Chapter 1

Days, months and years

We still do not fully understand exactly how and when Earth and Moon formed in the tumultuous youth of our solar system. A widely accepted theory [1] says that about 4.5 billion years ago, a Mars-sized planetoid collided with Earth and that our Moon is the product of that collision. Presumably, details of Earth's and Moon's orbits are the result of that violent encounter. After things calmed down, still in Earth's early history, stable patterns of motion became dominant. Since then, the ensuing repetitive and nearly constant changes between day and night, between lunar phases, and between seasons have provided powerful natural rhythms that have deeply influenced the evolution of life in general and human development in particular. Obvious manifestations are the myriad biological clocks found in plants and animals alike[1]. Bioregulators may be tuned to 24-hour or circadian cycles, the lunar cycle (for example via the tides), seasonal changes (witness the blooming of asters), or complex combinations (such as the lifecycle of the 17-year cicadas). The natural cadences of time also seem to be reflected in several early manmade structures. The Newgrange (Irish: *Si an Bhru*) mound in County Meath, Ireland, the Stonehenge circles in Wiltshire, UK, or the Kukulkan temple in Chichen Itza on the Yucatan peninsula of Mexico are a few examples. Today we can still experience what people must have felt hundreds and even thousands of years ago when the solstice or equinox Sun aligns with these impressive constructions. While speculations and even fabrications are not uncommon in connection with ancient sites such as these, the influence of astronomically observations on their layout is quite certain. The regularity and reliability with which the Sun marks the same extreme points on the horizon during the course of a year is at the heart of these remarkable edifices. There is also evidence that structures were built in antiquity that aligned with either planets or stars, the other objects that move in regular patterns across the sky [2]. Their influence on earthly events is certainly not of the type that astrology

[1] In this book, I will bypass entirely this fascinating and still not fully understood type of clock.

doi:10.1088/978-1-6817-4096-6ch1

wants us to believe, but the impact of the stars on human thinking, and in particular our thinking about time, is undeniable and strong. One of the oldest artifacts depicting celestial objects is a bronze disc that was found in Nebra, Germany (German: *Sternscheibe von Nebra*), with an estimated age of around 3500 years. On that disc there is, next to the prominent crescent Moon, an arrangement of dots thought to represent the star cluster of the Pleiades in the constellation of Taurus (see figure 1.1).

Well before material artifacts were made that imitated the configuration of the lights in the sky or marked certain aspects of their motion, the rhythms of the seasons—whether directly by their manifestation in the form of dry and wet or warm and cold periods, or indirectly by patterns of animal migration or plant growth—must have left lasting impressions on the human mind. The relentlessness and regularity of that change, even if accompanied by variations and fluctuations, is a powerful indication that something must be 'out there' to keep this rhythm alive. What people across the world and through the ages believed to be the moving agent is a long and fascinating story, but not the one being told here. What we will dwell on is the contemporary version of this story, a version that combines many previous strands into a coherent saga of the Universe. This new tale tells of white dwarfs and red giants, of black voids swallowing everything venturing too close to their edges, and of a mysterious dark substance guiding the ebb and flow of things. It contains chronicles of collapsed stars that spin around their own axis like furious dervishes of old, but with a precision rivaling that of modern atomic clocks. Regarding our closer neighborhood, the saga contains the residue of ancient mythologies letting gods and

Figure 1.1. Bronze disk, ca 3500 years old, showing waning Moon, Sun and stars—possibly including the Pleiades (between the Sun and Moon). This image has been obtained by the author from the Wikimedia website where it was made available [by name of uploader] under a CC BY-SA 3.0 licence. It is included within this article on that basis. It is attributed to Anagoria.

goddesses of old live on. In their modern version, we can actually see the faces of Jupiter, Saturn, Mars, Venus and Mercury. We have even touched them and have sent messengers to explore their smallest wrinkles, to find out what stuff they are made of and how they became what they are today. In the old stories, the 'dance of the planets' used to be a stately, highly controlled affair confined to our solar system. The new epic reveals that billions of such systems exist whose choreography follows a few basic rules. Nevertheless, each family of planets with its central star—or sometimes two or even more—is different and unique. Our present saga of the Universe is incredibly wide in scope, rich in detail, and exact in its utterances. We can calculate the paths of the planets or artificial satellites with extraordinary precision. We can do that because we have discovered that general and fundamentally simple laws determine the paths and because we have found ways to turn that knowledge into specific predictions of the future. We have come a long way since the Oracle of Delphi.

Gravity is the force that governs motion in the Universe and one of the fundamental laws alluded to above specifies that the strength of gravity falls off with the square of the distance. As a consequence of this law, the path of any two gravitationally interacting objects belongs to the geometric class of conic sections, i.e. it is a curve generated by the intersection of a cone and a plane. Depending on the angle between plane and cone axis, the cut generates a hyperbola, parabola, ellipse or circle. For example, the rotation of planets around their host star occurs in elliptical orbits, often only slightly eccentric, i.e. almost indistinguishable from a circle. Historically, the argument went in the opposite direction: careful observational data (Tycho Brahe) were used to deduce that planets move in elliptical orbits (Johannes Kepler), which in turn led to the discovery of the gravitational force law (Isaac Newton). In any case, all the natural rhythms are consequences of this simple relationship between force law and the geometry of orbits, i.e. between dynamics and shape.

Despite the simplicity of the law governing it, the motion of Sun, Moon and planets also exhibits subtle complexities that for centuries have befuddled both casual and careful observers of the sky. In the end, though, the root cause for this complication turned out to be straightforward and reducible to three major contributions: more than one rotation is involved, the various rotations do not occur in a single plane and the ratios of their respective periods are non-integer. All this gives rise, for example, to seasonal changes of the length of the day, the need for leap years, the difficulty to fit lunar months into years, and a somewhat surprising abundance of definitions of what a day and a year really are. On a finer scale, there are both long- and short-term variations of the basic periods, which have produced further difficulties for anyone trying to come up with a comprehensive model of the motion of the heavenly bodies. The additional variations are caused by the gravitational influence of celestial objects other than the Sun, or the vagaries of the rotation dynamics of Earth's core, to name just a few causes. An example of the former effect is the slow motion of the celestial pole, while the recent practice of inserting leap seconds into our atomic clock-controlled time-keeping is a response to the latter—at least in part. Around the start of the current millennium, a large number of books were published on this very topic—how our calendar, how our clocks, how our sense of time is informed by both natural, objective causes and by social and historical circumstances. While the latter, intriguing aspect is yet another issue that is not pursued in this short book, the former will be. So, how do we describe the celestial clock from our modern perspective? In a moment, we will begin to look into this question. First, we need to introduce some specific terms and units of measurement to characterize the distance between two objects in space, their orientation relative to a given direction and/or plane, and the length of time elapsed between two events.

1.1 A first discourse on measurement, units and precision

At the time of writing, only three countries[2] in the world have not yet officially adopted the metric system, which I will use throughout this book. These decimal-based units are also known as SI units (the acronym stands for the French *Système Internationale*). In particular, we will employ the SI base units of length (meter, unit symbol 'm'), angle (radian, 'rad'), and time (second, 's'). Multiplicative scaling of the base units by factors of 1000 or 1/1000 and attaching associated prefixes (mega-, kilo-, milli-, micro-, etc) provides an expedient way to extend the units to larger and smaller ranges (see also table 3.1). However, in the context of time, additional non-metric units are convenient and commonly used, most notably the units of years, weeks, days, hours and minutes. Likewise, the vast distances encountered in astronomy are usually indicated in non-metric units, such as the astronomical unit, the parsec and the light year.

To measure is to compare with a standard. For example, the distance between two points is determined—at least in principle—by placing an agreed upon standard, say a meter stick, end to end between the two points as many times as is necessary. Clearly, this recipe fails if you want to know the distance to the Moon, the Sun, or any star or galaxy. Whether we are able to reach them directly or not, what can be measured for any two points is their angular separation. Take two stars in the night sky or two points on opposite sides of the Moon's equator. Independent of their distance from us, the two lines of sight connecting an observer with the two points subtend a well-defined angle that can be measured with a simple mechanical instrument such as the sextant. Conversely, two angles specify completely and uniquely the location of any one point in the sky. As with any other definition, we have to choose a reference frame in which the locating angles are to be quantified. One straightforward choice is the local horizon augmented by the four cardinal directions (north, west, south and east). Next, draw an imaginary line, the celestial meridian, in the sky that connects the zenith with any of the four cardinal points on the horizon—say south. Then the first angle measures the vertical angular height of the star above the horizon and the second the horizontal angular separation from the meridian. The first angle ranges from 0° for a point at the horizon to 90° for any object located at the zenith. Stars and anything else on the celestial hemisphere

[2] Liberia, Myanmar and the USA.

below the horizon will be associated with negative angles ranging up to −90° for the nadir. The values of the second angle are anywhere between 0° and 360°. For completeness, we also have to specify the direction in which the second angle is being determined along the horizon—say in the east to west direction. This now allows an unambiguous specification of the location of any point in the sky. Of course, the angular values of an actual object depend on the place (and the time) at which the observation is made and observers at different locations will have to convert their specifications. Therefore, astronomers have adopted a 'standard' location from which celestial navigation is conducted. Locally measured angles are routinely transferred to this 'global' reference frame. The discussion of what exactly this reference frame is can wait until we describe the structure and dynamics of our solar system in more detail. Incidentally, position on the surface of the Earth is determined by a pair of angles that are completely analogous to that used to pinpoint locations on the celestial sphere. In the case of terrestrial navigation, the two angles are called latitude and longitude. The former is the angular distance from the equator along the north–south direction (with positive values between 0 and 90° for the northern hemisphere and negative values for the southern hemisphere). Longitude is the angular distance in the east-to-west sense from a reference arc that joins the north and south poles, and which by convention passes through a certain point at the Royal Observatory in Greenwich, UK.

The division of the circle into 360 parts or degrees goes back to the Babylonians and the similarity to the numbers of days in the year is probably not a coincidence. In the SI unit system, an alternative way to express a given angle gives rise to the unit of the *radian*. Since the circumference of a circle of radius r is $2\pi r$, the ratio between the circumference and radius is 2π, which is equivalent to the full angle of 360°. Any smaller angle θ given in degree can be specified by the ratio of the corresponding arc to the radius. The value of this ratio is the magnitude of the angle θ in units of radians. For example, an angle of 60° is the same as 1/6 of the full angle, i.e. $\pi/3$ rad. The metric base unit 1 rad corresponds to an angle of $(360/2\pi)° \approx 57.3°$. Much of the dynamics we encounter in our upcoming discussion is rotational motion—as opposed to movement along straight lines. If the center of the path lies outside the moving body, we speak of an orbit (like that of a satellite about the Earth). If the spin occurs around an axis that passes through the object itself, we refer to that motion as rotation. In either case, a full completion is identical to a change of 360° (or 2π) in angular position of any given point of the body.

Without further ado, set seven days as one week, one day (unit symbol: d) as equal to 24 hours (h), let one hour be exactly 60 minutes (min) long and have each minute contain 60 seconds (s). As already alluded to above, conversion between the time units of year and day is non-trivial and depends on the exact definition of the two quantities. Let us define for now the length of a day as the average time span between two subsequent moments of the Sun being at its highest point in the sky, i.e. from noon to the next noon. Likewise, we adopt the notion that a year is the average time span between two subsequent summer solstices, i.e. moments of the setting Sun crossing the horizon at the most northerly point. Then approximately 365 ¼ days fit into one year. In general, science insists on precise terminology and the preliminary definitions above of the day and the year are two examples. This insistence on precision might be perceived as needless, ('we all know what a day is'), if not outright annoying. But it turns out to be necessary. On closer inspection things are often more subtle or more complex than expected. For example, the two definitions just mentioned include the word 'average', which seems superfluous, yet it is not. As we will see soon, the length of the day and the year is not constant.

With the above definitions, we also have a first set of conversions between time units, which we can write succinctly as follows:

$1\text{d} = 24\,\text{h} = 24\cdot 60\,\text{min} = 1440\,\text{min} = 1440\cdot 60\,\text{s} = 86\,400\,\text{s}.$
$1\text{a} \approx 365.25\,\text{d} = 8766\,\text{h} = 31\,557\,600\,\text{s}$ (note that the precise value is $1\text{a} = 31\,556\,925\,\text{s}$)

In other words, even with a daily 8-hour beauty sleep you can while away almost 60 000 hours in a year or conduct about 2 million back-to-back countdowns—your choice. In order to get a feeling for the implied precision when the specification of the duration of one year is uncertain by one second, it is helpful to consider the precision a clock needs to have to be capable of detecting this last second in the year. Suppose our clock ticks at one-second intervals. Since there are about 30 million seconds in one year, each clock period needs to repeat with a precision that is better than one part in 30 million. While this is no challenge at all for atomic clocks (see chapter 3), only the best mechanical clocks (see chapter 2) can accomplish such a feat.

At present, the basic unit of time is the second, defined as the period of the light wave emitted by cesium (Cs) atoms under certain well-specified conditions. Historically, the base unit of time was the day, defined by astronomical observations, and hours, minutes and seconds were derived quantities. When the atom-based definition of the second was introduced, it was chosen so as to preserve the actual length of the corresponding time spans. This is accomplished with the second fixed as the duration of a bit more than 9 billion (to be precise, 9 192 631 770) oscillations of the cesium atom radiation. In essence, the two sets of equations above are the definitions of the longer units of minute, hour, day and year. It will be useful for our upcoming discussions to imagine that we have available a perfect stopwatch that ticks at a uniform rate and allows us to determine the time span between two events with arbitrary precision. This notion of measuring time spans with clocks appeals to 'common sense', but actually needs refinement, which we will provide in chapter 4 (on relativity). For now, it will suffice.

Adopted worldwide, the Gregorian calendar gives another answer to the question of how many days there are in a year by defining one ('regular') year to be exactly 365 days and adding one day for leap years. Leap years are those whose number is divisible by four, but omitting three out of four century years, namely those that are not divisible by 400 (e.g. the year 1900 was *not* a leap year, but 2000 was). Thus, the Gregorian calendar has a repeat length of 400 years. While certainly not the only one, the Gregorian calendar is a useful and relatively simple tool that comes to terms with the conundrum of incomensurability and lets us organize the flow of time into a pattern of countable units, keeping in sync with the seasons—at least for the next few thousand years. Following the Gregorian calendar, we can calculate the average number of days in one year in the following way. First add up all the days in 400 years if no leap years existed ($400 \cdot 365 = 146\,000$). Then add one day for each leap year, disregarding the century rule (+100). Now subtract one day for all centuries not counted as leap year (−3) for a grand total of 146 097 days in 400 years. Thus, there are 365.2425 days on average in one Gregorian calendar year. Compare this to the current best value of 365.242 189 days in one tropical year, which is the time span to complete one cycle of the four seasons: the annual difference of 0.000 311 d or about 27 s accumulates to a slip of one day after $1/0.000\,311 \approx 3215$ years. Not bad at all.

Is a calendar, Gregorian or otherwise, a clock? In other words, does it use a 'device' to mark certain basic time intervals, and does it then indicate how many of those units have elapsed between two events? Maybe it is stretching this definition a bit, but I think one can argue that calendars are indeed some sort of clock. The time-beating core is simply the Sun's journey in the sky—which of course means Earth's rotation coupled with its orbit around the Sun. Measuring the elapsed time is accomplished by counting days 'labeled' in a unique manner, for example in the Gregorian calendar by year, month and day, or in the Mayan long count by providing a combination of numbers for the teeth of interlocking gears. Calendar systems have a repeat length, after which the same combination recurs. In this sense, all calendars reflect a cyclic view of time. One long, repeating count in the Mayan calendar is equivalent to 1 366 560 days or about 3741 years, which is similar to that of the Gregorian calendar. For the latter, we have seen that it slips by one day after about 3215 years, which came about because of a mismatch of about 27 s to the average time it takes Earth to complete a trip around the Sun. One might say, then, that the Gregorian calendar, viewed as a clock, runs untrue by about that short time span in one year when compared to the time told by the Sun–Earth system itself. Just to put this into perspective: state of the art atomic clocks perform at such an astonishing level of precision that it takes them millions of years to be off by one second. In chapter 3 we will see how to quantify clock precision with the help of the so-called Allan variance, which, amusingly, also permits a loose comparison of time pieces as diverse as atomic clocks and megalithic monuments such as Stonehenge or Newgrange, which themselves might be viewed as calendar-clocks marking off units of years.

1.2 What we see in the sky—stars, planets, Sun and Moon

Whether via direct or scattered rays, our Sun causes most natural light effects, such as red sunsets, green flashes, blue sky or the whole color spectrum of the rainbow, and the light reflected by our Moon and the planets. There are a few exceptions, including lightning bolts, the occasional meteor or comet and, of course, the fixed stars. Lightning, comets and meteors are light shows that people probably felt more frightened of than entertained by. Although quite frequent, these light effects lack the precision and regularity to contribute much to our story of natural timekeepers—notwithstanding the (near) periodicity with which a number of meteor showers peak in intensity. It is the dance of the Sun itself and of the Moon and the stars, whether fixed or wandering, that had a profound influence on our understanding of the world in general and of time in particular—and our ways to measure it. These lights in the sky have intrigued people for a long time, and not only because of their practical utility. How could it not be so? Imagine that you are living in Egypt, four or five thousand years ago. It is night. Some lingering kitchen fires disturb the darkness, but just a short walk outside the village the sky is pitch dark. The bright splendor of the stars in the night sky must have been mesmerizing—it still is.

Stay awake one night and you are treated to a slow-moving light show (provided you can find a sufficiently dark place): in unison and in an east-to-west direction, the stars complete part of grand circles around a common point in the sky, the celestial pole[3]. Even the fixed stars are not fixed at all. At dawn, they become invisible against the Sunlight scattered by the atmosphere. However, come dusk, the stars reappear, one by one, at the points where a straightforward extrapolation of last night's motion would predict them to be. In other words, the fixed stars rotate through the sky in full circles[4] with a common center and with their positions relative to each other unchanged. This twirl is prosaically called the diurnal stellar motion. During daytime, the Sun completes an arc in the same direction and around the same center. After centuries of fits and starts, we understand that the motion of Sun and stars across the sky is caused by Earth's rotation around its own axis, hence the common center. Earth's spin axis points to the celestial pole. The latitude of any place on Earth determines the declination of the celestial pole. Should you be at the north or south poles, the celestial pole is exactly overhead. Stargazers at the equator can easily infer that there are two celestial poles—one in the north and the other in the south, both near the horizon. If you are more patient and pay attention to the sky all year long, you will further see that the particular star constellation that rises above the horizon in a given direction changes over the course of one year. After that, the same sequence repeats.

Much more restless are the wandering stars or planets (Greek: $\pi\lambda\alpha\nu\eta\tau\eta\sigma$ (wanderer)). Mercury, Venus, Mars, Jupiter and Saturn are visible to the unaided eye and thus they were already known in antiquity. They are not part of any stellar constellation, they participate only approximately in the diurnal stellar motion and their relative motion is not simple. Relative to the fixed stars, they speed up, slow down and sometimes even reverse direction in so-called retrograde motion (see figure 1.2). While Mars, Jupiter and Saturn range freely across the sky—albeit confined to a narrow band around the ecliptic—Venus and Mercury never stray far from the Sun, and sometimes disappear behind it. Probably because of the tight connection to the Sun and its bright appearance as morning and evening star, Venus had a special significance in Mayan culture. Astronomer–priests observed the planet closely. The Dresden codex, one of only four known books written by the Maya, very likely in the Chichen Itza region [3], contains details of Venus' motion in the

[3] With a circumpolar path of radius less than 1°, Polaris, the north star, is very close to the celestial pole on the northern hemisphere. No visible star is near the southern celestial pole.

[4] During mid-winter, when the Sun does not rise above the horizon north or south of the polar circles, you can actually observe the full circle of the stars.

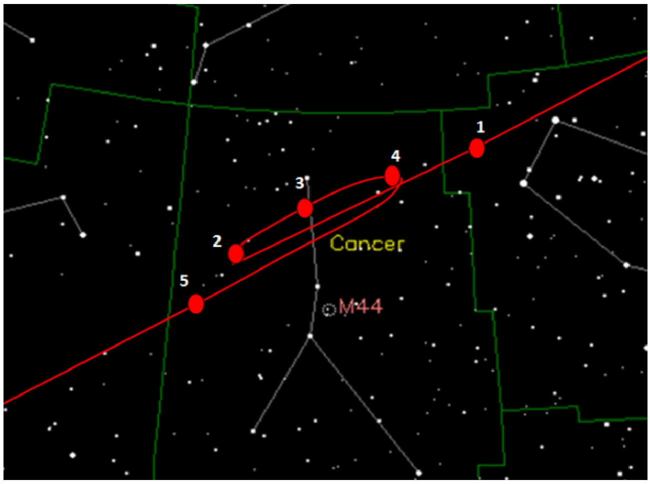

Figure 1.2. Observed retrograde motion of Mars (points 2–4) relative to the constellation Cancer in 2009–10. Five sequential observations of Mars' position in the sky are shown by the numbered dots.

sky, including its synodic period of 584 days (the current best value is about 583.92 d). The synodic period of a planet marks its return to the same spot in the sky and ranges from 116 days for Mercury to 780 days for Mars. For the other planets, this interval is much closer to the length of the Earth year, ranging from 399 days for Jupiter to 368 days for Neptune.

For much of human history and across cultures all over the world, the brightest light in the sky—our Sun—has been at the center of both worship and observation. It does not take much sophisticated effort to realize that the Sun's path changes significantly over the course of a year. First, the daily east-to-west arc varies with the seasons in length and height above the horizon. In doing so, the high point of the arc moves up (summer) and down (winter) along a skewed, slender figure of eight called the Sun's *Analemma*, whose slant and shape varies with the observer's latitude. Secondly, the arc changes position relative to the background of the stars. Imagine for a moment (as I alluded to in the preface, we will do a fair amount of imagining) the Earth had lost its atmosphere, which would clearly be a nuisance. Adding insult to injury, there would no longer be any beautiful sunsets or rainbows, and the sky would always be black. The latter aspect, however, would have the advantage of the stars being visible 24 hours a day; the diurnal motion would be in full display even when the Sun was above the horizon. At any given moment, it would be obvious which starry constellation was in closest proximity to the Sun, as we would be able to see the Sun and stars moving along their respective circles in synchrony all day long. That is to say in approximate synchrony, because every day the Sun lags a tiny bit behind the motion of the stars. Relative to the stars, the Sun moves eastward. Thus, the stellar backdrop against which the Sun is seen changes during the course of a year. After one year, though, the cycle repeats. Luckily, we still have oxygen to breathe. Therefore, the moments around sunrise and sunset are fleeting when we can catch a glimpse of a few stars and Sun together. However, since the pattern of the fixed stars is indeed fixed, we can still figure out 'in which house' (as astrology likes to call it) the Sun presently resides. The apparent year-long path of the Sun in the sky against the background of stars defines a plane in the sky, called the *ecliptic*, whose axis is tilted by about 23.5° against the axis of the diurnal motion. During our ride on the merry-go-round we call the solar system, when we look towards the center where the Sun is, the far distant background scene of stars is constantly changing as we complete one circle.

If you are very patient and pay attention to small details, you can also see something very subtle—the celestial pole moves relative to the stars. It shifts ever so slightly—and with it the point on the horizon where stars rise or set that are sufficiently far from the pole. Earth's spin axis describes a circular path against the fixed stars and it takes, in round numbers, 26 000 years for one completion. This phenomenon—equivalent to that of a wobbling top—is primarily a result of the gravitational tug by the Sun and Moon on the Earth's bulge around the equator. As a result, both the north and south celestial poles wander by about 1.4° every century. While not all that much, it still amounts to almost three times the angular size of the Moon. Astronomers refer to the long-term celestial pole motion as 'axial precession' or 'precession of the equinoxes', a term introduced because the shift is observable as the westward movement of the spring and autumn equinoxes along the ecliptic relative to the fixed stars. Because of Earth's pole motion, which stars if any are closest to the celestial poles changes over time. Currently, we find the star Polaris (aka α-Ursae Minoris) closest to the North celestial pole. When the last glacial period was still in full swing about 14 000 years ago and you looked up in the sky to find your way home at night, the brightest star near the pole would have been Vega in the constellation Lyra [4].

And then there is the Moon. Its angular size, quite accidentally, is very nearly the same as that of the Sun. As beautiful as the stars and the Milky Way are, our Moon outshines them all, albeit with secondhand light. Waxing, waning and distinct features in the face of the Moon add to the allure. No wonder tales about the Moon and, importantly in our context, attempts to build a lunar calendar can be found in many cultures. The time span between two full Moons (or any other lunar phase) is called the synodic month. On average, it lasts roughly 29.5 days but varies by as much as 18 hours during the year. Since about 12 ½ lunar cycles occur in one solar year, lunar-based calendars are very difficult to synchronize with the solar rhythm.

The Sun and Moon follow independent paths and the Moon's arc across the sky is slanted relative to the ecliptic, thus allowing for crossings. Occasionally the Moon indeed moves in front of the Sun, obscuring it partially or even fully. On the other side of its orbit, the Moon can dive into the shadow cast by the Earth, causing a lunar eclipse. Solar and lunar eclipses occur not infrequently but they lack simple periodicity. Nevertheless, attempts to predict, for example, total solar eclipses started surprisingly early, even if initially they were not blessed with much success. As early as 2300 BCE, the Chinese emperor apparently expected solar eclipses to be foretold. At least that is how it appears, since two astronomers at the imperial court were beheaded for failing to predict the eclipse of that year [5]. Despite this strong disincentive to pursue the issue, by the 4th century BCE Chinese astronomers had learned to predict solar eclipses by observations of the relative position of the Sun and Moon. As we will see later in the book, in modern times a solar eclipse has played an important role in our advancement of the understanding of time. Yes, you guessed it, I am referring to the solar eclipse of 1918 confirming Einstein's prediction of light bending by gravity. In chapter 4, we will discuss another wrinkle concerning the Moon's role in the context of general relativity.

Associated with the motion of the Sun along the ecliptic is a very small, but nowadays easily measurable, variation of the angular size of the solar disc. In January, the Sun's angular diameter is found to be slightly more than 3% larger than it is in July. Let us discard the highly unlikely hypothesis that the Sun cyclically changes its actual size in response to or at least in synchrony with the seasons on Earth. Then the periodic change in angular size hints strongly at a small change in the distance between the Sun and the Earth—a little closer than average during the winter in the northern hemisphere, a little further away in the summer. By the way, appearances notwithstanding, there is *no* measurable change in the angular size of either the Sun or the Moon during the course of one day. They do *not* increase with approach to the horizon, although refraction in the atmosphere does distort the image.

Let us now return to the question of how the natural units of the day, the year, or the synodic month arise by looking at the make-up of the solar system, the ultimate mechanical clock. It may appear as a useful and probably necessary detour to describe the spatial properties first in order to discuss the temporal ones—but it is a detour nonetheless. However, this seemingly indirect route already hints at a deep connection between time and space, as fully revealed only in Einstein's relativity theories (see chapter 4).

1.3 What the solar system looks like

Once upon a time, we thought that we had figured it out. Earth is at the center of the Universe with the Sun, Moon, a few planets (the wandering stars) and the fixed stars circling around. The various celestial light sources are attached to 'crystalline' spheres[5] that are nestled like spherical Russian dolls and that glide smoothly around each other in simple harmonic motion at their own unchanging speed. The Moon is on the sphere closest to Earth, and then come Mercury, Venus and the Sun, followed by Mars, Jupiter and Saturn, and finally there is the firmament with all the fixed stars. This, in a nutshell, is the cosmological model that was developed in ancient Greece and survived in one form or another for more than a thousand years. The model is efficient, as it reproduces all of the salient observations with just a few explanatory ingredients. In addition, it is flexible enough to accommodate, with some tinkering, additional details such as the retrograde motion of the planets. Single spheres, rotating around the Earth with constant angular speed, just cannot explain it. However, if the planets are allowed to move in smaller spheres attached to

[5] To modern ears it sounds strange and naïve to have whole planets suspended by such a fragile substance as quartz. But that is in hindsight—the early view of the 'lights in the sky' was just that: lights in the sky. One can argue—I think convincingly (see e.g. [6])—that the geocentric model is an early scientific attempt to come to terms with observation, 'explaining' nature without recourse to supernatural powers.

the large sphere that carries the planet (so-called epicycles), then the observed motion can be reproduced. Adding more and more epicycles, at the cost of its original simplicity and elegance, the evolving geocentric model was able to duplicate all relevant observations as late as the 16th century. In addition, the model has the benefit of being consistent with the immediate impression of our senses, telling us that we are standing still and that the Sun, Moon and stars are moving around us. It was a truly revolutionary and at first fiercely resisted insight that the observed motion of the Sun is only apparent and that it is the Earth that swings around a (nearly) stationary Sun.

Here is what we now know about our solar system and the stars. Not Earth but the Sun is at the center of the dance. Earth is but one of a huge number of objects orbiting Sol, our central star. The Sun dominates the total mass of the system: the estimated contribution to the total mass by all other objects is only about 0.14%, in other words the solar mass is about 700 times larger than the mass of everything else in the solar system combined. We can classify these other objects according to size (or mass) and orbital shape into categories (asteroids, comets, planets, etc) whose boundaries are not entirely fixed—witness the demotion of Pluto. Planets are the objects found in the inner part of the solar system that are massive enough to be round due to their own gravitational force as well as to be able to sweep clear the vicinity of their orbit. They move, all in the same directions, around the Sun in near circles with radii that differ by a factor of almost 80. In terms of the Earth–Sun distance, which we introduce here as one astronomical unit or 1 au, the distance of the innermost planet Mercury is about 0.39 au, while that for Neptune, the outermost planet, is about 20 au. In the context of this book, the planets and their satellites are our immediate concern. Of the former, the solar system contains eight, with Earth being the third rock from the Sun. It started with Galileo (see figure 1.3) and by now, astronomical observations have discovered 178 moons to orbit six of the planets. Venus and Mercury do not have known companions, and Earth and Mars have one and two, respectively. All planets

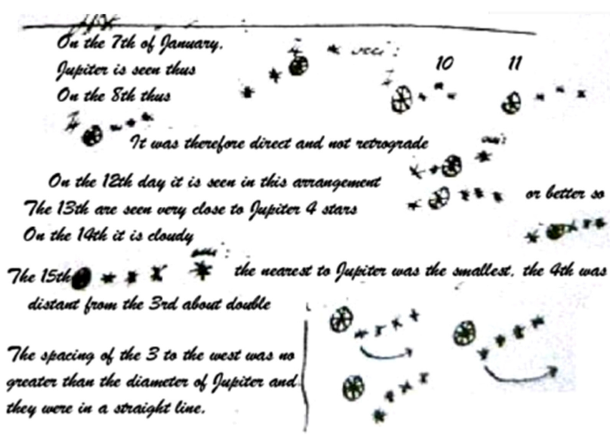

Figure 1.3. Page from Galileo's manuscript detailing the day-to-day alignment changes of Jupiter's four largest moons. Source: Regents of the University of Michigan (CC-BY-SA-4.0).

revolve around the Sun in non-crossing elliptical orbits, with the Sun in one focal point. The orbital planes of the planets are somewhat tilted against each other. Planetary satellites are also moving in planes and these are tilted against the orbital plane of their respective hosts. Finally, the planets turn around an axis through their own center. These axes are tilted yet again relative to the planet's orbit. However, this widespread tendency to be askew is, for the most part, only a matter of degree and does not lead to random orientation. Deviations from the mean of the planetary orbit tilt in particular are not very large. The corresponding axes of the planets Mercury to Uranus vary around the average by only about ±3.5°. In other words, all planets move in the same sense in a disc around the Sun. Most planets rotate around their inner axis in the same sense as they orbit the Sun. However, there are two odd balls—Venus is spinning retrograde and Uranus' polar axis lies almost inside the orbital plane. The next section will discuss in more detail how the various orientations and alignments can dramatically affect how time goes by as measured by the course of the Sun in the sky.

All time-keeping derived from celestial observations is ultimately rooted in rotational motion, be it via orbits around an external center, namely the Sun for the planets or the host planet in the case of satellites, or via intrinsic rotation around their innermost core. Rotation is repetitive and consistent with a view of time as being cyclical. A linear and directed quality of time—which has been dubbed the 'arrow' of time—arises from the uniqueness of history and thus from the breakdown of complete periodicity. At first and even second glance, celestial dynamics is cyclic. However, when you apply high precision to short-term studies or exercise patience and consistency in conducting long-term observations, you will discern deviations from purely repetitive motion. The former approach is a modern one, while Babylonian and Egyptian early stargazers were already applying the latter thousands of years ago.

1.4 What a day makes

We have seen that Earth orbits the Sun in one year and rotates around its own axis in one day, and that the Moon orbits Earth in about 29.5 days. These terse statements define three separate units of time. When we say that something has completed 'one rotation' or 'one orbit', we mean that the angle measuring the position of the corresponding object or feature has changed by 2π radian or 360°. Although the Sun and stars track similar circles in the sky, we specifically use the Sun's motion to define the length of one day. Returning to the same spot in the sky marks an average time span of 24 hours or 86 400 seconds. One might expect that substituting a star, any star, for the Sun and measuring the star's completion of one round in the sky would delineate exactly the same length of time. Let us take a good clock and measure. What we actually find is that the stars complete their daily circles faster than the Sun does—by about 235.9 s, a little less than four minutes. The geocentric model built on celestial spheres recognized that fact by letting the stellar sphere move a bit swifter than the sphere of the Sun. Originally, this adjustment was not implemented on account of the small difference per day, but rather because of the annual shift of the Sun relative to the constellations of the Zodiac. In any case, we now have two different definitions of the time unit 'day'—the solar day of 24 hours and the sidereal day of about 23.934 hours. Multiplying the difference between solar and sidereal days by the number of days in one year yields the annually accumulated difference. Interestingly, we find this difference to be equal to the length of one additional sidereal day (365.2425 × 235.9 s = 86 161 s = 23.934 h). In other words, during the course of one year, i.e. one complete orbit around the Sun, there is one more revolution of the Earth around its axis relative to the stars than there are rotations as measured relative to the Sun. This is no coincidence.

Thanks to a string of discoveries made around the beginning of the 21st century, and starting as early as 1988 with the first confirmed sighting, we know that at least in our own galaxy planets are abundant. It appears that planets might even be a necessary by-product of star formation. Imagine that we have perfected our instruments and that we can not only measure the motion of planets around their central star with great precision, but also the spinning around their own axes, as well as any geometrical aspect of their motion, such the eccentricity of the orbit and the tilt of the rotation axis relative to the orbital plane. In order to illustrate the difficulties faced when deriving a calendar from the observed changes of the Sun and the Moon, let us consider a few idealized cases. We have not (yet) actually observed these cases, but none would violate any known law of physics. Ideal systems just make it easier to reveal the fundamental issues. Who knows, maybe there are some civilizations out there that developed their calendars and clocks in such environments.

Suppose that one of the exoplanets orbits a star (let us call them Aplanet and Asun) in a mathematically perfect circular path. Unlike Earth, Aplanet's spin axis is oriented exactly perpendicular to the plane of orbit around Asun. Also unlike Earth, the exoplanet does not have a satellite. For observers such as us looking at the system from afar two obvious time scales exist: the time for Aplanet to complete one circle around Asun (one Aplanet year or one Ayear for short) and the time for Aplanet to finish one rotation around its own axis, a period we will call one Aday. The magnitudes of the two time scales are completely unrelated to each other because they are determined by accidental details of the formation of this star–planet system. Furthermore, the orbital and spin rotation could be in either the same or the opposite direction. Whatever the ratio, Aplanet cannot sit still in its place relative to Asun. Gravity would pull it very quickly towards the center and poor Aplanet would be gone in a flash—literally. The possibility exists, though, that Aplanet does not rotate at all around its own axis. In that case, would there be night and day on Aplanet? Suppose an Aplanetian living at the equator of Aplanet firmly plants a stick vertically into the ground. Because there is no rotation, the stick will always point into the same direction in space—by accident towards us[6]. From the stated assumptions it follows that we happen to be in the plane defined by the orbit of the

[6] There is a slight cheat here. I am describing the situation as if the geocentric model were correct. However, the distances to exoplanetary systems are so large that it makes no significant difference whether the stick points to us or to the Sun. To be entirely in the clear we can position our viewing platform somewhere on the line between Asun and our Sun.

distant planet. Therefore, at certain moments Aplanet will be exactly on our line of sight and in front of Asun and half an Ayear later exactly behind it. While Aplanet is in front of Asun from our point of view, the stick points away from Asun. Thus, the area surrounding the marker is dark. When the Aplanet is behind Asun, the stick—still pointing towards us—now also points towards Asun and the area is lit. Our ingenious Aplanetian will measure the length of day and night to be exactly equal to the period we measure for one orbit. A different question is how she could ever know for sure that Aplanet is in orbit. Other than the Asun's slow ascent and descent to and from the zenith, nothing is changing from one day (or year) to another. Thus, by the way, there is no clear distinction on Aplanet between annual seasons and daily weather, which presumably does exist, driven by the varying irradiation of the atmosphere during Asun's protracted rise and fall in the sky.

At the exact moment when Asun is at the zenith, the vertical stick will not cast a shadow. After that, Asun will move along a great circle in the sky, with the stick casting an ever-longer shadow. Once Asun touches the horizon, the light dims and, after some twilight, the fixed stars appear in the dark sky. Because Aplanet is the lonely companion of Asun there are no 'wandering stars', and no other planets to look at. What will the fixed stars look like? Remember that our assumption is that of no rotation relative to us, the distant observers who are in turn essentially stationary in relation to the fixed stars. Looking up overhead along the direction of the stick, a certain star might be visible. To aim the stick at another star, you would have to tilt the stick somewhat. No matter which star you choose, the stick will always point at the chosen star and will not wander away. On Aplanet, there is no diurnal stellar motion. Of course, the stars grow dimmer and eventually become invisible as dawn approaches and the sky grows lighter (Asun appears from below the horizon exactly opposite from where it sank below the horizon half an Ayear ago). If the Aplanetians have evolved on a planet without atmosphere, the stars remain visible. So here is an interesting finding: the central star of Aplanet is seen as 'moving' in the sky while the stars do not move at all. They are truly 'fixed stars'—at least over times scales that are short compared to any large-scale motion of this mini solar system. A would-be Acopernicus would face a daunting task to discover what we can see clearly, namely that Aplanet orbits Asun. But even in this instance there is a way to observe orbital motion—see the discussion of stellar aberration in section 4.2.

In terms of sidereal days and solar days, we can succinctly describe the above scenario by stating that, in the absence of planetary rotation, for each orbit there is one solar day and zero sidereal days (one might also say that the sidereal day never ends and thus is infinitely long). Could there be the reverse situation, so that there is an everlasting solar day and one sidereal day per orbit? The answer is 'yes'. Instead of discussing another hypothetical exoplanet (in which case I would be tempted to go backwards in the alphabet and name them Zsun and Zplanet), we will train our sight on a much nearer spectacle to illustrate what can happen when orbiting and rotation are combined. Meet Mr P!

Occasionally, Mr P enjoys a somewhat unusual ritual in the local park where he walks around a statue in a precisely choreographed manner to celebrate the cosmic dance of planets. Here is how he does it.

Beginning simple, the first act is dedicated to planets whose solar days and years are indistinguishable, just like the Aplanet we just visited. Mr P starts at a certain distance away from the statue, looking straight at it and simultaneously at a tall building, located much further away. From here on, Mr P fixes his gaze solely on the distant building, never to lose sight of it or change the direction in which he looks at it while he circumnavigates the statue. In this scenario, the statue will obviously disappear behind Mr P's back after half of the circle. In this analogy, the Sun (aka statue) shines during half of the planet's orbit (aka Mr P's circle); during the other half, the Sun is below the horizon and it is night. In other words, one solar day is exactly equal to one year and the sidereal day is infinitely long, since the stars never change position in the sky (or the building is always located in the same direction). This choreography is essentially the scenario just described for the Aplanet–Asun system.

Act two shows how a sun can stay permanently still in the sky. Throughout his circular path around the statue, Mr P now steps sideways and thus always looks at the statue. In doing so, the distant building gradually moves out of his sight (of course, he and not the building is doing the moving). After half a circle, the house has disappeared behind his back, only to reappear into his exact central vision after one full circle. Does Mr P turn around his own axis? Anyone watching from the building will certainly think so. From that vantage point, one sees Mr P's front at one point, then one of his sides, his back, the other side, and, upon returning to the initial spot, once more his front. So, yes, relative to the 'background' defined by the building, he does rotate around his axis. From Mr P's perspective, the statue does not move—the Sun remains stationary in the sky all the time. However, the building and all other points at the horizon move in a circular fashion—the stars are moving in grand circles while keeping their relative distances and relative positions constant.

Specifically in the case considered, orbiting and rotation take place in the same direction, say both clockwise when viewed from above. What would happen if they occur in opposite directions, but with the same period? Assume Mr P faces the statue and building when he starts his dance. After half a period, he has completed half a rotation around his axis and thus now looks away from the building. He has also finished half an orbit and thus looks again towards the statue. In between, after one quarter of a period, he will face away from the statue. During each round-trip, he will face twice towards and twice away from the statue, but only once towards and away from the building. In other words, there are two solar days but only one sidereal day per sidereal year. We do not have to travel to distant systems to find a planet like this: Venus in our solar system is very nearly performing such a dance (see figure 1.4). As already mentioned, Venus does rotate in a retrograde direction, and does so very slowly. Even today, the reasons for this odd behavior are unknown.

We have seen that the mere act of orbiting without any intrinsic rotation already yields one solar and zero sidereal days per revolution. In the absence of orbiting, the number of sidereal and solar days is equal. Therefore, in combination the two numbers differ by one. Depending on the relative sense of rotation and orbit, the 'extra' solar day from the orbital motion is either added to or subtracted from the number of days arising from the rotation around the planet's axis. How many days

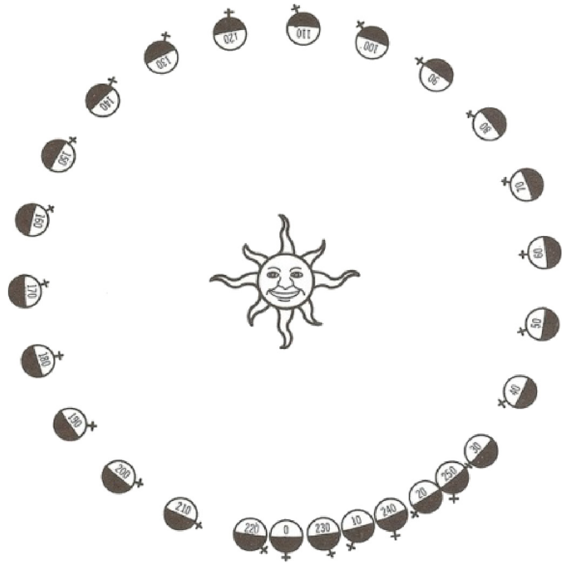

Figure 1.4. The orbital position and rotation of the planet Venus shown at 10 Earth-day intervals from 0 to 250 days. The position of the point of the surface that was the antisolar point at day zero is indicated by a cross. Reproduced from [7].

fit into the time span of one orbit obviously depends on the relative magnitude of the two associated periods. Because of gravitational influence and the internal friction of planets and their moons, the various periods do change slowly over time. In the Earth–Moon system, this interaction has led to a strict 1:1 synchrony of the Moon's rotational and orbital periods—Mr P's second dance.

In either of the two illustrations, a clear reference frame exists—the distant building, the distant stars—against which the definition of motion becomes clear and unique. In particular, the rotation of the planet around its own axis is evident and well defined. It is crucial that such a reference frame exists, a condition that turns out to be far from trivial. Our galaxy, for example, rotates around its center, carrying the solar system and us with it in a gigantic circle that completes in one *galactic year* (about 225 to 250 million years). Therefore, our vista of extragalactic objects and their alignment with the stars of our galaxy changes dramatically in one galactic year[7]. On much shorter time scales, proper, relative motion of stars in our galaxy is nowadays easily detectable. Once we accept that the stellar diurnal motion is only an apparent motion caused by Earth's rotation, the fixed stars are a good first choice for a space-fixed reference frame. It will do fine as long as we are not concerned with finer details or large-scale structures and long-term dynamics.

[7] Interestingly, much of cosmic history can be conveniently expressed with this time span. In galactic year units the age of the cosmos is about 61, that of our sun about 18, and the Cambrian explosion of life forms happened less than three galactic years ago. As useful as the SI units are, systems or processes can often provide 'natural' units that are more commensurate with the prevailing scale.

As we will see soon, Newton picked exactly this reference frame to represent an assumed 'absolute' space.

1.5 Discovering the laws of motion

In order to be consistent with the available data the geocentric model underwent significant changes, in particular the addition of nestled off-center circles to accommodate the complex motion of the planets. However, in the effort to stay abreast with observation, the model had become very cumbersome. Copernicus did *not* improve the quantitative agreement between theory and observation when he asserted—but initially did not dare publish—that the motionless Sun is at the center of the Universe or that while the Moon does revolve around the Earth, Earth itself and the other five known planets orbit the Sun. He certainly did not help the matter by asserting that the planetary orbits are circles with harmonic proportions of their radii. In fact, because Copernicus continued to insist on the circle as the foundation for all celestial motion, he encountered the same problems as the geocentric model when it came to incorporating the known and complex motion of the planets. Nonetheless, his work was radical and Copernicus agreed only after extensive prompting and prodding to publish his theory in the book *De Revolutionibus Orbium Coelestium*. When this indeed revolutionary manuscript eventually came out in print in 1543, it did not contain any discussion about what might *cause* the motion of the planets or what could sustain their orbits. In that sense, Copernicus was still quite close to the Ptolemaic worldview of eternal spheres that simply exist and need no further explanation.

Galileo was the first to see them—four tiny specks next to Jupiter (see figure 1.3 once more). One moment they lined up this way, a few days later a different way. Then they disappeared, only to come back into sight somewhat later. The change repeated like clockwork. With the help of the recently invented telescope, Galileo had discovered Jupiter's largest moons. He also found craters on our own Moon, stars that are invisible to the unaided eye, sunspots and Saturn's rings. The booklet *Sidereus Nuncius* ('Starry Messenger'), published in 1610, summarizes these remarkable observations. They shook Galileo's faith in the geocentric worldview and led him to support publicly the Sun-centered cosmology of Copernicus, a stance that famously got him into quite a bit of trouble. There is much more to Galileo's work than otherworldly discoveries. His astute examination of everyday objects and imaginative experimentations led him to insights that are part of the foundation of modern science—in particular and most relevant in the current context, the science of motion. Up until then the Aristotelian view held sway, in which two kinds of motion exist: one for ordinary matter, whose properties rest entirely on four elements, and another for the heavenly crystalline spheres, which rotate eternally without change and are made of a fifth, ethereal element (aptly called quintessence). Specifically, natural things released from the same height fall faster when they are heavier. Aristotle is silent when it comes to the behavior of objects *during* free fall. To be fair, this was difficult to observe, given the tools available at the time. Today we can use movies of the process played back in slow motion. Galileo did something equivalent—he slowed down the motion itself with the help of wooden ramps with variable tilt and smooth, straight grooves along which a polished brass sphere could roll downward. When the tilt angle is small, the sphere moves very slowly, allowing us to measure details of its motion. Galileo's arguments took off from the change in behavior with increasing tilt angle, extrapolating to free fall as the asymptotic limit of very steep incline angle. Incidentally, as a stopwatch Galileo used water flowing out of a vessel via a thin pipe into a measuring cup. The weight of water accumulated in the cup, determined with a balance, is then a measure of the elapsed time. From his careful measurements, Galileo found the correct relation between time and speed, namely that the speed of objects in free fall increases in proportion to the time spent since their release from rest. He also surmised that all objects fall at the same rate once drag and friction have been removed.

The final breakthrough to a modern science of mechanics had to wait for Isaac Newton, who was born in 1643, within a year of Galileo's death. It also had to wait for another level of precision and completeness in the astronomical database, as well as a refinement of the Copernican model. Between 1686 and 1687 Newton published his theory of dynamics (and so much more) in the three-volume *Philosophiae Naturalis Principia Mathematica*. In the preceding decades, others—some of the giants on whose shoulders Newton famously proclaimed to be standing—had accumulated a large body of ideas and reliable observational data on the motion of various objects. As already mentioned, Galileo in particular had measured the details of objects falling to the ground on Earth, including projectile motion. The Danish astronomer Tycho Brahe (1546–1601) is responsible for a similarly accurate and precise body of data related to objects in permanent free-fall around the Sun (aka the planets) or Earth (aka our Moon). Brahe was a colorful man with a rather colorful biography that includes losing part of his nose in a duel. He was also fabulously wealthy. On top of that, he managed to coax the king of Denmark and other patrons into giving money for an array of buildings and specially designed star gazing tools. The result was Uraniborg, at the time the most sophisticated center for astronomical observations, at least in Europe. Although no telescope was at their disposal, Brahe and his assistants were able to measure a vast number of ascension and declination angles of the planets with unprecedented precision. Simultaneously recording the time at which the planet's position in the sky was observed allowed the eventual reconstruction of the planet's motion from the data. In particular, extensive and precise data for the motion of Mars became available. The high quality of the observational data attracted a lot of attention, including that of Johannes Kepler, who in 1600 joined Brahe as an assistant in Prague. Brahe's new position as the imperial astronomer of the Bohemian king had a lot to do with deteriorating relations with his own monarch at home (as I said, Brahe lived an interesting life). Kepler's motivation was quite simple by contrast: he wanted access to Brahe's records. Of course, the senior astronomer did not just hand over his studies, but rather offered the junior partner a position in his firm. As one of the first tasks, Brahe asked Kepler to try to solve the retrograde puzzle of Mars using the detailed observational records. The challenge turned out to be more difficult than either one of the two had probably thought. It took Kepler more than a decade to solve the puzzle, but he did solve it. It turned out that observed facts, in particular Mars' apparent backward motion, were much more readily explained on the basis of all planets moving along *elliptical* orbits, rather than the harmonic circular motion invoked from antiquity up to Kepler's time, most famously by Copernicus. With that simple change, Copernicus' heliocentric model became a powerful tool to describe even minute details of the witnessed motion of the wandering stars. Kepler was able to summarize his findings into three rules (the first two in 1609, the third in 1618).

(1) Planets move along paths that have the shape of an ellipse, with the Sun in one of the focal points.
(2) An imaginary line connecting a planet and the Sun sweeps out equal areas in equal times, regardless of the planet's location on the ellipse.
(3) The squared time for completing one orbit—as judged relative to the fixed stars—is in proportion to the cubed major axis of the ellipse.

The clockwork in the sky begins to surrender its mystery, permitting much clearer views of some of its inner workings. The celestial clock has several hands, all moving in elliptical paths around the Sun as the central pivot. Mercury completes approximately 120 solar round trips for each Saturn orbit. So you can think of Mercury as being a half-minute hand, while Saturn shows the hour. Knowing the ratio of the periods of a given pair of planets for completing one full round around the Sun, you can compute from the third rule the ratio of their respective distances from the Sun—or vice versa. According to the second rule, planets move at unequal speeds along their elliptical orbits. Specifically, they move fastest when they are closest to the Sun and slowest when furthest away, i.e. near the 'empty' second focal point. That this is so follows from the fact that the connecting line to the Sun is shortest in the former instance and longest in the latter. Thus, in order to sweep out equal areas, the planet's speed has to adjust accordingly and as stated above. These examples show that by measuring time, spatial aspects of the solar system become apparent and geometric portions reveal temporal relations. A linkage between time and space already exists in classical physics. How this connection becomes even tighter in Einstein's twin theories of relativity is the topic of chapter four. For now, let us get back to our immediate story.

Typically, answers in physics—or in science in general—lead to more questions. How come the planets are so good at knowing Kepler's rules? I mean, they did not take driving lessons and just learned how to behave in traffic. A better question would be, what drives them to behave this way? During Newton's time, the general and somewhat vague notion had taken hold that gravity choreographed the dance of the planets. Several people (Newton himself names Wren, Hooke and Halley) had suggested that the gravitational force diminishes with increasing separation, specifically as the inverse square of the center-to-center distance. But this was based more on guessing than on quantitative arguments. At the same time, Kepler's proposition of elliptical orbits gained more widespread acceptance and some even tried, without success, to join the two ideas. So when they met at Cambridge in 1684, Edmond Halley (yes, the comet is named after *him*) asked Newton what he imagined the shape of orbits would be if the planets were attracted by the Sun with a force that weakens as the inverse square distance? Then not-yet Sir Isaac replied without much hesitation that he had mathematically proven that those orbits would have to be elliptical. At that moment, he could not quite find the pertinent notes, but not to worry. In three months' time Newton sent a paper (*De motu corporum in gyrum* or *On the motion of bodies in orbit*) mathematically proving that Kepler's elliptical orbits are indeed a consequence of the gravitational force law. After that, the floodgates opened. Within two years, Newton, with continued prompting and financial support from Halley, had published the three volumes of his masterpiece, the *Principia*, which is undoubtedly one of the most influential science books ever written. For our purposes, we are interested in the three laws of motion, which we express in the following, paraphrased form, with their modern symbolic expression given in parentheses.

First law — Unless a non-zero force acts on them, objects remain at rest or in uniform motion along straight lines (v = constant if $F = 0$).

Second law — When a non-zero force acts on an object, the velocity of the object changes in proportion to the strength and in the direction of the force and in inverse proportion to the mass of the object ($\mathrm{d}v/\mathrm{d}t = F/m$)

Third law — To every action there is a reaction of equal strength and opposite direction. This is in particular applicable to two objects interacting with each other ($F_{\text{1-on-2}} = -F_{\text{2-on-1}}$).

As self-evident, limited, or boring these statements might appear to some, they express deep and important insights. Much of modern engineering and manufacturing owes its efficiency, if not its entire existence, to these equations. More importantly for the context of this book, they also incorporate assumptions about space and time, which will be the topic of the next section.

1.6 Absolute space and time

One of the benefits of Newton's equations of motion is their practical utility. I tell you the position and velocity of an object at one moment in time. In addition, I specify the force acting on the object in a certain region of space. With that information, Newton's second law enables you to predict the position and velocity of the particle at *any* other time, as long as the object does not leave the region. You can predict the future—for example, the date and time of the next lunar eclipse. The more precisely and accurately you can identify the force and initial conditions, the more precisely and accurately do you know the entire path of the object. Space probes are another example. Whether they journey 'only' straight to the Moon or more ambitiously to Pluto on an intricate trajectory involving several gravitational slingshot maneuvers, the flight plan to get them from A to B is based on the procedure outlined above. The Jet Propulsion Laboratory in Pasadena in California is home to an astonishing database and set of equations, the Ephemerides [8], that model the dynamics of our solar system—planets, satellites of planets, asteroids and comets. At its core are Newton's equations of motion and an ever-growing set of positional observations. Think of this model as a digital orrery, a virtual painting of the detailed comings and goings of Mercury and Mars, Venus and Neptune, Titan and Europa—the whole pantheon of Greek mythology resurrected in a computer.

In addition to their inherent utility—certainly not only for astronomers—the laws of motion contain basic ideas about space and time that are still relevant today. The first law stipulates that in the absence of a force, objects either stay at rest or move with an unchanged velocity. Is that reasonable? You managed to get a window seat on your next flight to Paris (lucky you) and the tarmac is rushing with increasing speed past you as the jet is about to take off. If Newton's laws are valid, then you are compelled to say that a force acts on the tarmac, heck on the entire ground plus airport and all. Of course, you well know that no such thing takes place. While Newton had no inkling about airplanes, he was well aware of the existence of accelerated frames of reference such as the one just described. The wording of the first law indicates that his theory only applies to reference frames in which acceleration of objects comes about because of an actual force. These systems also go by the name *inertial reference frame*[8]. The interior of an airplane taking off is *not* an inertial reference frame. Newton understood the motion of inertial reference frames as occurring in absolute time and relative to absolute space, and therefore a special reference frame exists that is at rest relative to space. Space is the container in which all objects exist, the stage on which all events unfold. A universal clock against whose rhythm the pace of every change is measured tells absolute time. Both space and time exist apart from and independent of any material object. Not everyone bought into these concepts, including Gottfried Wilhelm Leibniz (1646–1716), a polymath and philosopher who took issue with the idea of space and time existing in isolation. Despite such opposition, the concept of absolute space and time took hold of physics for the next 200 years and more until the beginning of the 20th century, when a patent clerk … well, I am getting ahead of myself. That part of the story has to wait.

Newton's theory of motion has more to say about time: it offers a new approach to the notion of the 'now', the instantaneous moment of (or in?) time that is analogous to a mathematical point in (or of?) space. Newton fashions this contribution using a new tool, the mathematics of differential calculus, which he invented for the explicit purpose of applying his theory to concrete real-life situations. In the above summary of the three laws, I included the modern mathematical expressions that express the same content in a highly condensed way—in particular, the differential calculus expression dv/dt for the rate of change of the velocity, i.e. acceleration. Of course, *rate* was and is a familiar concept even without calculus. We all know—or at least should know—that our bank account will deplete if the spending rate is larger than the rate at which money flows into the account. The level in a leaky bucket will rise if you pour in water faster than it flows out. As long as the rate of change is constant, we can also easily calculate the overall change. Suppose you accelerate your car from rest at a rate of 10 m s^{-2} (or a in general). After 10 seconds (or t in general), your car will run at a speed of 100 m s^{-1},

[8] In chapter 4, we will explore the significance of inertial and accelerated reference frames in the context of Einstein's theories of special and general relativity. Also, for a good explanation why the first law is not just a redundant consequence of the second, see, for example, [9].

about 100 km h^{-1} (or $a \cdot t$ in general). Easy, but not realistic. If the acceleration were truly constant, the speed of the car would grow above all bounds, which clearly does not happen—for better or for worse. In reality, due to air drag and the limitations of the motor, the acceleration will diminish with time and eventually the car will reach its top speed. Calculations become more challenging when rates are not constant and that is precisely why and where we need differential calculus. If the acceleration of our car varies, how can we figure out how fast the car runs at any given moment and how far it has traveled since the start at time $t = 0$? If you are willing to dispense with accuracy, you can use a simple recipe to find approximate positions and velocities at discrete moments of time t_1, t_2, t_3 separated by a short time interval Δt (i.e. $t_2 = t_1 + \Delta t$, $t_3 = t_2 + \Delta t = t_1 + 2\Delta t$, etc). Just as abbreviations in cooking recipes (tbsp = tablespoon, tsp = teaspoon, °C = degree Celsius, etc) are useful and common, we make use of writing a_n for the value of the acceleration at time t_n, with $n = 1, 2, 3 \ldots$ enumerating the time steps. In order to simplify matters further, let us assume that the motion is along a straight line (we hold on to the steering wheel and hope for the best). With that, the recipe reads as follows.

1. As the king told the white rabbit, begin at the beginning—where you know position and velocity. For our example of the car, we have $x_1 = 0$ and $v_1 = 0$, respectively. Because force and mass are specified everywhere, you also know the acceleration at that moment.
2. From the data in step 1, calculate the velocity at the next moment t_2 as the sum of the previous velocity and the velocity change due to the acceleration acting during the intervening time interval, i.e. $v_2 = v_1 + a_1 \cdot \Delta t$. Note that this assertion is not exactly correct, because the acceleration is *not* constant. Therefore, a better choice for the acceleration would be the average acceleration between times t_1 and t_2. However, if the time interval is short, the error will be small. Because the car is initially at rest, i.e. $v_1 = 0$, we cannot use the exact equivalent expression for the new displacement. Instead, we must right away use the average velocity or (less accurate) the velocity at the end of the step: $x_2 = x_1 + v_2 \cdot \Delta t$.
3. Now that we know the approximate values for the position and velocity at t_2, we can proceed in just the same way as in step 2. With every new step, you propagate position and velocity forward until you reach the end of the allotted time—and, yes, you do stop there.

As already mentioned, this recipe will lead to errors in both predicted final position and final velocity. However, by using average values in the step propagation and shrinking the intervals, these inaccuracies will decrease. The power of calculus consists in the mathematically provable property that the error will be zero in the limit of shrinking the time interval Δt to zero. In this limit, intervals of any quantity—not just time—are said to be infinitesimal[9] and will be denoted by the letter ‘d’ rather than ‘Δ’.

[9] Full disclosure: infinitesimal changes or differentials are not ordinary mathematical functions and we should treat them with proper respect and courtesy. Having said that, I will probably be negligent, and ask mathematically inclined readers for leniency.

Infinitesimal changes are called differentials. So the infinitesimal change dv of velocity during the infinitesimal time interval dt due to an acceleration a at some moment of time t, can be written as dv = a·dt. Consequently, the instantaneous acceleration a is the ratio of the two differentials dv and dt:

$$a(t) = \frac{\mathrm{d}v}{\mathrm{d}t} = \underset{t_2 \to t_1}{\text{limit of}}\left(\frac{v_2 - v_1}{t_2 - t_1}\right) = \underset{\Delta t_2 \to 0}{\text{limit of}}\left(\frac{v(t + \Delta t) - v(t)}{\Delta t}\right).$$

The machinery of calculus will not be of concern to us in this book. Occasionally, I will use results without giving proof. In general, it is sufficient to appreciate that instantaneous rates of change are well defined and have specific values that are usually finite. Newton's genius is certainly not restricted to the invention of calculus, but that creation in itself was a formidable achievement. It enabled him to compare his theory with experimental and observational data—a central aspect of all of modern science.

Here, then, is the precise manner in which Newton introduces a new aspect into the concept of the temporal 'now'. Instantaneous quantities—such as velocity or acceleration—are the result of shrinking the ratio of finite differences to a 'point', the mathematical limit of contracting the denominator of the ratio to zero. In particular, any rate is the limit of a finite difference ratio with a time interval as the denominator. Therefore, the expectation that theoretical models using calculus can mimic reality is equivalent to the assumption that we can divide time into shorter and shorter intervals without limit. While this sounds reasonable and may even be true, for any physical property of matter the corresponding statement is incorrect. As an example, take the amount of a given substance—say a piece of wood. Use your finest axe and chop the piece into splinters. Then take a sharp knife and carve away. Collect the finest chips or better yet some of the dust you produced. Under a microscope, you may see a few tiny slivers that are still wood. However, if you go further, you will notice that you introduce a fundamental change. For starters, wood is a conglomerate of substances that eventually separate into different molecular structures. With the help of refined tools, you can even tear the molecules apart. Being a physicist, you are inclined to claim that from the beginning the wood was only a collection of atoms and so the division has as of yet *not* produced anything different. It will, though, when you start splitting the atom, and we all know that you can. Here is another example showing that after a finite number of divisions the character of the chopped-up quantity actually changes. Forgot to put the fabric softener into the dryer? You know you did when the sock you take out clings to the shirt or your hand because of static electricity. A standard axe will not do, but there are ways to break up electric charge, or at least to look for its natural denomination. Robert Millikan (1868–1953) was the first to do so and to find that (1) a smallest unit exists and (2) all electrical charges are integer multiples of this *elementary* charge. Today, an extremely precise value for this charge exists (if you must know: 1.602 176 6208·10^{-19} in SI units, with an uncertainty of about six parts in a billion). Present experimental evidence, as well as what is currently the best theory of matter—the standard model—agree that electrons and thus their charge are indivisible. Of course, the indivisibility of charge or the change of character when splitting the nucleus or any other example does not preclude that space and/or time can be continuously divisible. We do not know. For practical purposes, it also does not matter, because we do know that we *can* split the fleeting second into many, many fine segments whose duration we can measure with exquisite precision. Thanks to atomic clocks and other developments (see chapter 3), we now have the ability to determine time intervals as short as attoseconds—one billionth of one billionth of one second—and there is no evidence of time changing its character. Whatever that would mean, anyway. Similar fundamental questions (see chapter 6 for a few examples) are sometimes questions with vague or possibly even meaningless answers. Time is a slippery 'thing'.

1.7 A bit of mechanics—or why the solar system is stable

What is it that keeps the rhythm alive? What is the magic that keeps the grand cosmic clock running for billions of years? It appears that after the youthful universe went through a phase of massive inflation (see chapter 6), it has become thrifty and now spends only what it owns—at least to the degree that we can count the coinage. That type of currency comes in several denominations: energy, linear momentum, angular momentum, electric charge and some attributes that are more exotic. For closed systems, the overall 'amount' of these properties is 'conserved', i.e. the total sum of the values for the individual parts of the system is unchanged over time. It is not quite clear what happens in black holes, but otherwise the Universe is a closed system. Subunits of the Universe, like our solar system or the Milky Way, are only approximately closed and change does occur. However, as long as the exchange of energy, momentum, etc is not large compared to the amount present, the time scales for change measure into the millions of years and more. Of course, at certain times, for example when a supernova marks the end of the life cycle of a star, change is drastic and swift.

How can a planet slow in the rotation around its own axis or in its orbit around the Sun? One of the consequences of the laws of motion is that a force acting perpendicular to the direction of motion deflects the object without changing its speed. Conversely, any force that is not exactly perpendicular will either increase or decrease the speed. Only for elliptical orbits with the gravitational attractor, the Sun, in one of the focal points is the force component tangential to the orbital motion in just the right way to keep the orbit stable. If planets were to experience any additional force opposite to their direction of motion, commonly referred to as drag or friction, they would slow down, spiral closer to the Sun, and eventually crash—just as a low-orbit satellite exposed to a planet's atmosphere does. The vast majority of dust, debris and chunks of matter that existed in the formation period of the solar system were absorbed long ago, mostly into planets and asteroids. By now there is just not enough stuff left over to influence the planets significantly. Another potential source for drag could be the corona and the plasma halo of the Sun. Yet even Mercury as the planet closest to the Sun is in no danger of crashing into the Sun any time soon.

Even so, the stability of the planetary orbits is not mathematically absolute, mostly because of the gravitational pull of large planets like Jupiter and Saturn and/or of their own moons, which are much smaller but also much closer. These gravitational forces are acting along lines that at any given moment do have a component that is tangential to the planet's orbit. In the long run, though, this direction is averaged out and does not lead to a long-term net increase or decrease of the speed, only to quasi-periodic variations of the orbital and rotational parameters. For Earth, the so-called Milankovitch cycles are an expression of the complexity of these variations. Leading terms of the overall variation include the already-mentioned precession of the rotation axis, as well as a change of the eccentricity of Earth's orbit. Typical periods for the dominant variation are tens of thousands of years, but all changes combine to convoluted and aperiodic motion. Overall, then, the length of the sidereal periods of planetary orbits are nearly unchanged, since they only depend on the semi-major axis, which at least to first order is unaffected by the perturbations. Thus, while running in a slightly bumpy way, the great clock in the sky runs true and for a long while without the need for rewinding.

1.8 Not quite a clock yet—the Antikythera mechanism

In 1901, a small Greek ship carrying sponge divers back home from a not-so-successful stint in the Western parts of the Mediterranean Sea was caught in a storm. The crew sought shelter at the coast of the small island of Antikythera north of Crete. When the divers decided to explore the underwater neighborhood, they did not find sponges. Instead, they discovered what turned out to be the remains of a ship that had fallen victim to another storm about 2000 years earlier. The find created quite a stir in archeological circles, since it was the first discovery of an ancient shipwreck. The cargo scattered on the bottom of the sea included sculptures and glass vessels of various kinds, giving valuable clues for the date of the ship's journey, which was estimated to be around 150–100 BCE. Among the scattered parts on the seafloor, the divers found fragments of a geared mechanism that archeologists later proclaimed to be an astrolabe—an instrument used to measure vertical angles for navigation and stargazing. Soon afterwards, the crated pieces were stored in the Museum of Athens, and forgotten almost for good. Half a century later, a careful reexamination [10] led to the startling insight that the mechanism was actually a sophisticated mechanical device capable of keeping track of a variety of periodic events—both on Earth and in the heavens. The machine allowed calendric calculations, such as determining the times of the Olympiad. Even more astonishing, this analog 'computer' was capable of mimicking the motion of the Sun and Moon, including the prediction of eclipses. Possibly, although this is unclear, the device could also show the position of the five planets known to the Greeks. Nothing even remotely resembling such a mechanism was (or is) known from that early time. The level of sophistication of the overall design and construction, in particular of the gears, and the inherent precision of this earliest known astronomical simulation mechanism were unmatched until the fourteenth century. Nevertheless, this remarkable instrument was not a clock. It did not produce repeatable, constant units of time, and neither did it otherwise allow measurement of the duration of elapsed time. There is also no known later mechanism that could be seen as has having benefitted from this early invention. On the other hand, there is something transcending the uniqueness of this machine. Like Stonehenge or Ptolemy's crystalline spheres, the Antikythera mechanism is an attempt to reproduce and capture at least aspects of the natural rhythms surrounding us. Using purely mechanical means, it acknowledges that such rhythms exist and succeeds in mimicking them. To my mind, this is the general hallmark of dynamic model building and specifically an engagement with the notion of temporal change. The following chapter investigates the next steps of this journey, with a focus on the mechanical devices that we properly call clocks.

References

[1] Hartmann W K and Davis D R 1975 Satellite-sized planetesimals and lunar origin *Icarus* **24** 504–15

[2] Šprajc I 2017 Astronomy, architecture, and landscape in prehispanic Mesoamerica *J. Archaeol. Res.* https://doi.org/10.1007/s10814-017-9109-z

[3] Barnhart E L The First Twenty-Three Pages of the Dresden Codex: The Divination Pages

[4] www.space.com/21719-vega.html
Sharrah P C 1975 Pole stars of other planets *J. Arkansas Acad. Sci.* **29** 62–3

[5] http://image.gsfc.nasa.gov/poetry/ask/a11846.html

[6] Kuhn T S 1957 *The Copernican revolution: Planetary astronomy in the development of Western thought* (Cambridge: Harvard University Press)

[7] Koenib L R *et al* 1967 Handbook of the Physical Properties of the Planet Venus NASA SP-3029

[8] The homepage of the JPL is a good starting point to explore: http://ssd.jpl.nasa.gov/?ephemerides

[9] www.quartets.de/acad/firstlaw.html
Pask C 2013 *Magnificent Principia: Exploring Isaac Newton's Masterpiece* (Amherst NY: Prometheus Books)

[10] de Solla Price D 1974 Gears from the Greeks. The Antikythera mechanism: a calendar computer from ca. 80 BC *Trans. Am. Phil. Soc.* **64** 1–70