century. The scenario of the Heat Death enjoyed great popularity. It was meant to signal that the whole universe would eventually consume the amount of useful available energy, making it impossible to sustain temperature differences. At the end of the universe the total amount of energy was still unchanged but the balance between useful and useless energy had shifted towards thermal equilibrium. The sun, for instance, would no longer sustain life on Earth. Thomson (1852) spoke of 'a universal tendency in nature to the dissipation of mechanical energy' and concluded his survey with the ominous warning that 'the Earth was and will again be unfit for human habitation'.

L. Boltzmann, like W. Thomson, lifts the notion of entropy to a cosmological level in an attempt to identify the arrow of time. Whilst it is generally accepted that entropy increases to a maximum in a closed system, like a container of gas molecules, Boltzmann assumes that the Second law can be applied, under certain reservations, to the whole universe.

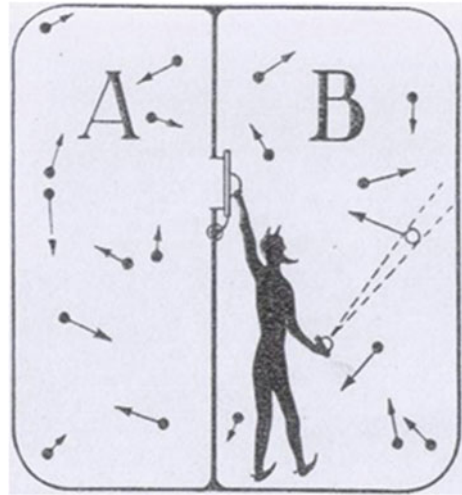
People have been amazed to find as an ultimate consequence of this proposition that the whole world must be hurrying towards an end state in which all occurrences will cease, but this result is obvious if one regards the world as finite and subject to the second law. (Boltzmann 1904, 170; cf. Uffink 2001)

It will emerge later that this assumption is questionable. Boltzmann himself pointed out that the direction of increasing entropy could be identified with the arrow of time only in certain parts of the universe.

There must be in the universe, which is in thermal equilibrium as a whole and therefore dead, here and there relatively small regions of the size of our galaxy (which we call worlds), which during a relatively short time of eons deviate significantly from thermal equilibrium. Among these worlds the state probability increases as often as it decreases. (...) a living being that finds itself in such a world at a certain period of time can define the time direction as going from less probable to more probable states (the former will be the "past" and the latter the "future") and by virtue of this definition he will find that this small region, isolated from the rest of the universe, is "initially" always in an improbable state. (Boltzmann 1897, 242; cf. Boltzmann 1895a, 415; Sklar 1992, Chap. 3; Zwart 1976, Chap. VI, Sect. 3; Whitrow 1980, Sect. 7.3; van Fraassen 1970, 90ff)

Boltzmann holds that the arrow of time lies in the direction, in which a small region of the universe moves closer to equilibrium. Such small regions house life-supporting planets like the Earth. Boltzmann proposes that the surrounding universe is in thermal equilibrium but there are local, entropy-decreasing fluctuations; our world is in an ordered state, away from equilibrium, because it exists in an entropy-reduced fluctuation state. Although the entire universe languishes in a state of Heat Death, in which time no longer exists, the inhabitants of low-entropy fluctuation

Fig. 2.17 Maxwell's Demon, Wikimedia Commons



states see their time as anisotropic and may conclude that their ‘universe’ displays an arrow of time (cf. Boltzmann 1923, Sect. 90).

2.8.5 *Maxwell’s Demon*

Boltzmann, however, soon encountered problems with the thermodynamic notions of entropy and disorder, which forced him to reinterpret the Second law in a statistical manner. The understanding of the Second law as a probabilistic, rather than a deterministic law, in turn had serious consequences for the identification of the arrow of time with entropy. This shift constituted the beginning of the transition from thermodynamics to statistical mechanics.

In a famous thought experiment, involving ‘a being with superior faculties’, James Maxwell attempted to show that the Second law of thermodynamics only possessed statistical validity (Maxwell 1875, 328–329; cf. Earman and Norton 1998, 1999; Leff and Rex 2002; Hemmo and Shenker 2010). 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. 2.17). 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 principle, which holds with overwhelming probability for systems composed of many particles. 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 averages over micro-states (of molecules) but leaves us ignorant about their individual properties (position, momentum):

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; cf. Myrvold 2011)

Physicists ever since have attempted to show that Maxwell's demon cannot achieve his aim; and that the demon himself is subject to the Second law.¹⁰ For if the demon succeeded, our energy problems would be solved:

Machines 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)

One reason why the demon cannot escape the effects of the Second law is that, 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.

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, Sect. 46.3)

¹⁰ 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 Baeyer (1998); Leff and Rex (2003).

Nevertheless, J. C. Maxwell achieved his aim of ‘picking a hole in the second law’. The Second law is a statistical principle, which raises several questions for the notion of time:

- *Reversibility*: It is not a violation of the Second law if all the molecules in the container in the course of time return to their initial state. But this return is unlikely to happen in the lifetime of the universe. Yet there is a fundamental split between the reversibility of the micro-states and the overwhelming irreversibility of the macro-states. What then is more fundamental: the temporal symmetry of the dynamic laws, which govern the micro-states or the temporal asymmetry of the macro-states, with which we are familiar?
- *Arrow of Time*: If the second law is a statistical law, is it still possible to associate the arrow of time with an increase in entropy? Do the Boltzmann fluctuations mean that the arrow of time reverses when entropy decreases? And is the universe on a trajectory towards a Big Chill or a Big Crunch, a return to its original state? Is the universe symmetric rather than asymmetric?

As will be shown in [Chap. 4](#) such questions cannot be answered within classical thermodynamics. They require tools as furnished by quantum cosmology. But this Chapter first completes its survey of the contributions of physics to the notion of physical time.

2.9 Time in Modern Physics

It is important to make a sharp distinction between an asymmetry of the world *in* time from an intrinsic asymmetry *of* time itself. (Davies 1994, 120; italics in original)

It has been argued that there is a need for physical time and that physical time, because of the uniformity in the sequence of events, on which it is based, is measurable time. Both the theory of relativity and quantum theory made significant contributions to our understanding of measurable time, because they threw further light on the notions of regularity and invariance.

2.9.1 *The Measurement of Time in the Special Theory of Relativity*

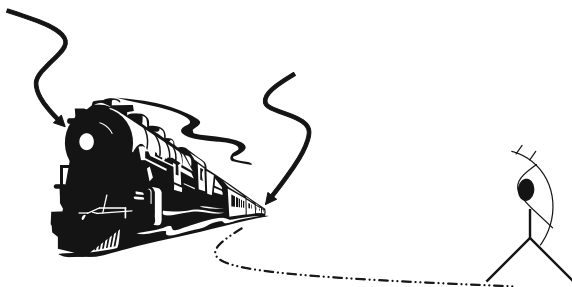
Much of classical physics is concerned with absolute reference frames, in which the laws of motion are embedded. Reference frames can be characterized as coordinate systems, which are in inertial, non-accelerated, motion with respect to each other. The coordinate axes provide the units of spatial and temporal length. Prior to Einstein, many thinkers, from Aristotle to Newton, assumed that absolute reference frames existed, against which all other, merely relative motions could be measured.

These absolute reference frames were often regarded as privileged in comparison with the mere relative motions. In Greek cosmology, for instance, the position of the motionless Earth, near or at the centre of the universe, constituted such a privileged position. The natural motion of earthly objects, like falling stones, was to ‘strive’ back towards the Earth, since the ‘centre’ was their ‘natural’ place. The ‘fixed’ stars also provided absolute reference systems, against which the orbits of the planets could be tracked. Although around the time of the Scientific Revolution the Aristotelian theory of motion had long been abandoned, Newton still formulated both space and time as absolute reference frames, against which all other, merely relative motions are to be measured. Such systems are called ‘absolute’ because they furnish the ultimate unchanging standards, against which all other motions can be gauged. Recall the two ‘intelligent’ molecules floating randomly in their container world. When they meet, the question ‘where are you?’ makes little sense as long as they only have their relative motions and positions to orient themselves. An answer like ‘within your sight’ or ‘at arm’s length’ would not be very informative. But if on the inside of the container two axes are painted: one vertical ‘y’ axis along the side and a horizontal ‘x’ axis along the floor, it is easy for the two intelligent molecules to position themselves inside the container. The ‘x–y’ system on the inside of the container would serve as an absolute reference frame, compared to their relative motions (Fig. 2.11). Such ‘absolute’ reference frames were often regarded as fixed—like the stationary Earth in Greek cosmology—and immaterial, like Newton’s philosophical notion of absolute container space and the metaphorical river of time (cf. Smart 1956). They were non-dynamic, since they served as a canvass against which the material world evolved (cf. Rynasiewicz 2000; Friedman 1983). And although Leibniz did not agree with Newton that these absolute reference frames were immaterial, he regarded them, in agreement with classical science, as *universal* in the sense that all observers in the universe agreed with these standards.

As science progressed in the 19th century it failed to find evidence for the existence of absolute reference frames. At first, 19th century physicists believed that an all-pervasive ether filled all of space, which served as an absolute scale, against which all relative motions on Earth could be measured. Such an ether-filling medium would fulfil the same functions as Newton’s absolute space, since it would constitute an absolute reference frame, which is both regular and invariant. However, physicists could not find any experimental evidence for the ether—the famous Michelson-Morley experiment (1887) drew a null result. The experiment showed that the existence of such an ether or absolute reference frame could not be detected. If there was no evidence of absolute space, there was no evidence of absolute time either. Absolute motion could not be detected.

At the beginning of the 20th century, Einstein’s famous Special theory of relativity (1905) showed further that such absolute reference frames were not needed for the formulation of the laws of physics. As we shall see (Chap. 3), the Special theory of relativity reawakened interest in the Parmenidean-Kantian view of time and it requires a reinterpretation of the Leibnizian notion of order. The Special theory of relativity does not apply to cosmology but it laid the foundation

Fig. 2.18 Relative simultaneity. Only for the human observer on the embankment do the bolts of lightning strike simultaneously



for the General theory of relativity, which applies to large-scale cosmic structures. In both theories the time used is clock time, since the effect of motion on various kinds of clocks is taken into consideration. The Special theory is not, strictly speaking, a revolutionary theory. It is best thought of as an extension of classical mechanics. But it had a revolutionary effect on the classical notion of time in the same way in which the General theory had an effect on the notion of space. Both theories work with the notion of space–time. The Special theory had an immediate impact on the measurement of time. Einstein used light clocks because his theory of relativity shows that mechanical clocks are no longer reliable indicators of universal time. The impact was threefold—it concerned (a) the simultaneity of events; (b) the synchronization of clocks and (c) the universality of time.

1. In one of his famous thought experiments Einstein considers a fast-moving train whose front and rear ends, as it passes an embankment, are hit by bolts of lightning (Fig. 2.18). Two observers witness the event. One sits in the middle of the train, the other observes the events from the embankment. Only the embankment observer will see the two events as simultaneous, whilst the train passenger experiences the events as non-simultaneous. Within their respective frames the events occur regularly but they are not invariant across the two frames moving inertially with respect to each other. The reason for this curious situation is the finite, constant velocity of light. The observer on the train ‘rushes’ towards the ray of light coming from the front but ‘runs away’ from the ray coming from the rear end of the train. The two signals hit the embankment observer’s retina at the same time. Nevertheless, this situation does not give rise to relativism, since the observers can calculate the respective values of the arrival times of the signals. The rules they use are known as the Lorentz transformations. The effect of these rules can be compared to the transformation rules used by Archimedes and Eratosthenes, who—according to a previous thought experiment—used their knowledge of astronomy to calculate the highest point of the sun in their respective locations.

Einstein’s train thought experiment illustrates the so-called ‘relativity of simultaneity’, which is a consequence of the Special theory of relativity. In this illustration, an event, which is simultaneous for the embankment observer is no longer simultaneous for the train-bound observer. Thus Einstein abandons the

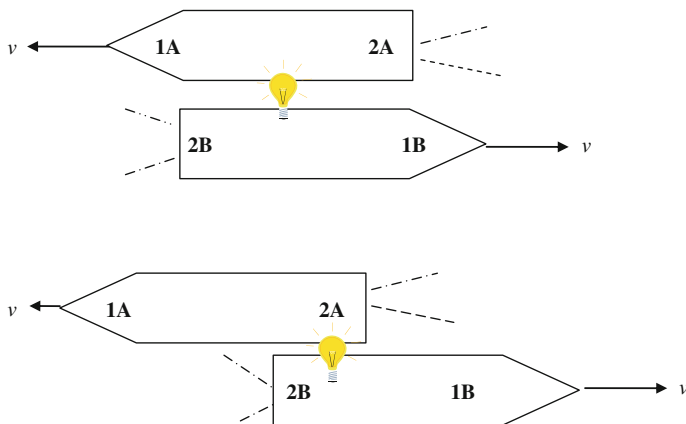


Fig. 2.19 Disagreement about the synchronization of clocks in two reference frames in relative motion to each other. This is an example of the path-dependence of clock time (adapted from Gamow 1971, 13–14)

universal simultaneity assumption in classical physics. Relative simultaneity, as Einstein insists, is one of the arguments for the relativity of time.

Events which are simultaneous with reference to the embankment are not simultaneous with respect to the train, and *vice versa* (relativity of simultaneity). Every reference-body (co-ordinate system) has its own particular time; unless we are told the reference-body to which the statement of time refers, there is no meaning in a statement of the time of an event. (Einstein 1920, 26; italics in original)

2. Consider some space travellers who experiment with the clocks, which are attached to their space rockets. Their rockets serve as inertial reference frames, which travel near the speed of light. There are four observers, each one occupying a corner of the spaceship (1A, 2A and 1B, 2B, respectively) travelling in opposite directions (Fig. 2.19). They have set their clocks internally according to the method of light clocks, discussed below. They now wish to compare their clocks. When the spaceships' two midpoints coincide, a light signal is set off, which propagates towards the front and rear ends respectively. As light travels with finite constant velocity, and the spaceships travel relative to each other, the observers 2A and 2B will be closer to the light source than observers 1A and 1B.

- The signal will reach observer 2A earlier than 1B who will receive it after some additional time. According to 2A the clock of 1B lags behind time.
- The signal will reach 1A later than 2B, so that 1A will conclude that 2B's clock is ahead of time.

As their clocks are synchronized within their respective spaceships, observers in A will claim a delay between the watches of the other observers in B, and vice versa. Both A and B are correct from their respective points of view. It makes no sense to ask whose clock is 'really' showing the right time. The notion of absolute

simultaneity has vanished. Clearly the clock readings are perspectival. There are other effects on clocks, to be discussed later, which lead to similar consequences.

3. Finally, consider the loss of the universality of temporal measurements, which was inherent in Newtonian physics. Recall that Leibniz agreed with Newton that time ticks at the same rate for every observer in the universe. They agreed that time was universal although they disagreed about whether it was also absolute (independent of material events). Einstein abandoned the universality assumptions of classical physics. When clocks do not seem to tick at the same rate for all observers, it becomes difficult to agree on the temporal lengths of events. By abandoning the classical universality notion, he also abandoned the classical invariance notion. Thus the clocks run regularly for inertial observers within their own reference frames but *across* inertial reference frames the respective clocks are seen as running either ‘fast’ or ‘slow’, depending on the velocity of the frame, and the viewpoint from which they are observed. Spatial and temporal measurements become perspectival features of inertial reference frames, a phenomenon known as ‘time dilation’.

From the point of view of the measurement of time, the most striking consequence of these thought experiments is the loss of universality and invariance, though not regularity. The classical notion of *universality* of clock readings is lost because every reference frame carries its own clock but if these clocks are in inertial motion with respect to each other the clock readings will not agree. The *invariance* across different reference frames is lost because clocks, as judged from respective reference frames in inertial motion with respect to each other, are considered to tick at different rates (time-dilation). In other words, clock readings are path-dependent. Clocks record the length of a time-like curve connecting two events. But clock time retains its *regularity* within a particular reference frame—according to observers in these particular frames, the clocks run ‘normally’. It seems that the only invariant clock time, which still exists in the Special theory of relativity, is the clock time measured along the world line of each particular reference frame (Denbigh 1981; Chap. 3, Sect. 7; cf. Petkov 2005, Sect. 4.9; Hoyle 1982, 94). However, the loss of the universality and invariance of clock time is compensated for by new types of invariance. For Einstein postulated that the velocity of light, c , is invariant in all inertial reference frames. In 1676 the Danish physicist Olaf Rømer had determined the velocity of light by observations of the Jupiter moons; he found the 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). In his concern for invariance Einstein shares the same worry as Newton who postulated absolute time and space as a background for the formulation of the laws of motion. These ‘invariants’ are one reason why Einstein’s theory of relativity is not a relativist theory. What Einstein means is that the theory of relativity needs to recover invariant relationships, despite the perspectival nature of spatial and temporal measurements. Absolute yardsticks like the ether, absolute space and time are replaced by a number of invariant relationships, which may serve as the basis for the measurement of physical time.

2.9.2 Invariants of Space–Time: Speed of Light

The best known invariant relationship is the invariance of the speed of light, c , which Einstein raises from a mere empirical finding in classical physics to a postulate of his theory. The invariance of c means all observers, who move inertially with respect to each other in what is known as Minkowski space–time, measure the same constant speed of light. The speed of light is *neither* dependent on the velocity of the emitting body *nor* on the direction, in which the light ray is emitted. To fully appreciate the significance of this result, compare it with an ‘ordinary’ situation, as in Fig. 2.18, but in which a ‘shooter’ is now positioned on top of the moving train whilst an observer on the embankment watches the action. The shooter first aims his rifle in the direction of the moving train and then in the opposite direction, and fires. Let the train move at 30 m/s and the bullet at 800 m/s. What is the velocity of the bullet? According to classical physics, to determine the velocity of the bullet the two velocities need to be added, according to a particular rule. This rule is known as the addition-of-velocity theorem: $w = v + v'$. As the shooter is part of the moving train system—the moving train is his reference frame and he does not move with respect to it ($v = 0$)—the bullet has a velocity, v' , of +800 m/s in the direction of the train and a velocity of –800 m/s in the opposite direction. But the observer on the embankment finds himself in a different reference frame, which is at rest with respect to the moving train. Applying the addition-of-velocity theorem, he will calculate a combined velocity of +830 m/s in the direction of the train— $w = 800 + 30$ m/s—and a combined velocity of +770 m/s in the opposite direction— $w = 800 - 30$ m/s. But if this classical theorem is applied to the velocity of light, c , the embankment observer will arrive at a contradiction with the postulated constancy of c . To see this, let the person on top of the train shine a torch, instead of firing a rifle, in both directions. The light rays from the torch travel at a velocity c . The embankment observer will add the velocity of the train, v , to the velocity of light, c , and s/he will subtract the velocity of the train from the velocity of light, and hence will arrive at different values for the two situations. If the value of c really is a constant ($\approx 3 \times 10^8$ m/s), then it must have the same velocity in all situations. Hence the addition of velocities theorem of classical physics must be wrong for it yields superluminary velocities ($c + 30$ m/s) in one scenario and subluminary velocities in the other ($c - 30$ m/s). Einstein proposed that the old theorem $w = v + v'$ be modified to reflect the constancy of c :

$$w = \frac{v + v'}{1 + \frac{v \cdot v'}{c^2}} \quad (2.4)$$

This new velocity theorem yields the old theorem in the limit when v and v' are much smaller than c . But it never renders a velocity greater than c . Let the train move at a velocity near the speed of light, c_1 , and let the train-bound observer again shine a light, c_2 , in both directions. On applying the new velocity theorem:

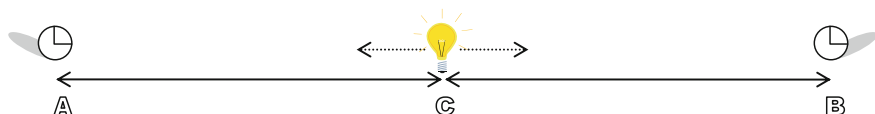


Fig. 2.20 Synchronization of distant clocks

$$w = \frac{c_1 + c_2}{1 + \frac{c_1 c_2}{c^2}} = \frac{2c}{1 + 1} = c \quad (2.5)$$

the embankment observer will calculate a combined velocity of c for the two situations, in agreement with the postulate. For the train-bound observer $w = c$.

Einstein also used light clocks to compensate for the lack of synchronization of moving clocks (Einstein 1905, 30). How can two clocks, which are separated in space, be synchronized? An observer, C , can be placed at midpoint between the two clocks at A and B . A and B emit light signals towards C . The two clocks are synchronized when the signals arrive at C at the same time (Fig. 2.20). These two clocks will be in synchrony according to this procedure, if the condition $t_A - t_C = t_B - t_C$ is satisfied.

The use of light clocks also allows a definition of the simultaneity of two events in a given reference frame. Imagine that observers A and B , after having synchronized their clocks, move to positions A and B , where they explode bombs. This is a version of the train thought experiment. C , who is equidistant between A and B , will see the light from the two explosions at the same time and hence will conclude that they exploded simultaneously. Thus two events in a given reference frame are simultaneous, according to this procedure, if the light signals from the events reach an observer halfway between the events at the same time, according to his clock.¹¹ But, as has been emphasized, two events, which are simultaneous in one reference frame, S (like the train platform or spaceship A), are not simultaneous in another reference frame, S' (like the moving train or spaceship B).

As we shall see in Chap. 3, this train thought experiment may be dubbed the *locus classicus* for the prevailing view that the passage of time, according to the Special theory of relativity, is a human illusion, since every reference frame records its own clock readings, which depend on the velocity of the frame. There are ‘as many times, as there are reference frames’ (Pauli 1981, 15). There are also as many ‘Nows’ as there are reference frames, with the consequence that there is no universal Now. It seems, then, that time cannot be a feature of the physical universe. For the slicing of four-dimensional space–time into a ‘3 + 1’ perspective—three spatial and one temporal dimension—depends on the velocity of the reference frame, to which respective observers are attached. If the passage of time

¹¹ The apparently simpler method of moving clocks, which have been compared and synchronized in one location, to different positions, does not work because moving clocks run slowly according to the time dilation result of the Special theory of relativity (see Tipler 1982, 941; Torretti 1999a, b, Chap. V).

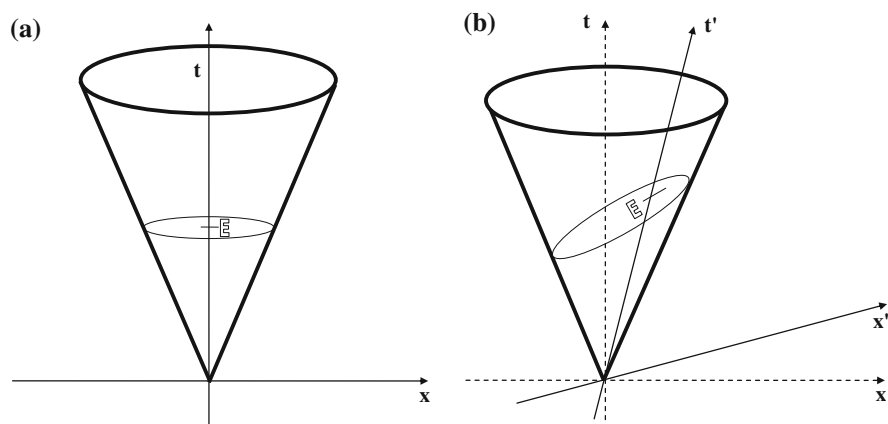


Fig. 2.21 a Stationary observer and simultaneity plane to E b Moving observer and simultaneity plane to E . The simultaneity plane has the same inclination as the x' -plane. Note that the light cone does not tilt

is a human construction, then the physical universe must be a Parmenidean block universe. Heraclitean flux is an illusion. This reasoning drove many scientists and philosophers to accept a Kantian view of time.

Due to the absence of an observer-independent simultaneity relation, the Special theory of relativity does not support the view that ‘the world evolves in time’. Time, in the sense of an all-pervading ‘now’ does not exist. The four-dimensional world simply is, it does not evolve. (Ehlers 1997, 198)

But does this perspectivalism justify the claim, often made in connection with such thought experiments, that time is a mental construct and that the physical world is an atemporal block universe?

2.9.3 Further Invariants in Space–Time

Although temporal and spatial measurements are perspectival in the Special theory of relativity—there are as many times as there are reference frames in inertial motion—there exists nevertheless an invariant four-dimensional space–time structure—Minkowski space–time—which is the same for all inertial observers.

Minkowski space–time is a geometric construction, which reflects both the perspectival aspects of the theory and its invariant structure. The invariant structure is expressed by the invariance of the speed of light, c and the space–time interval, ds .

Consider, first, how the perspectival features of the geometric representation are captured in space–time diagrams. In Fig. 2.21a, b the simultaneity planes of the

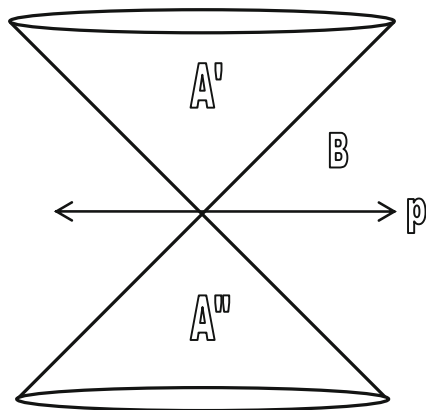


Fig. 2.22 Basic space-time structure of Minkowski space-time. The horizontal line defines the present; the inclined lines define the limits of the light cones, which no material particle, moving at less than the speed of light, may penetrate. Event A'' can affect the present time, and the present time can affect the future event A' . But the event B , outside of the lightcone, cannot be reached by any known material processes. This structure is the same for all observers: light cones do not tilt

respective observers are inclined at oblique angles, which reflect the different velocities of the reference frames to which each observer is attached. The observers disagree about the simultaneity of events, as well as their duration and the spatial extent between them. Nevertheless, they are able to compute each other's measurements by the employment of the Lorentz transformations, which are therefore invariant.

Second, the invariance of the speed of light, c , for all observers invests the geometric construction with a causal structure (Fig. 2.22).

From each event, p , a light cone diverges into the future, and a light cone converges from the past onto p . These light cones can be connected by signals, and hence an earlier event can have a causal effect on a later event, within the light cone, but events lying beyond the boundary of the light cones cannot be reached by either light or other electromagnetic signals. The invariant structure of Minkowski space-time means that the temporal order of events is fixed for all inertial observers, even though they judge the simultaneity of events differently.

Chapter 3 will investigate the philosophical consequences of Minkowski space-time, in particular the claim that the perspectivalism of temporal measurements means that time is a mental construct in the Kantian sense, and that the physical world is atemporal.

We shall also see that there are two different ways of interpreting c , giving rise to two different views of Minkowski space-time. The *geometric* view is the standard approach, in which c is a postulate, which defines the geometric limit of the light cones, within which all material particles trace out their world lines. But c can also be seen as an optical signal and this defines an *optical* approach. From the point of view of an optical approach, which leads to a light geometry, c propagates with a finite velocity between events in space-time and is subject to

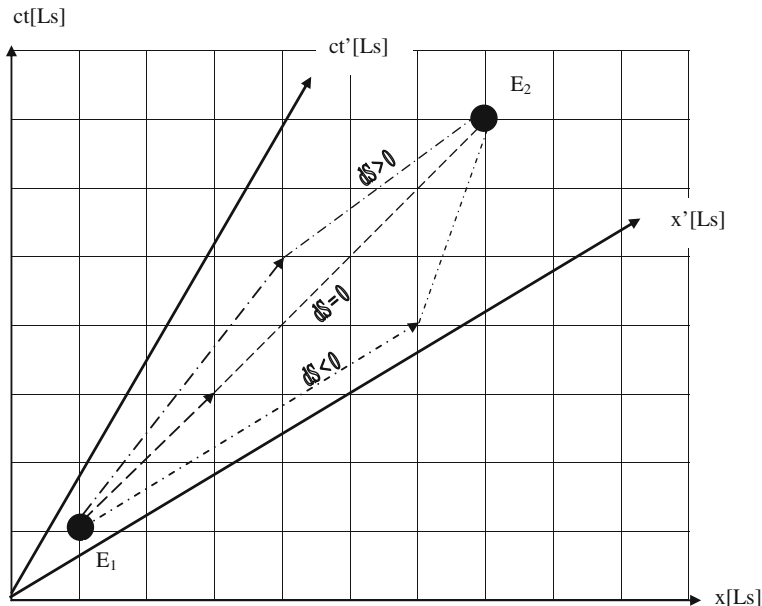


Fig. 2.23 Various world lines lead from event E_1 to event E_2 , as seen in two different reference systems moving relative to each other. The primed systems moves at $v = 0.6c$, which means that its temporal and spatial axes are inclined at 31° . Each square of the grid has a unit of ten. The length of each world line is computed from the equation $ds = \sqrt{c^2 dt^2 - dx^2}$, where $c = 1$. Although the coordinates of each event are different in each reference system, because of the effects of time dilation and length contraction, the invariance of the space–time interval— $ds = ds'$ —means that its length is the same in each reference frame

thermodynamic considerations. These two approaches lead to very different interpretations of Minkowski space–time and the notion of time (Chap. 3).

A further invariant relationship is the space–time interval, ds , which starts from the well-known fact that separate measurements of spatial and temporal lengths in Minkowski space–time are subject to the phenomena of time dilation and length contraction. This has the result that observers, who are attached to different reference frames, moving inertially with respect to each other, will not agree on (a) the simultaneity of two events (Fig. 2.18), (b) the temporal duration of events (Fig. 2.23), (c) the spatial lengths of rods. For these observers the clocks run regularly within their respective frames (proper time), but are not judged as ticking invariantly across the observers' inertial frames (co-ordinate time). However, if the observers unite space and time, in Minkowski's words, then they will agree on the space–time interval ds (Fig. 2.23) such that

$$ds = \sqrt{dx^2 + dy^2 + dz^2 - c^2 dt^2} = ds' = \sqrt{dx'^2 + dy'^2 + dz'^2 - c^2 dt'^2} \quad (2.6)$$

Although the observers record different clock times in their respective frames, the space–time separation, ds , of events in space–time is invariant. This means that observers will agree on a basic succession of events in space–time, as long as the possibility of a causal signal connection between them exists. Thus they will disagree on the simultaneity of events but agree on their order and space–time separation. The time order of events is invariant within the respective light cones.

Whilst ds and c are the best-known invariant relationships in the Special theory, and often used in philosophical arguments against the block universe, they are not the only invariants. In the geometric approach ds and c appear as geometric constructions, which either set the limit of the light cones or specify a space–time interval between events, which is invariant for all time-like related frames. But from the point of view of an optical approach, which leads to a light geometry, c propagates with a finite velocity and is subject to thermodynamic considerations. This means that the space–time interval ds , which guarantees invariance, must be grounded in empirical relations, still to be specified.

2.9.4 *The Measurement of Time in the General Theory of Relativity*

As is well-known, soon after the completion of his Special theory of relativity (1905), Einstein realized that this theory had limitations, which needed to be overcome. The first limitation was that the Special theory still gives undue preference to inertial reference frames. It extended the Galilean relativity principle from mechanical to electrodynamic phenomena but it was a ‘principle of the physical relativity of all *uniform* motion’ (Einstein 1920, Chap. XVIII, 59; italics in original). In order to include non-uniform motion in the principle of relativity, Einstein extended the special principle of relativity to a general principle.

All bodies of reference K , K' , etc. are equivalent for the description of natural phenomena (formulation of the general laws of nature), whatever their state of motion. (Einstein 1920, 61)

Thus the General theory of relativity deals with uniform *and* accelerated motion and assumes the equivalence of inertial and gravitational mass. The Special theory of relativity is only valid in the absence of gravitational fields. But gravitational fields have an effect on the acceleration of particles. In a famous thought experiment Einstein concluded that the force, which particles experience as gravitational acceleration, could be attributed to the local curvature of space–time. The second limitation of the Special theory was that it treated Minkowski space–time as an undynamic geometric structure. It affected the behaviour of measuring rods (length contraction) and mechanical clocks (time dilation) but the material world had no effect on the space–time structure. In the General theory the space–time structure itself becomes, in Wheeler’s words, fully dynamic in the sense ‘that matter tells space–time how to curve and curved space–time tells matter how to move’. For

Einstein space–time is ‘not necessarily something to which one can ascribe a separate existence independently of the actual objects of physical reality’ (Einstein 1920, Note to 15th edition).

What effect does gravitation have on the measurement of time? The ticking rate of a clock will be affected by its position in the gravitational field. The stronger the gravitational field, the slower its rate. If two clocks are placed in a gravitational field, which varies in strength, the two clocks will not tick at the same rate. A clock at the bottom of a mountain will run more slowly than a clock at the top of a mountain, because of the gravitational gradient between top and bottom—this phenomenon is called *gravitational time dilation*. Generally, clocks at rest in a strong gravitation field run more slowly by a factor of $1 + GM/c^2 r$ than in a weaker field (Clemence 1966). If a clock moves with uniform velocity through a gravitational field both effects—time dilation due to its velocity and its position in the gravitational field—have to be taken into account. In a famous experiment—the Maryland experiment (1975–1976)—atomic clocks were flown around the Earth in commercial airlines at a constant height and their time was compared to atomic clocks stationed on Earth. Compared to the earth-bound clocks, the moving clocks experienced two effects: due to the velocity effect the airborne clocks ran more slowly by 6 ns; and due to the gravitational effect they ran faster than the earth-bound clocks by 53 ns. The net difference is $53 - 6 \text{ ns} = 47 \text{ ns}$ (see Sexl and Schmidt 1978, 37–39; Thorne 1995, 100–103).

Thus the General theory reaffirms the relativity of temporal measurements. The space–time of General relativity also seems to suggest a fundamental, timeless world. Einstein himself affirmed that the General theory had deprived time of the ‘last vestiges of physical reality’ (Einstein 1916, Sect. 3). As we shall see, in the General theory the Lorentz transformations of the Special theory are replaced by more general coordinate or manifold transformations (so-called diffeomorphisms). Yet there are no transformations between coordinate systems without the retention of some invariant structure. In the case of the General theory, this is still the space–time interval, δs , but in the presence of gravitational fields it takes the general form:

$$ds = g_{ik} dx^i dx^k,$$

where the g_{ik} are functions of the spatial coordinates, x^1, x^2, x^3 and the temporal coordinate x^4 (cf. Wald 1984).

The measurement of time in the General theory then reemphasizes the point, which reverberates throughout this chapter. On the one hand, there is a strong tendency in modern physics to relegate the notion of time to a Kantian ‘pure form of intuition’ and to consider the physical world to be a Parmenidean block universe; on the other hand, there are a number of invariant relationships, whose impact on the notion of time has not yet been properly appreciated. The question, which will be considered in the following chapters, is what effect the criterion of invariance will have on the notion of time in modern physical theories and whether it allows an inference to a Heraclitean universe. Although the General theory

applies to the large-scale structures of the universe, it does not seem to change the basic doubt that some writers may harbour about the objective passage of time. Will this doubt persist if the attention turns to the very small and probes the measurement of time in quantum mechanics?

2.10 The Measurement of Time in Quantum Mechanics

Effects like the relativity of simultaneity, the path-dependence of clocks in inertially moving systems and gravitational fields, have seduced many philosophers and physicists into modelling reality as a Parmenidean block universe, in which time seems to be reduced to a mere human illusion. But the question remains whether all these macroscopic considerations about the passage of physical time are not thrown into doubt when the microscopic world comes to the fore. Clocks are atomic systems to which quantum mechanics applies.

Quantum mechanics—the physics of the atom—illustrates perfectly well the tension between flux and stasis, which so far has been associated with the theory of relativity. The fundamental equation of quantum mechanics—the Schrödinger equation—is said to be time-symmetric, whilst the measurement process of quantum systems leads to irreversibility. This difference marks a further reason, apart from the theory of relativity, why many people are tempted to opt for stasis, instead of flux: the fundamental equations of physics, like the Schrödinger equation, are time-reversal invariant. Thus the same tension persists on the microscopic as on the macroscopic scale. Our experience of the world around us is temporal: we observe almost exclusively asymmetric phenomena. They give us the impression of the objective passage of time. But on a more basic level of fundamental equations physical systems are perfectly time-symmetric. Can this tension be resolved?

What, again, does temporal symmetry mean? The Introduction briefly hinted at a distinction between the time-reversal invariance of solutions and that of laws. To illustrate the time-reversal invariance of solutions, imagine you were shown a film of some ordinary physical process, like the orbit of planets around a central body. You are then asked to decide whether the film was running forward or in reverse. If you cannot decide—and in the case of orbiting planets it is impossible to decide—then the process is time-reversal invariant. If, however, you were shown a film of a vase, which falls from the top of a table and smashes to pieces on the floor, you would find it very natural to reject the suggestion that the backward-running film was showing a physically possible situation. Yet the laws of statistical mechanics 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 in the lifetime of the universe. The time-reversal invariance of laws means that the equations of physics admit—as physically possible—processes, 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 processes. 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 condition of the time-reversed process becomes the time-reversed final condition of the initial process. To illustrate, consider an object, which freely falls to the ground from a certain height, h . If the fall time, t , is given, the height, h , is determined by the equation:

$$h = \frac{vt}{2} \quad (2.7)$$

(where h is the distance the object has travelled after time, t , and v is its final velocity.) Let the object fall for 10 s near the surface of the Earth, then its final velocity ($v = gt$) is approximately 100 m/s (where $g \approx 10 \text{ m/s}^2$). Then height, h , is

$$h = \frac{100 \text{ m/s} \cdot 10 \text{ s}}{2} = 500 \text{ m} \quad (2.8)$$

But imagine that instead of running forward a backward running clock is used, so that it counts -10 s for the fall. Then the height traversed is -500 m , which is physically possible, since the object may fall from the top of a tower (set $h = 0$) to the bottom of the tower ($h = -500 \text{ m}$).

The two notions of time-reversal invariance are distinct:

- A time-symmetric solution may not have been generated by a time-reversal invariant law. According to the cyclic theory of time, the past and future states of the world look identical and an observer would not be able to tell whether the observation of Plato and his friends lay in the past, the present or the future. And yet, as we found, the cyclic theory of time is incoherent.
- A time-reversal invariant law may have asymmetric solutions, as is the case in quantum mechanics. The Schrödinger equation is an example of a deterministic, time-symmetric equation (under certain transformations), which governs the linear evolution of a quantum system. But the measurement of quantum processes produces asymmetric solutions (for all practical purposes).

The Schrödinger equation describes the rate of change of the wave function, Ψ or the state vector, $|\Psi\rangle$ with respect to time.¹² There is only one time coordinate for all quantum systems (time is not an operator in standard quantum mechanics) so that the parameter, t , stands for a universal, external time in the Newtonian sense. The expression $|\Psi\rangle$ is an example of the so-called ket notation. These

¹² Good descriptions of basic quantum mechanics can be found in Penrose (1995), Chap. 5; cf. Penrose (2005), Chaps. 12, 23, 30. See also Box IV, Chap. 4 for a description of basic quantum mechanics. Note that the Schrödinger equation is not invariant under the application of a simple time-reversal operator, $T: t \rightarrow -t$ (i.e., if $\Psi(t)$ is a solution, $\Psi(-t)$ may not be a solution) but it is time-reversal invariant under the Wigner transformation, which involves complex conjugation of the wave-function to express the reversal of momentum.

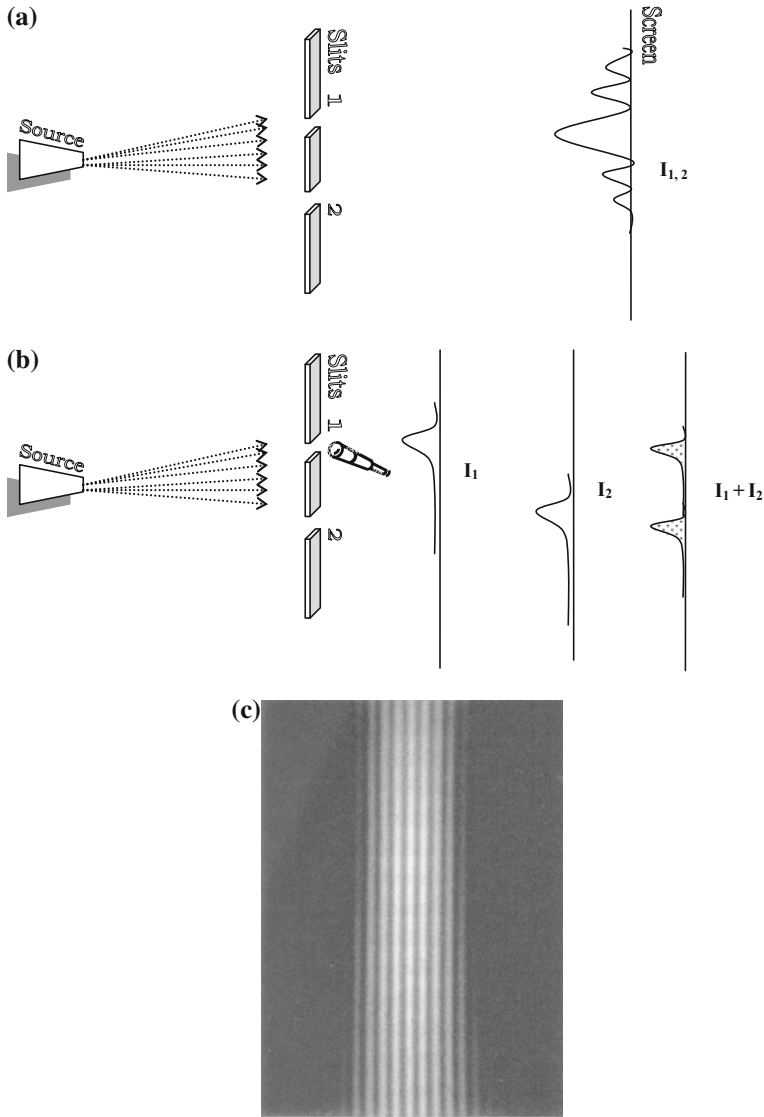


Fig. 2.24 **a** Wavelike behaviour of electrons, when both slits are open. Even when electrons are sent individually through the apparatus they still record an interference pattern on the screen, thus confounding the view that they behave, individually, like macro-particles. The individual particle goes into a superposition of states and travels through the slits in a superposed state. See **c**. **b** Particle-like behaviour of electron blast aimed at two slits. Depending on whether slit 1 or slit 2 is open, patterns I_1 or I_2 will form respectively. Alternatively, when a device is used to measure through which slit the electrons travelled, they will again display particle-like behaviour and form two peaks ($I_1 + I_2$), thus destroying the interference pattern. **c** Interference pattern observed in a laboratory. It should be noted that between 1959 and 1962, Jönsson (1974, 2005) carried out the real double-slit experiment with electrons and found the characteristic interference pattern

notations express ‘the entire weighted sum, with complex-number weighting factors, of all the possible alternatives that are open to the system’ (Penrose 1995, 259). For instance, quantum mechanics allows an electron to be at two different locations, A and B , simultaneously; these two alternatives are expressed in the quantum state $|\Psi\rangle$ as

$$|\Psi\rangle = \alpha|A\rangle + \beta|B\rangle \quad (2.9)$$

where α and β are complex numbers, which, when squared, express probability coefficients, which satisfy the condition $|\alpha|^2 + |\beta|^2 = 1$. The state $|\Psi\rangle$ is in a *linear superposition* of the two states $|A\rangle, |B\rangle$. The linear superposition of the two states indicates, according to the mathematics of quantum mechanics, that the quantum system is allowed to occupy these two states simultaneously. The Schrödinger equation then describes the deterministic evolution of such a system (in an abstract Hilbert space):

$$H\Psi = \frac{i\hbar\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi + V\Psi \quad (2.10)$$

where H is the Hamiltonian of the system and the expressions on the right express kinetic and potential energy, respectively. The Schrödinger equation is linear, which means that if two states, say $|A\rangle$ and $|B\rangle$, evolve after time t , to new states, say $|A'\rangle$ and $|B'\rangle$ respectively, according to the equation, then any superposition of these two states before t

$$\alpha|A\rangle + \beta|B\rangle \quad (2.11)$$

would, after time t , evolve to a new superposition

$$\alpha|A'\rangle + \beta|B'\rangle \quad (2.12)$$

which is equally linear. This whole evolution is deterministic and time-reversal invariant: the Schrödinger equation states the rate of change of the system with respect to time in the sense that the forward and time-reversed evolutions give rise to physically admissible processes. That is, if $\Psi(x, t)$ describes the evolution of a quantum system, then $\Psi^*(x, -t)$ also describes an evolution, where the time reversal operation is defined as $\Psi^T(x, t) = \Psi^*(x, -t)$, which means that the time-derivative operator $i\hbar\partial/\partial t$ in Eq. (2.10) is replaced by the expression $-i\hbar\partial/\partial t$.

Whilst the Schrödinger equation is time-symmetric in the specified sense, the basic irreversibility in quantum mechanics arises from the measurement process. Such measurement processes have an effect on interference patterns, which often serve to illustrate the behaviour of quantum-mechanical particles. The simplest way to observe the interference patterns is to consider the famous double-slit thought experiment, which normally supplies an example of the wave-particle duality involved in quantum mechanics. In this experiment a stream of electrons is targeted at a partition with two slits. When both slits are open the electrons will behave like waves and form characteristic interference patterns on the recording

screen (Fig. 2.24a, c). It is tempting to think, on an analogy with classical particles like sand grains, that half of the particles will travel through slit 1 and the other half will choose slit 2. But this way of thinking is refuted by a modification of the experimental set-up: instead of a stream of quantum particles, it is now individual particles, which are sent towards the screen. The surprising result is that the individual particles will slowly build up an interference pattern on the screen, in the same way as a stream of particles. The accepted explanation is that each individual electron goes into a superposition of states: $\Psi = \alpha|slit1\rangle + \beta|slit2\rangle$, where again α and β are complex numbers, whose squared moduli express the relative probabilities of the measurement outcomes. The interference pattern will disappear when an attempt is made to measure through which slits the electrons travel to the screen, for instance by shining a light on them. The interference effects, associated with wave-like behaviour will be destroyed, and the measurement screen will display two peaks (Fig. 2.24b), as they would be produced if the electron stream were replaced by a blast of sand particles.

The standard quantum mechanics of particles displays a dual character. There is on the one hand the linear evolution of the system in an abstract Hilbert space, encoded mathematically in the Schrödinger equation. It gives us information about the potentiality of the system or the number of possible states, the system is allowed to occupy, according to the rules of quantum mechanics. The system evolves deterministically according to the time-symmetric Schrödinger equation from one state to a later state or to a superposition of such states. The second fundamental aspect of quantum mechanics arises when measurements are made on such systems. In such cases the so-called ‘collapse’ of the wave function occurs. Any measurements, carried out on the system, ‘collapse’ the wave function from potentiality to actuality. The measurement process forces the wave function to abandon its options, expressed in the wave function Ψ or the state vector $|\Psi\rangle$, and reduces it to one observable state. In the measurement process the superposition of states ‘collapses’ to one actual, observable state. There is much uncertainty about the dynamics of this collapse, or whether it takes place at all. This question will be taken up in connection with the arrow of time; the present chapter will proceed to describe the state vector reduction, which leads to an ‘unsuperposed’ state of the quantum system under consideration.

This state reduction happens when a superposition, like $\alpha|A\rangle + \beta|B\rangle$, is faced with a measuring device, which is designed to register a particular outcome. Consider a photon in state $|A\rangle$, which impinges on a half-silvered mirror (Fig. 2.25). This action results in a superposition of states $|C\rangle, |B\rangle$. Part of the photon is transmitted, $|B\rangle$, whilst another part is reflected, $|C\rangle$.

The resulting state is a superposition of the transmitted and the reflected part:

$$|A\rangle \rightarrow |B\rangle + i|C\rangle \quad (2.13)$$

(where i is due to the phase shift that occurs between the two beams, indicating that $|C\rangle$ is reflected). What happens to this superposed state as it progresses through the

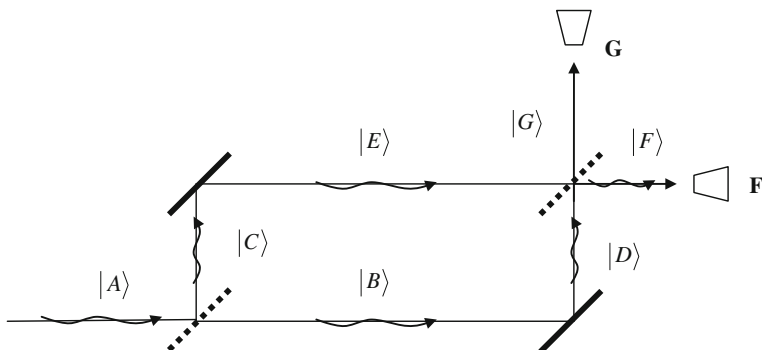


Fig. 2.25 Photon interference. The two photon parts interfere at the second half-silvered mirror, top right, such that only the transmitted part, $|F\rangle$, is registered, whilst the detector **G** does not register (adapted from Penrose 1995, 267; see Griffiths 2009, 120)

experimental apparatus? Both the $|C\rangle$ and $|B\rangle$ states will be reflected by silvered mirrors in their respective arms so that $|B\rangle$ evolves, according to the Schrödinger equation, into $i|D\rangle$ and $|C\rangle$ into $i|E\rangle$. After this evolution the superposed state $(i|D\rangle + i|iE\rangle) = i(|D\rangle + |E\rangle)$ is set up to encounter another half-silvered mirror, which brings the two beams together. The evolution is such that $|D\rangle \rightarrow |G\rangle + i|F\rangle$ and $|E\rangle \rightarrow |F\rangle + i|G\rangle$. At this stage the two photons interfere with each other in such a way that the entire state emerges in state $|F\rangle$ and is registered at detector **F**, whilst detector **G** receives no photon.¹³ Thus the two beams have come together to reduce the superposed state to just one actual registered state $|F\rangle$. The measurement of the superposed state results in the destruction of the interference effects. The measurement produces a basic asymmetry because the superposition is destroyed and the photon is registered in an unsuperposed state. Thus a contrast exists between the microscopic evolution of the system according to the time-symmetric Schrödinger equation, and the macroscopic asymmetry associated with the destruction of the interference effect and the emergence of one actual observable photon state. This contrast is the general situation in quantum systems:

What would happen in a more general situation such as when a superposed state like $w|F\rangle + z|G\rangle$ encounters these detectors? Our detectors are making a *measurement* to see whether the photon is in state $|F\rangle$ or in state $|G\rangle$. A quantum measurement has the effect of magnifying quantum events from the quantum to the classical level. At the quantum level, linear superpositions persist under the continuing action of U-evolution [unitary evolution]. However, as soon as the effects are magnified to the classical level, where they can be perceived as *actual* occurrences, then we do not find things to be in these strange complex-weighted combinations. What we *do* find, in this example, is that *either* the

¹³ Mathematically, the detection at **F** and the non-detection at **G** is explained because the entire state $i|D\rangle + |E\rangle$ evolves to $i(|G\rangle + i|F\rangle) - (i|G\rangle + |F\rangle) = i|G\rangle - |F\rangle - |F\rangle - i|G\rangle = -2|F\rangle$, where the factor ‘2’ has no physical significance (see Penrose 1995, 262; cf. Griffiths 2002, Sect. 26.2; Carroll 2010, Chap. 11) Physically, **G** does not register because interference effects cancel each other out.

detector at F registers *or* the detector at G registers, where these alternatives occur with certain probabilities. The quantum state seems mysteriously to have ‘jumped’ from one involving the superposition $w|F\rangle + z|G\rangle$ to one in which just $|F\rangle$ or $|G\rangle$ is involved. This ‘jumping’ of the description of the state of the system, from the superposed quantum-level to a description in which one or other of the classical-level alternatives take place, is called *state-vector reduction*, or *collapse of the wave function* [**R**]. (Penrose 1995, 263; bold and italics in original)

The algebra involved in the determination of the measured, observable state, after the collapse of the superposed state, indicates that reduction to a state of actuality is both regular and invariant, whether observers are moving at ordinary or relativistic speeds. Although they may disagree about the temporal ordering of the measurements when they move at relativistic velocities, this disagreement has no bearing on the outcome (Penrose 1995, 294; cf. Fn 8, Chap. III). Thus the observation of asymmetric irreversibility is not dependent on a particular perspective, which the observer may adopt. And its regularity is revealed in its predictability.

The measurement process in quantum mechanics is associated with a fundamental asymmetry, as is revealed in the destruction of interference effects. The asymmetry involved in the destruction of interference patterns is accompanied by an increase in entropy.

A consideration of the measurement of time has thus delivered several indicators, which seem to suggest an objective ‘flow’ of time—the expansion of the universe on a cosmic scale, entropic increase on a macroscopic scale and quantum–mechanical measurement processes on a microscopic scale. Those and other indicators of the nature of time shall be investigated in [Chaps. 3, 4](#). Meanwhile this chapter concludes with some remarks about the chosen focus on the measurement of time and on some permissible inferences regarding the nature of time.

2.11 Why Measurement?

Time is nothing by itself. Lucretius, *De Rerum Natura* I, 459–464

This chapter has mostly focussed on the association of time with cosmology because it illustrates two necessary requirements for the measurement of time: regularity and invariance. These two features accompany the story of temporal devices, from Galileo’s water clocks to Einstein’s light clocks. Even though clocks are ultimately atomic systems, the measurement process in quantum mechanics leads to an irreversible asymmetry. The measurement of time—in the sense of measuring a less periodic process by a more periodic process—pervades the sciences. Periodicity is not just a feature of physical systems; it is also an aspect of biological time. To illustrate the pervasiveness of periodicity consider some biological aspects 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 hour-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; Treisman 1999). 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 1 min 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.

It has been emphasized in this chapter that the measurement of time was at first closely associated with periodic cosmological processes. This association continues into modern physics. For a long time cosmological phenomena, like planetary motion, served as a basis for defining the day and the year, despite Newton’s misgivings about irregularities in the motion of the Earth. The identification of periodic patterns—the rotation of the Earth or lunar cycles—eventually gave birth to various kinds of clocks. Of particular importance for the measurement of time was the invention of the mechanical clock.

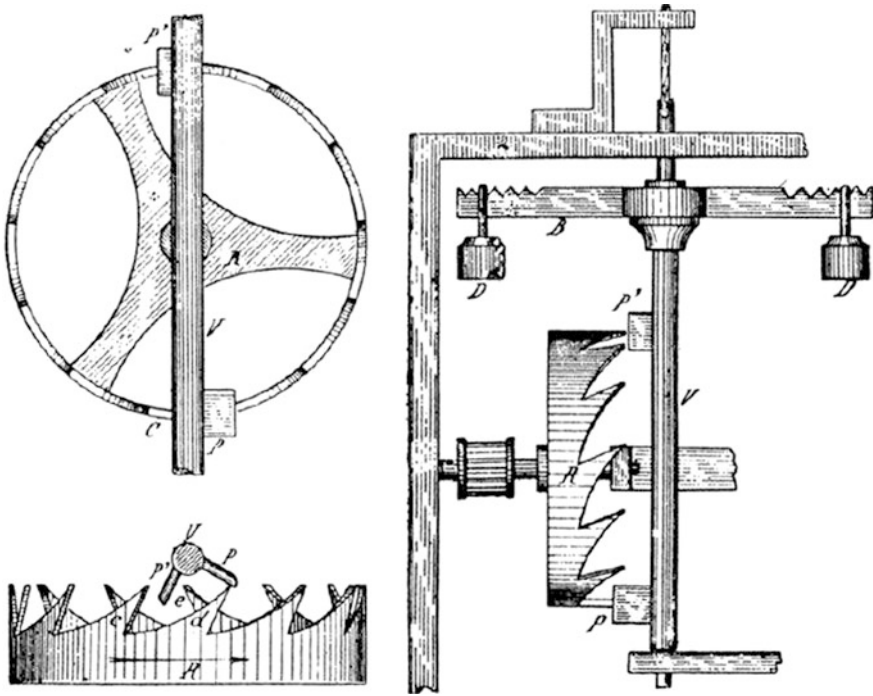


Fig. 2.26 Verge-and-foliot escapement, Wikimedia Commons

One of the most surprising facts about the history of the clock is that neither the inventor nor the exact date of the invention of the mechanical clock are known. But there is good reason to believe that the invention occurred between 1280 and 1300 A.D., somewhere in Europe since an extant document of 1271 forecasts the invention of a mechanical chronometer (Thorndike 1941; Dohrn van Rossum 1992, 89f). There is reason to believe that it was not the clock, which created an interest in time measurement in Europe; rather the interest in time measurement led to the invention of the mechanical clock (Landes 1983, 58; Dohrn van Rossum 1992, 90f; Wendorff 1985, 10, 85, 135 f, 142, 541; Whitrow 1989, 102f; 1980, Sect. 4.7). Several points are noteworthy about this invention:

- What was revolutionary about the mechanical clock was not that it employed mechanical parts. Water and astronomical clocks did so too. The crucial invention, which made the mechanical clock possible, was the 'verge' and 'foliot' escapement (Whitrow 1989, 103). A horizontal bar (foliot) was pivoted at its centre to a vertical rod (verge), on which two pallets hang (Fig. 2.26). These engaged with a toothed wheel (driven by a weight suspended from a drum), which pushed the verge one way and then the other, causing the foliot to oscillate. The wheel advanced, or 'escaped' by the space of one tooth for each to-and-fro oscillation of the foliot (Whitrow 1989, 103–104; Dohrn-van Rossum

1992, 52, Landes 1983, 7–8, 68). This innovation allowed a subdivision of the ‘flow’ of time into numerous equal, measurable parts (Wendorff 1985, 139; Landes 1983, 11). Furthermore, this mechanism made the clock independent of the elements: as the mechanical clock was weight-driven, it was impervious to frost (unlike the water clock) and could be used (unlike the sundial) through the night and on cloudy days (Landes 1983, 7). This clock allowed the measurement of a less periodic process—say, the flow of water—by a more periodic process—the oscillation of the foliot.

- The search for more precise oscillations eventually led to the replacement of the mechanical clock by atomic clocks, of which the first model was built in 1948 in the United States. In 1967 the first pulsar was discovered. Pulsars are fast rotating neutron stars, which emit electromagnetic radiation with such regularity, that in some cases the regularity of the pulsations is as precise as that of atomic clocks. Further refinements followed in the 1950s when the rotating Earth was displaced as the fundamental timekeeper by ephemeris time (1956), which provided a new definition of the second. Ephemeris time marks the transition from 1/86,400 of the mean solar day—a definition of mean solar second according to the daily rotation of the Earth—to 1/31,556,925.9747 of the tropical year in 1900—a definition according to the annual rotation of the Earth around the sun. Then, in 1967, the definition of the ‘second’ was finally detached from celestial clocks. A ‘second’ is now defined as 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom.
- The 1990s saw the introduction of GPS-time, and the sale of radio-controlled watches, which are accurate to 1 s in a million years. Modern time reckoning is still as much committed to cosmology as ancient Greek cosmology, although with a shift in emphasis. The General theory of relativity showed that clocks in gravitational fields suffer gravitational time dilation. The General theory is a theory of the large-scale structures of the universe, which, however, has certain limitations. For instance, it cannot explain the chaotic conditions, which govern the very earliest part of the universe. Modern quantum cosmology attempts to answer some of the questions, which are left open by the General theory. One such question concerns the famous ‘arrow’ of time. Any question regarding the temporal ‘orientation’ of the universe must be answered by turning to the area of quantum cosmology, which attempts to combine the General theory of relativity with quantum mechanics (Chap. 4). These areas are highly relevant to a study of the ‘march of time’.

The measurement of time supplies us with many clues as to the nature of time. Many prominent thinkers, throughout the ages, have relied on clock time to reveal the ‘nature’ of time. Furthermore, a focus on the measurement of time is neutral with respect to any preconceptions, which may be entertained with respect to the temporality or atemporality of the physical world. It is often asserted that quantum mechanics and the theory of relativity ‘prove’ that the physical world is a block universe, that only the Parmenidean view is compatible with our fundamental

theories. But we should investigate, before we accept this view, which inferences we are allowed to draw from scientific theories. The measurement of time involves both symmetric and asymmetric aspects, and both aspects are often used to draw inferences regarding the nature of time in the light of scientific discoveries. The question arises, which inferences we are allowed to draw from these various phenomena.

2.12 On Permissible Inferences from Scientific Theories

It is very common in the literature to see philosophical inferences about time drawn from scientific theories:

These two ideas of time, the unidirectional and the symmetrical, I have for brevity called “one-way” and “two-way” time. In going from the very simple science of mechanics to the very complex science of psychology, we must change from two-way to one-way time. (...)

The thesis (...) is that throughout the sciences of physics and chemistry, symmetrical or two-way time everywhere suffices. (Lewis 1930, 38)

But is it really the case that the block universe follows, say, from the Special and General theories of relativity and their central notions of relative simultaneity and time dilation, as many claim? (Weyl 1921; Gödel 1949; Lockwood 2005; Petkov 2005; see Weinert 2004, Chap. 4.) Is it really the case that a ‘symmetrical or two-way time’, with no trace of an objective passage of time can be inferred from many branches of science? In order to tackle these questions it is useful to distinguish several types of inferences, which are drawn from scientific theories.

- *Deductive* consequences follow from principle theories, like the theory of relativity and thermodynamics. According to Einstein, principle theories are based on fundamental postulates—like the principle of relativity, the light postulate or the Second Law of Thermodynamics—from which particular predictions can be derived, which are then tested against reality. Given Einstein’s principles, it follows deductively from the Special theory that for systems in motion mechanical clocks experience time dilation and rods suffer length contraction.
- *Inductive* consequences follow from statistical theories, like Darwin’s evolutionary theory or statistical mechanics. If, however, a theory is statistical in nature, then it is obvious that all its consequences will be probabilistic too. Statistical mechanics affirms that a system is more likely to evolve from ‘order’ to ‘disorder’ than in the opposite direction. But exceptions are compatible with the statistical regularity, so that the inference follows inductively.
- But neither *stasis* nor *flux* follow either deductively or inductively from the theories, with which these views are associated. Neither Parmenidean nor Heraclitean views are built into the principles of the respective theories. The theory of relativity and statistical mechanics do not include in their principles any reference to stasis or flux. Dynamic or static views of time are more

philosophical consequences of these theories. So a further type of consequences should be considered: *Conceptual* consequences follow neither deductively nor inductively from scientific theories but can be maintained with certain degrees of plausibility. That is to say that these conceptual consequences are more or less compatible with these theories whilst others are incompatible. For instance, a Newtonian notion of universal time would today be regarded as incompatible with the results of the theory of relativity.

Why should such issues be pursued? Firstly, it is important to clarify what inferences can be drawn from scientific theories as to the nature of time, space–time and the arrow of time. Secondly, the notion of time and associated notions, like the anisotropy of time, are of central importance to human existence. Thirdly, in the face of certain wide-ranging claims, which are taken to be direct consequences of modern space–time theories, it is important to consider quite generally, which claims can be regarded as plausible conceptual consequences.

One problem in deciding this issue is that scientific theories are complex structures, which consists of a cluster of different components: mathematical formalisms and models—which ensure a representation of the formalism and built a bridge between the mathematical abstraction and the physical reality—as well as predictions and observational statements, and philosophical consequences. What the models represent are structural features of the systems under consideration. The physical world consists of various systems, of various degrees of complexity, and the structural models represent these structures. A system consists of a number of components—called *relata*—which are bound into a system by the relations, which hold them together. The relations are typically encoded in the laws of nature. The solar system is such a system, consisting of planets, satellites and moons, orbiting a central sun according to the laws spelt out by Kepler and Newton. According to Newton’s explanation, the planets’ orbits are the vectorial result of the first law of motion and the law of gravity acting together to keep the planets in their orbits. A structural model of this system then represents the structural features of a planetary system. There is no claim that every bit of the model corresponds to some feature of reality. The claim is only that the model structure fits the physical structure of the system modeled. Fit is confirmed by observations, measurements and the corroboration of predictions. The conceptual consequences of the theories, in which models are embedded, are not necessarily confirmed by the fit between model structure and physical structure. There are debates—*asymmetry versus symmetry*, *stasis versus flux*—which are not subject to empirical tests, but subject to plausibility arguments.

But if that is the case the problem of realism arises. Should we infer, for instance, from the impressive predictive success of the Special theory of relativity that the underlying model of four-dimensional space–time can claim existence too? And hence that the passage of time is a human illusion! There have certainly been numerous arguments from the success of such theories to ontological claims about the reality of space–time. For instance V. Petkov has claimed that none of the observable phenomena of the Special theory would be possible if space–time did

not exist. And a static space–time implies the unreality of time. For Petkov unobservable space–time exists in its own right as much as the observable predictable consequences of the Special theory. Such authors go even further and claim that the Special theory implies fatalism and the denial of free will (Petkov 2005; cf. Lockwood 2005). Others have used the block universe to claim that time travel into the past is possible, despite the paradoxes it involves (Lewis 1976).

The problem is that these authors treat claims about the ontology of space–time and the unreality of time as deductive consequences from the theory of relativity when they are in fact conceptual consequences. As shall be explored in [Chap. 3](#) alternative models of the Special theory have been proposed, which do not lead to these consequences.

Such conceptual consequences, even if they are only subject to plausibility tests, are worth pursuing because scientific theories do include such less testable claims. Scientists regularly make claims about the nature of time, based on an assessment of how compatible such conceptual consequences are with the established facts and theories. One such claim is that time in the physical world does not exist and that consequently both the passage and the arrow of time are an illusion.

It is precisely because of the prevalence of such claims that their compatibility or incompatibility with scientific theories should be investigated. Human life is intimately related with questions about the nature of time. Whilst in the past Greek metaphysics provided answers to such questions, this task has now fallen to science and philosophy. Science nowadays is frequently associated with worldviews. It is often the case that the results of scientific theories—in particular the results of the Special theory: time dilation, length contraction and relative simultaneity; quantum theory: especially the symmetry of its fundamental equations and thermodynamics: the symmetry of probabilities with respect to time—are presented as evidence for the conceptual consequences, especially the atemporality of the physical world. This chapter, however, has focussed on considerations of the measurement of time, for several reasons. The measurement of time is a ubiquitous phenomenon, which touches on all the central issues involved in the discussions of the nature of time. The starting point of the measurement of time allows a survey of the relevant issues, without prejudice for one particular view. The decision is ultimately not one of pure empirical testing. The conceptual consequences are simply not material for refutation. The issue is one of degrees of compatibility of the amount of empirical data with the two prevailing views: flux or stasis, symmetry or asymmetry.

In this chapter we adumbrated these ideas from the point of view of the measurement of time and its two requirements. In the following two chapters we shall consider the conceptual, philosophical consequences of our physical theories in greater detail. [Chapter 3](#) will focus on the opposition between flux and stasis, as it arises in the theory of relativity. [Chapter 4](#) will consider the opposition between the symmetry of the fundamental laws and the asymmetry of the observable world around us, as it arises in thermodynamics, statistical mechanics and quantum cosmology. When it comes to the notion of time and associated notions, the question ultimately boils down to which criteria to choose to interpret the nature of

time. If we choose t -invariant theories and relative simultaneity, we are tempted to opt for stasis; if we choose asymmetric physical processes and invariant relationships, we are tempted to opt for flux (cf. Denbigh 1981, Chap. 4, Sect. 5). The author will not hide from the reader that an examination of these issues has persuaded him of a Heraclitean, temporal view of the physical world. But this conviction seems at present a minority view in need of a justification and it is worth investigating the reasons why we should favour flux over stasis and asymmetry over symmetry.

The March of Time

Evolving Conceptions of Time in the Light of Scientific
Discoveries

Weinert, F.

2013, IX, 284 p., Hardcover

ISBN: 978-3-642-35346-8