Project Physics 2

Travis Norsen

December 31, 2008

Preface

Let me try to explain what this book is, why I wrote it, and how I think it can and should be used. To begin with, it is a text designed to be used in the second semester of a more-or-less standard "freshman physics" course. It presupposes a familiarity with the standard topics covered in the first semester of a college physics course – kinematics, projectile motion, circular motion, Newtonian dynamics, momentum, and energy. The book uses calculus and is therefore probably best used in a "University" (calculus-based) as opposed to "College" (non-calculus-based) Physics course, at least for institutions large enough to have such a distinction. But I have used this material extensively and successfully at a small college where there is only one "freshman physics" course and many of the students have not had calculus.

Many of the topics covered in this text are standard ones for this kind of course – e.g., rotational kinematics and dynamics, gravitation, the physics of gases, and some thermodynamics. But this is not just another ordinary textbook, with its boring and (to the student) seemingly arbitrary progression from one chapter to the next. Instead, the standard topics have been tightly integrated into two broad historical arcs, covering the origins, development, and applications of two great theories of pre-20th century physics: Newton's Theory of Universal Gravitation and the Atomic Theory of Matter. The standard topics are therefore supplemented and integrated with significant additional material that might normally be found in a history of science book, but is (too) rarely seen in science textbooks.

The original purpose of this re-organization of the curriculum was simply to make the standard material more interesting, by placing it in a historical context and giving the course as a whole a sense of drama and mystery – one might say, a plot. I think it succeeds on this front. But the inclusion of historical material serves additional goals as well, most importantly the inculcation in students of a realistic understanding of science and scientific method. In virtually all other disciplines (in the humanities, arts, and social sciences) it is taken for granted that literacy in that field requires a firsthand knowledge of and appreciation for the important historical figures of that field. A proper education in philosophy, for example, simply requires that one has read Plato, Aristotle, Descartes, Hume, and Kant. What well-educated literature student has never read Shakespeare? There is a kind of irony (or perhaps tragedy) in the fact that the natural sciences – where contemporary work is most obviously and most hierarchically grounded on earlier historical discoveries – tend to educate students un-historically. This means that science students are largely asked to accept the claims they are taught without understanding their historical origins – i.e., without understanding the *evidence* which makes it scientifically rational to accept those ideas. In other words, science – too often – is taught as a kind of dogma which students are asked to accept on faith from an authority (the textbook or the teacher). The tragedy is that the sciences are precisely the area where such an appeal to authority and faith is unnecessary. Actually, the real tragedy is that students taught this way will never fully appreciate the difference between science and the (irredeemably dogmatic) ideas that also vie for influence over their lives.

Research in science education in recent decades has revealed that traditional lecture classes tend to increase students' (often already-existing) sense that science is all about memorizing equations without thought or question or understanding – i.e., that science is dogma. We have thus seen an explosion in the use of more inquiry-based and project-based approaches to structuring classrooms and class time. We have not, however, seen similarly radical restructuring of the *content* of such courses, as manifested most obviously in textbooks.

This isn't to say that there haven't been some important improvements. For example, Priscilla Laws' *Workshop Physics* curriculum is specifically designed to allow a course in which a traditional textbook plays, at best, a secondary role. And other modern texts (e.g., *Understanding Physics* by Cummings et al.) make some valuable attempts to connect the material in the text with experiments and projects which may be performed by students in class. Also, the "Tutorials" pioneered by the PER group at the University of Washington have made significant headway in encouraging and allowing teachers to spend some class time (often what used to be spent on "problem sessions") having students work in small groups to confront and master challenging qualitative and conceptual issues.

Still, for the most part, even the (in this sense) best current textbooks have tables of contents that are virtually indistinguishable from those of many decades ago. So one of the things I am after in this book is to try a more radical re-structuring of (at least this part of) the curriculum. This is not merely a small tweak to the standard sequence of topics, pertaining only to how class time is spent or whether end-of-chapter exercises are phrased in the third or second person. It is rather the result of stepping as far back as possible from the standard curriculum and asking: what should students at this level actually be learning and doing? And perhaps more importantly: what should students know and be able to do before they go on to more advanced coursework in physics (or elsewhere)? And perhaps most importantly: what should students know and be able to do as preparation for life in the real world (whether as scientists or not)?

My answers, in outline, are as follows. Students should know something about the broad historical development of the important theories in physics, and this historical knowledge should be tightly integrated with their technical, mathematical knowledge of those same theories. That is, they should genuinely understand why it is scientifically rational to accept and use the theories – and they should know some of the practical benefits, whether to technology or to subsequent science, of doing so. Students should be spending time actually doing real physics – working with raw data, participating actively in important derivations, performing experiments, thinking creatively about how to set up challenging problems, and occasionally making (or reproducing) important discoveries. Students should understand that plugging numbers into formulas is not physics. They should understand that science is fundamentally a way of knowing in which all claims are *ultimately* traceable back to empirical evidence. And they should begin to see the virtue in applying the methods of science to every aspect of their lives.

The historical structure of the text is, I think, an indispensable means to the achievement of these goals. This, however, is no history (or history of science) book. I am no historian, and (although I have made significant efforts to read all of the primary literature that is discussed in the text) the historical accounts found here are still heavily based on secondary sources and will no doubt suffer from many of the vices (such as Whiggishness and an unrealistic emphasis on the importance of theorists) often attributed to "bad history." My only defense against such charges is that this is not intended as history, but rather as *physics* – at a level that can and should be understood by every "freshman physics" student.

There is a more important sense in which this is not a history of science. History texts are typically written for humanities students, whose technical math skills are not strong. This is a book written for scientists and science students. Hence, our goal is not to teach history per se, but to teach the *technically rigorous* physics ideas (as they have evolved historically). Thus, instead of shying away from technical discussions and resorting to loose qualitative analogies, we tackle mathematically rigorous material head-on and expect students to do the same. If anything, the level of mathematical rigor in this text will be seen as higher than a standard freshman-level text. The equation-to-word ratio is probably a bit lower than in normal texts, but the fact that each equation and derivation plays some important role in the evolving plot raises the stakes: individual results cannot be taken (memorized, applied) in isolation, but must be fully digested as part of a coherent whole.

Let me say a few things about how I think this book should be used. It is designed to be used in – or more precisely, is designed to help bring about – a more "inquiry" based or Problem- (or Project-) Based-Learning (PBL) classroom environment. This has several aspects. First, the text itself is written in a way that students should be able to follow and understand. I see no need to spend valuable class-time having the teacher lecture on the material covered in the text (much as it would be regarded as preposterous for a literature professor to spend class-time reading Shakespeare to his students). Students should read the text outside of class – *before* class – thus freeing up class-time for more focused and purposeful Q&A periods, discussion of difficult concepts or derivations, working through derivations and discoveries that are merely sketched in the text, etc. I have included a number of "Questions for Thought and Discussion" at the end of each chapter to stimulate productive, open-ended discussions.

Each chapter also contains a number of end-of-chapter "Projects." These are typically very different from the end-of-chapter Exercises or Problems found in most texts. Since this book is designed for use in a physics course, a necessary pre-requisite for which is students' ability to do algebra, I simply take for granted that students can algebraically manipulate given equations and plug numbers into them. Those activities are appropriate for an algebra course, but they are not physics. The Projects in this book, by contrast, ask students to engage in real physics. For example, students are given raw data and asked to interpret it from the perspective of various theoretical models. They are asked to invent their own theoretical models. They are asked to fill in steps that were missing from important derivations sketched in the text. They are asked to use computers to solve equations which are too complicated to solve analytically. And they are asked to participate in and reproduce important moments of discovery.

These Projects are generally somewhat open-ended. They almost always require some amount of creative integration, by the student, of different pieces that were covered in the text. And very often they are designed to put the student in the shoes of actual scientists working with actual data and making actual discoveries. Thus, not just the main text of the book, but also the end-of-chapter assigned work, is designed to convey to students a much more realistic picture of science and its methods. Traditional plug-and-chug exercises teach students that deep understanding and creative thought are not required in science – that doing science consists of blindly grabbing pre-formulated "magic formulas" and then going through a rote procedure. None of the Projects in this book can be done blindly – without creative thought, genuine understanding, and hard work.

Note that the open-endedness, complexity, and sheer difficulty of (many of) the Projects in this book may require a shift in perspective when it comes to assessment and grading. Good students who genuinely understand the material, work in earnest, and actually think about what they're doing, may nevertheless not get "the right answer" at the end of the day. (Indeed, in some cases, there is no clear right answer.) That's OK. The shift in focus urged by this text – from physics as out-of-context dogmas, to physics as fundamentally a *method* of knowing – must extend all the way to how student homeworks and exams are graded. What matters most is that students understand what they are doing and have a coherent, rational, scientific justification for their approach and their conclusions. One of the crucial lessons of the history of science is that people on the losing side of scientific debates – those who had the wrong answers, as judged by future hindsight – should not necessarily be judged fools, and may not even have done anything wrong. (Often it is only *later discoveries* which show, finally and conclusively, that a given claim is definitely false.) Having learned this important lesson from history, we should surely extend the same courtesy to our students in the present.

I've already mentioned that I like to spend (at least part of) one class period each week in a kind of open discussion format, typically using students' answers to the "Questions for Thought and Discussion" as a jumping-off point. I've also explained that I envision the book being used with a project-based classroom structure. What that means specifically is that significant class-time each week should probably be spent working through the end-of-chapter Projects.

Sometimes the Projects are similar enough to material that was covered in the text that students (perhaps working in small break-out groups) could be immediately set to the task of working through them. Others are perhaps best approached in a whole-class-demonstration format, with the instructor taking the lead in working through the Project (but with lots of active participation from students). Weekly homework assignments should probably consist of several additional Projects. (If there is a particularly challenging or complex Project that I want to assign as homework, I sometimes give students time in class to start working on it – by themselves, or in groups, or with some assistance from me.) I also like to reserve a little bit of class time at the end of each week (or whenever the homework is due) for students to ask questions about the assigned Projects, to compare solutions with one another, etc.

Let me finally say something about the book's (still tentative) title. It was only long after embarking on this curricular project that I learned that, in some ways, I was re-inventing the wheel. Many before me have had the idea of spicing up science education by incorporating historical material. In physics, the best and best-known and most systematic attempt in this direction was Harvard's *Project Physics* (Cassidy, Holton, Rutherford, et al.) which began in the early 1960s (growing out of some earlier, related work, e.g., that leading to the splendid "Harvard Case Histories in Experimental Science") and produced textbooks and supplementary materials that stayed in print until at least 2002. I have benefited greatly from this work, and the current text owes much to it. But (despite its title), Project Physics was in no way project-based. Its scripted content was, I think, a significant improvement over standard texts, by virtue of its historical, inductive approach to the subject matter. But the work assigned to students was largely standard, plug-and-chug type exercises that gave students little opportunity to think creatively, use contemporary tools and develop contemporary skills, or participate in the discovery process in a genuinely first-hand way. Actually, the earlier editions of the curriculum (from the 1960s) are pretty good on this front. The more recent incarnations (with title *Understanding Physics*) are not only significantly dumbed-down relative to the original, they also lost the spirit of creative scientific inquiry that was, at the beginning, the whole point. (It is also sad and rather telling that the far more rigorous original text from the 1960s was intended for high school students, while the contemporary dumbed-down incarnation is intended for college students, even if not scientists and engineers.)

In any case, the (tentative) title is in some ways an homage to the original Project Physics course. It is particularly appropriate since one of my central motivations was to design a curriculum that left significant creative work to be done by students, in the form of the Projects. I also wanted a title that would make obvious that this is a book for use not in the first, but in the second, semester of "freshman physics." Hence: Project Physics 2.

Let me finally mention some of the other texts that I have learned from or leaned on in the preparation of this book. Malcolm Longair's *Theoretical Concepts in Phsyics*, which is intended for junior or senior physics majors, is perhaps the most similar book I've found to the current one and also one of the most inspirational to my project. In some ways, my whole project is a response to Longair's book: the topics he covers and the way he covers them are *so important and right* that they, I thought, should and must be done at the introductory level. The *Feynman Lectures on Physics* has also been inspirational to me... though what physicist wouldn't claim that? Feynman's legendary course was not organized historically and did not include open-ended Projects to be tackled by students in class and in homework. But I have attempted to imitate, as much as possible, Feynman's casual-yet-penetrating style, and the way he tightly integrates technical, conceptual, and historical material. The works of Stephen Brush, including especially his (and Gerald Holton's) *Physics: The Human Adventure* and his *Statistical Physics and the Atomic Theory of Matter*, have been particularly helpful. Thomas Kuhn's classic tome on *The Copernican Revolution* first sparked my serious interest in the history of science and subsequently helped me navigate through some otherwise-impenetrable primary texts. Cohen's *Birth of a New Physics* was a particularly influential example of how to explain the discoveries of Galileo and Kepler at an appropriate level. The *Harvard Case Histories in Experimental Science* were probably the most useful secondary sources for the second half of the book. And lots of other things too...

Let me finally say that this is very much a work in progress at this point. And so your feedback will be greatly appreciated!

Contents

Preface **i**

I Newton's Theory of Universal Gravitation **1**

1 Greek Astronomy **3**

1.1 Basic Observations . . . 3
1.2 Astronomical observations in more detail . . . 9
1.2.1 The Sun . . . 9
1.2.2 Moon . . . 12
1.2.3 Other Planets . . . 12
1.3 Measuring the distance to the Sun and Moon . . . 15
1.4 Ptolemy's Theory . . . 22
1.4.1 Epicycles . . . 22
1.4.2 Angular Kinematics . . . 25
1.4.3 Angular and Rectangular Coordinates . . . 26
1.5 The Precession of the Equinoxes . . . 27
1.6 Arguments against heliocentrism . . . 30

Part I

Newton's Theory of Universal Gravitation

Chapter 1

Greek Astronomy

Isaac Newton first published his theory of universal gravitation in his most important book, the *Principia Mathematica*, which came out in 1687. This book presented the entire structure of Newtonian mechanics (the three laws of motion, momentum, etc.) that was covered in the first half of your course. But it also extended the previously-familiar terrestrial notion of gravity out into the heavens and presented a detailed theory of how objects in the universe attract one another gravitationally. The primary evidence for this new theory was that it correctly accounted for a wide range of astronomical and terrestrial phenomena such as: the motion of the planets around the Sun, the motion of moons around planets, the motion of comets, the twice-daily rising and falling of the ocean tides on earth, and the fact that the earth's rotation axis changes its direction over a period of thousands of years. We'll discuss all of these effects as the semester progresses.

If we want to understand Newton's theory of universal gravitation, though, our first job must be to understand how the theory came about. And given the central role played in his thinking by the motion of the planets around the Sun, it makes sense to first try to understand how and why Copernicus proposed (way back in the first half of the 1500s) that the planets (including the earth) moved around the Sun. And in order to understand *that*, it makes sense to first try to understand how astronomers prior to Copernicus thought about what was going on in the heavens. So, for this first week, we'll need to do our best to forget everything we think we know about Newton and gravity and Copernicus and the fact that the planets go around the Sun, and start, as it were, from the beginning. One should think of this as an attempt to understand clearly the earliest seeds out of which Newton's theory of gravitation eventually grew.

1.1 Basic Observations

If you open your eyes and *just look* at how various things move, there are a couple of things that jump out at you. First, familiar terrestrial objects (like rocks and cars and people and trees) pretty much just sit there at rest unless something applies a force to make them move. A rock will just sit there on the ground – until or unless someone comes along and kicks it, in which case it will move some distance but then come again to rest. And it's clearly the same with lots of other things. Even people seem to have stillness as a kind of natural state: you can of course move around, but it takes some effort or energy to do so – and eventually (when you die) you'll lose the ability to exert that needed effort and hence stay at rest permanently. To summarize, it seems that things on earth have *rest* as their natural state, and that the unnatural state of *moving* requires some kind of *force* to be exerted. A budding physicist might try to summarize this with something like the following law:

$$F \sim v \tag{1.1}$$

where F represents the applied force, and v represents the velocity of an object. The idea is: if no force acts on an object, its velocity will be zero, and the faster you want it to move, the harder you have to push. Of course, we (who know about $F = ma$) know this is wrong. But if we forget about that and try to just consider familiar sorts of observations in a naive, unbiased way, Equation 1.1 is at least not crazy.

Let's then consider things up in the heavens such as the Sun, Moon, and stars. How do these things move? Probably the motion of the Sun is most familiar to you: it rises every morning in the east, travels slowly in a big arc across the sky during the course of the day, and then sets in the west. It seems plausible to guess that the Sun doesn't go out of existence each night and get re-born again in the morning. Rather, it just keeps going in a big circle around the earth – we just can't see it at night.

You may or may not already know that the stars move in a very similar way. First of all, the stars move in some sense *as one* – that is, the stars all move in the same way such that their relative positions with respect to each other are the same, night after night after night. (This is why there are recognizable collections of them such as constellations.) If you pick some one star and follow its motion throughout the night, you'll find that it does pretty much what the Sun does: it will rise in the east, travel in a big arc across the sky during the night, and then set in the west. And, just as with the Sun, it seems pretty obvious that the stars are "out there" during the day, too: the ones you can see at night are not visible during the day because they're below the horizon, and the ones that are in principle "out" during the day are simply too dim to see when the Sun is also out.

Actually, it's a bit misleading to say that the stars move the same way the Sun moves. Lots of them do. But stars in the extreme southern part of the sky just barely come up over the horizon to the south – they don't so much rise in the east and set in the west, as rise *just east* of directly south, and then set *just west* of directly south – and they're only up for a short period of time around midnight. And then the stars in the north behave rather differently, too: they go in the same kind of big circle we talked about the Sun and other stars going in, but (unlike the Sun and other stars) their circular paths *never dip below the horizon*. See Figure 1.1

An important point here is that there is a particular star – the *north star* or Polaris – which doesn't move at all. And all the other stars move in circles *centered on* the north star. The north star (assuming you live in the Northern Hemisphere – if you don't, you won't be able to see the North Star at all) is above the horizon to the north

Figure 1.1: A several-hour-long exposure taken from Hawaii, facing north, during the night. Note that the stars move in concentric circular paths – and hence leave circular "star trails" – centered at a certain point on the sky (the north celestial pole) that is very close to Polaris, the north star.

by an angle equal to your latitude. Hence, if you live down in the South, the north star will be $25 - 30°$ above the horizon, while if you live up near the Canadian border it will be $45 - 50°$ above the horizon. (If you lived at the North Pole, the north star would be directly up – $90°$ above the horizon!) Any star whose angle from the north star is less than the north star's angle above the horizon will never rise or set, but will always be "up." (Of course, you won't be able to see it during the daytime!) And those stars which are *farther* than this from the north star will rise and set like the Sun. And presumably there are even stars which are so far south that they *never* rise above the horizon. Indeed, if one were to travel south – to Mexico, say – there would be new stars visible in the southern part of the sky which weren't visible from Vermont.

We'll come back shortly to talk in more detail about the precise motion of the stars, Sun, Moon, and planets. But it is helpful at this stage to introduce the so-called "two sphere model" of the cosmos, to summarize what has been said so far. The model is just this: the Earth is a sphere, and then all the stars are (so to speak) painted on the interior of another big sphere which rotates around and around the earth. See Figure 1.2.

We'll develop this model in more detail shortly. For now, the main point is just this: in contrast to terrestrial objects like rocks and people which seem to have *rest* as their natural state, the stars (and Sun and Moon) seem to incessently, "naturally" *move in*

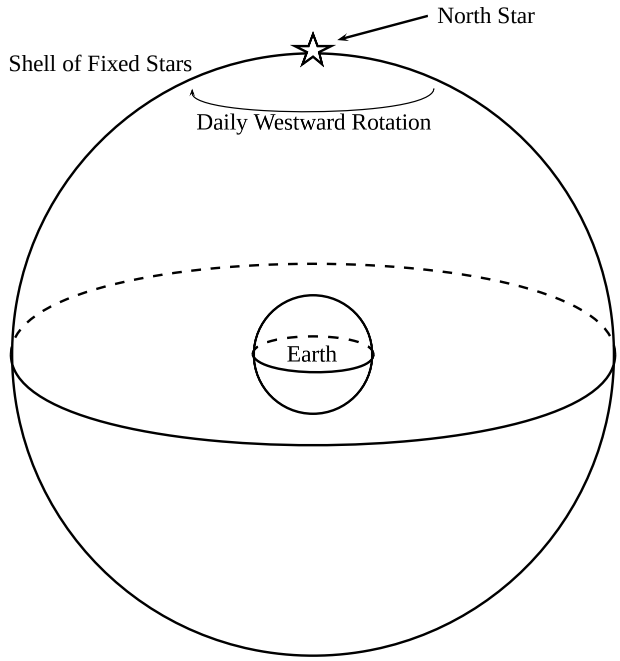

Figure 1.2: The two sphere model of the universe. The outer sphere is the shell of the fixed stars, which rotates westward once per day (or technically every 23 hours 56 minutes) around the Earth, which is at rest in the middle.

circles. This is a pretty fundamental difference, and it became, for the Ancient Greeks, the basis for the idea of a fundamental dichotomy between the earthly realm and the astronomical or heavenly realm.

It is worth briefly summarizing here what we might call Aristotle's Cosmology. This is just the basic worldview (including the heavens as part of the world) that was more or less accepted by Ancient Greek thinkers around the time of Aristotle (circa 350 BC) and was then orthodoxy until the Copernican revolution some 2,000 years later.

The whole cosmology is best illustrated with a picture. The first thing to note is the radial lines and concentric circles, which are meant to be a kind of polar-coordinates "cosmic graph paper." The earth sits at its center, and the other (outer) sphere contains the stars. The Greeks had figured out that the Sun was closer to Earth than the stars, and that the Moon was closer to Earth than the Sun, so the Sun and the Moon occupy intermediate positions between the Earth and stars. Indeed, the sphere of the Moon functioned as a kind of dividing line between heaven and Earth.

The Greeks believed in four terrestrial elements: earth, air, fire, and water. One can think of these as basically standing for what we now think of as the three phases of matter (solid, liquid, and gas), plus fire, which didn't seem to fall under any of those categories. (Actually, sometimes people say that "plasma" is a fourth state of matter, and fire is indeed a kind of plasma.) We talked above about the concept of a "natural

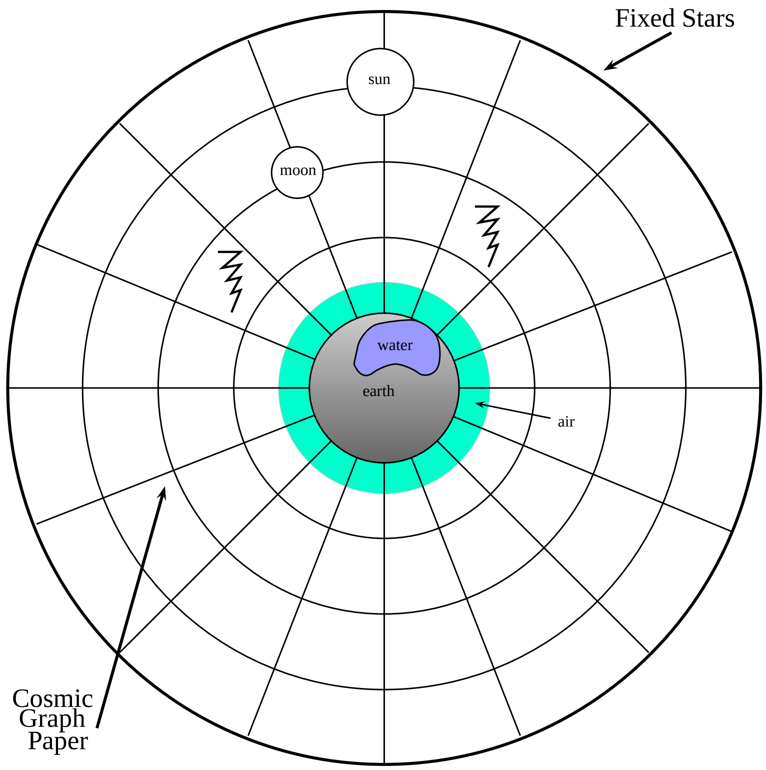

Figure 1.3: Simple sketch of Aristotle's universe. The matter (earth, water, air, fire, and aether) arranges itself relative to a fixed background space (the "cosmic graph paper") with the earth clumped up nearest the center, then water, then air, then fire, and finally (above the sphere of the Moon) the aether making up the Sun, stars, and other heavenly bodies (not shown).

state of motion" for various kinds of things. The Greeks also inferred, from familiar sorts of observations, that in addition objects had natural *places*. For example: rocks sink in water, whereas air bubbles up through water to the surface. Water poured from a pitcher falls down toward the ground, whereas fire tends to move up (and pull other things with it). It is not crazy to infer that everything spontaneously "wants" to move down, but that there's a kind of hierarchy: earthy things like rocks are, so to speak, more desperate to "get down" than water, which is in turn more desperate to "get down" than air, which is in turn more desperate than fire.

Positing something like that as a basic physical law can explain why the Earth has the basic structure that it has. It is essentially a big ball of (what else?) earth – a ball being the most efficient way for as much as possible of the earthy material to get as close as possible to the center of the universe (where the radial lines of the "cosmic graph paper" all converge). The water also likes to be as close as possible to the center of the universe (but gives way to earth) and so it tends to pool up in low spots on the surface

– hence oceans, lakes, etc. And then the air also wants to be as close as possible to the center, but it gives way to both earth and water and so forms a fairly uniform blanket around the earth called the atmosphere. And evidently there should be some fire up in the outer regions of the atmosphere – e.g., lightning!

Of course, all these materials aren't in equilibrium. Things are, apparently, constantly getting churned up. Water somehow gets pulled up into the atmosphere for a while (in the form of clouds) before eventually falling back down to earth; there is evidently some fire trapped down here in things like trees, which gets released back up toward its natural place in the process we call burning; volcanoes occasionally spew earthy and watery and fiery junk up into the atmosphere where it lingers for a while before slowly moving back toward its natural place; etc. But as a rough account of why the materials should more or less arrange themselves (from the bottom up: earth, water, air, fire) as they appear to do, this all makes good sense.

As already mentioned, the idea is that the outer sphere with the stars on it just rotates around and around and around. The Greeks actually believed in a *fifth element* called "aether" – basically the stuff that stars and other heavenly bodies were supposed to be made of. And just as the natural, unforced state for the terrestrial elements is rest (namely, rest in as close to its "natural place" as it can get to, given the constraint that all the other hunks of stuff are also trying to achieve their natural places, too), the natural state for aether is circular motion. So that explains why the sphere of stars just turns around and around.

The other heavenly bodies (Sun, Moon, and planets) move in a rough way just like the stars move – around and around in circles. But their motion isn't quite as *perfectly* circular as that of the stars. They drift, slowly, relative to the stars. The Greeks thought of this as a sort of imperfection, and thought that, as one moved in from the outer sphere of stars, the heavenly bodies were increasingly corrupted with a little bit of the terrestrial elements (e.g., it does seem like maybe the Sun has a lot of fire in it since it's so warm and so bright), resulting in motion that is roughly circular (due to the aether) but imperfectly so (due to the fire, earth, etc.).

In Aristotle's cosmology, there is also an interesting dynamical connection between the heavens and the Earth. He believed that the incessant rotation of the outer sphere of stars was *mechanically* responsible for pretty much all the other motion in the universe. The rotation of the stars pulled the planets around, which in turn pulled the Sun around, which in turn pulled the Moon around, and then the Moon in turn communicated this motion down into the Earthly (sub-lunary) region. This also has a kind of plausibility to it: for example, the sloshing of ocean water we call tides correlates with the motion of the Moon and Sun, as do the daily warming and cooling of the Earth, and yearly progression through the seasons. So the "churning" of the Earthly regions of the cosmos discussed above is maintained, according to the Greeks, by the motion of objects up in the heavens.

This is all really interesting, but for us it's just background. So let's move along and discuss now in more detail the motion of the Sun, Moon, and planets.

1.2 Astronomical observations in more detail

Let's jump off from the two sphere model mentioned above, and see how the Sun, Moon, and planets can be incorporated in a way that is consistent with their actual motion. Note, to begin with, that in this model the sphere of stars rotates in about a day (which we can define as the amount of time between noon on one day and noon on the next day). But it is not *exactly* a day. If one measures carefully the amount of time between the rising of some bright star (Sirius, say) on two subsequent nights, it will be about 4 minutes less than 24 hours: 23 hours and 56 minutes. This amount of time is called a "sidereal day." So, in this model, the outer sphere of stars rotates around and around, always and uniformly toward the *west*, once every 23 hours and 56 minutes.

Now let's discuss, in turn, the Sun, Moon, and planets.

1.2.1 The Sun

The first thing to say about the Sun is that it moves with the stars. If one only watches for a few days and/or doesn't watch too carefully, one would probably think that the Sun is just sitting on some particular spot on the sphere of stars, and hence rotating around with it as it rotates. But more careful observation reveals that this isn't quite right. The Sun moves just a bit each day relative to the stars – in particular, it slides just about 4 minutes to the *east* each day. Here's what that means: compared to (say) the rising of Sirius, the Sun will rise (on average) 4 minutes *later* each day than it did the day before. Or one can turn it around the other way. Since, in fact, we measure time in terms of the Sun, we could describe the relative motion of the Sun and stars by saying that the stars are all a little bit further *west* each night, compared to the previous night at the same time. So if, say, Sirius was exactly to the south at midnight on some particular night, it would be just a bit to the west of south at midnight on the next night. And how far to the west exactly? Four minutes to the west – meaning, the distance that Sirius moves across the sky in 4 minutes.

Note here an amazing numerical coincidence:

$$4 \text{ minutes} \times 365 \approx 24 \text{ hours} \tag{1.2}$$

What does this mean? Well, if one prefers to think of the Sun as moving (each day) a little bit east relative to the stars, it means this: after an entire year, the accumulated slightly-late-rising of the Sun (relative to Sirius) will have it rising late by exactly a day – which means, it'll actually be rising in exactly the same *place* relative to the stars, and at the same sidereal *time* as at the beginning of the year. In effect, the Sun will have "gone all the way around" and come back to its original location in the stars.

This periodic motion of the Sun is what defines the year: astronomically speaking, the year is the amount of time it takes the Sun to slowly wander all the way around a big circle through the stars (through the 12 constellations of the zodiac, actually) and come back to where it started.

So, during a year, the Sun moves through a closed circular path through the stars. There is a technical name for this path: it is called the *ecliptic*. Thus, the Sun is always

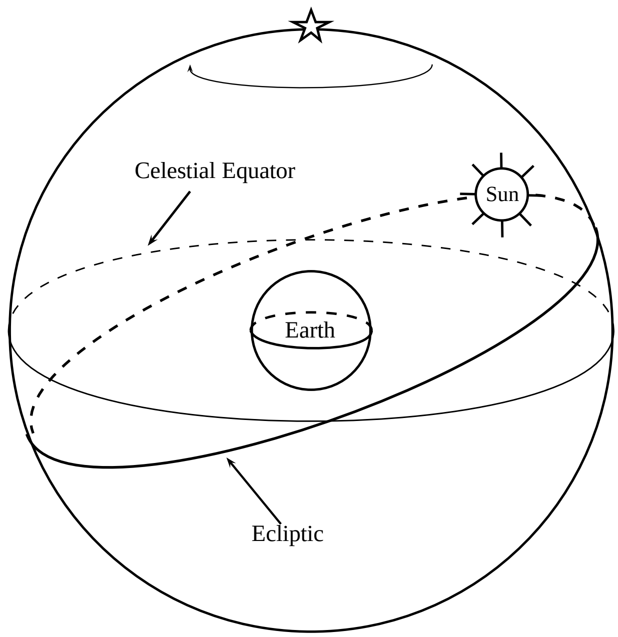

Figure 1.4: The two-sphere model of the universe, now including the Sun and the ecliptic (the Sun's path through the fixed stars). Note that the ecliptic and the celestial equator can be thought of as circles – specifically, intersections of the sphere of fixed stars with a plane passing through the center of the earth. The plane that cuts through the celestial equator and the plane that cuts through the ecliptic make approximately a 23.5° angle with one another. Thus, during the course of the year, the Sun is as much as 23.5° north of the celestial equator, and as much as 23.5° south of the celestial equator. It reaches these two extremes on the Summer and Winter Solstices, respectively, and goes back and forth between them in between, crossing the equator on the spring and fall equinoxes.

located somewhere on the ecliptic, but slowly moves to the east along the ecliptic during the course of a year.

It is useful to define another, related path through the stars called the "celestial equator." This is defined as all the points on the star-map that are exactly 90° away from the north star. If you are at the north pole of the earth, the north star will be straight above you and you'll just be able to see the stars on the celestial equator at the horizon. If you are at the earth's equator, the celestial equator will form a big arc across the sky, from directly east, to straight up above you, to directly west. If you're in the continental United States, the north star is about halfway between the horizon to the north and straight up, and the celestial equator forms an arc going from directly east, through a point about 45° up from the horizon to the south, and then over to directly west. I say all of this mostly to encourage you to get some kinesthetic feel for how the diagram in Figure 1.4 relates to the real world around you.

It is very important to understand that the ecliptic and the celestial equator aren't

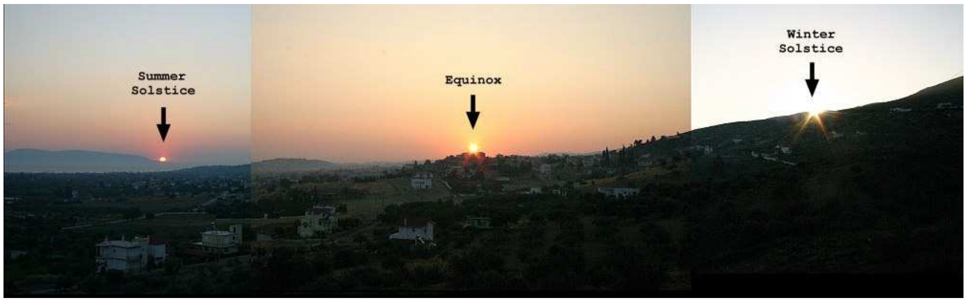

Figure 1.5: This is a composite of three photos taken at sunrise from the exact same location, at three different times during the year: the summer solstice, the winter solstice, and one of the equinoxes. The Sun is rising just to the east at the equinox, a bit north of east at the summer solstice, and a bit south of east at the winter solstice.

the same. If they were, then the Sun would basically follow the same path through the sky each day throughout the year. But they're not, so it doesn't. The ecliptic is "tilted" by about 23.5° relative to the equator. So one day each year, the Sun is a whole 23.5° south of the celestial equator – which means it will be very low on the horizon even at noon, and will rise and set considerably south of (respecitvely) east and west. This day is called the winter solstice. Since the arc of the Sun between rising and setting is considerably less than half a circle, the length of the daytime (the time between rising and setting) is considerably less than 12 hours.

There is then a corresponding day about 6 months later (the summer solstice) when the Sun is 23.5° further *north* than the celestial equator. On the summer solstice, the Sun rises and sets considerably north of east and west, and its arc takes it almost directly overhead at noontime. And since its arc is considerably greater than half a circle, the time between the rising and setting of the Sun is considerably greater than 12 hours.

During the rest of the year, the daily arc of the Sun slowly interpolates back and forth between these two extremes. The midpoints of this yearly cycle – when the Sun is at one of the two points where the ecliptic *crosses* the celestial equator – also have special names: the vernal (or spring) and autumnal (or fall) equinoxes. On these two days, the Sun rises and sets *exactly* to the east and west, and there is exactly half a day (12 hours) between Sunrise and Sunset.

To summarize: it is the motion of the Sun along the ecliptic which gives rise to the *seasons*. During the summer, the days are long and the Sun is close to directly overhead at noon, so it tends to be warm. And during the winter, the Sun is low on the horizon even around noon and the days are short, so it tends to be cold. The yearly progression of temperatures and weather (and all of the biological and ecological phenomena this cycling gives rise to) can be explained by the fact that the ecliptic is tilted relative to the celestial equator!

(Just for the record, some of what I've said here only applies to the middle latitudes of the northern hemisphere. The seasons work a bit differently if you go far enough south or north. But I'll leave that for you to puzzle out.)

The take-home point here is pretty simple: the Sun basically moves each day the same way the stars do, but not exactly. We can abstract out the extra motion of the Sun by thinking about its motion *relative to the stars*. And this extra, relative motion is also pretty simple: the Sun moves around in a circle, the ecliptic, with a period of exactly one year.

1.2.2 Moon

In all but a few details, the motion of the Moon is exactly like the motion of the Sun. The Moon shares the (rough) daily motion of the Sun and stars, and also shares with the Sun a more subtle motion with respect to the stars: it, too, moves along the path through the stars called the ecliptic. The only difference is that the Moon moves eastward along the ecliptic *faster* than the Sun. Where the Sun takes a whole year to complete its circuit around the ecliptic, the Moon takes only about a month (27.3 days to be precise).

The Moon exhibits one other unique feature, too: phases. During its monthly circuit around the ecliptic, the Moon alternates between "full" (when the entire circular disk of the Moon is illuminated) and "new" (when the entire circular disk is dark). These phases are readily explainable by the assumption that the Moon's light is not intrinsic, but reflected light from the Sun. The Moon presents as full when it is just opposite the Sun in the sky, such that, from here on earth, it is precisely the bright side of the Moon that is visible. New Moon occurs when the Moon is very close to the Sun in the sky, meaning that the side of the Moon that is illuminated faces away from earth, with only the dark side being "visible" from here. And so forth for all of the intermediate (crescent, half, gibbous) phases. Note that this explanation requires that the Moon be closer to the earth than the Sun.

It is worth mentioning here another occurrence involving the Moon: eclipses. There are two types. A lunar eclipse happens when the Moon passes through the shadow cast by the earth and thus appears dark for a short period of time right around full Moon. (Do you see why a lunar eclipse can only happen at full Moon? It doesn't happen *every* full Moon because the Moon's path isn't *exactly* along the ecliptic – it is rather within a couple of degrees of the ecliptic, but this is enough that most of the time it doesn't pass directly through the earth's shadow.) The other type of eclipse is a solar eclipse. This occurs when the Moon gets right between the earth and the Sun, so that the view of some or all of the Sun is blocked. And, again, this doesn't happen *every* time there is a "new" Moon, because the Moon's path is only roughly along (within a couple of degrees of) the ecliptic.

1.2.3 Other Planets

The Moon and Sun have several things in common as against the stars. First, unlike the stars, they are not *fixed* in their positions relative to (other) stars. Rather, they move (more or less) slowly through the stars, along the ecliptic. And second, the Sun and Moon just look different than stars: stars look like little points of light, while the Sun and Moon both present a large disc.

Careful observation of the heavens, however, reveals several additional objects which *look* like stars (in the sense of the second point just mentioned) but which have the first point in common with the Sun and Moon. That is, these objects look like little points of light (though they are typically as bright as some of the brightest stars), yet their positions are not fixed. Like the Sun and Moon, they *wander*. There are five such "planets" (from the Greek word for "wanderer") that are visible to the naked eye and hence were known about by the Greeks: Mercury, Venus, Mars, Jupiter, and Saturn. Actually, not surprisingly, the Greeks tended to think of all *seven* of the wandering objects we've talked about as "planets."

Not only do these additional five planets, like the Sun and Moon, wander – they wander in much the same way. Each of them (in addition, of course, to sharing the daily rotation of the stars) moves with a roughly-steady eastward drift along the ecliptic. Mercury and Venus each move around the ecliptic in (on average) one year, just like the Sun. The other three planets take longer: about two years for Mars, about twelve years for Jupiter, and about thirty years for Saturn.

The Greeks basically just assumed that the correlation between distance-from-earth and ecliptic-period which held for the Moon and Sun, continued to hold for the other planets as well. So they inferred that the seven planets had distances from earth in the following ascending order:

- Moon
- Mercury, Venus, Sun (order ambiguous!)
- Mars
- Jupiter
- Saturn

Note that Mercury, Venus, and the Sun cannot be placed unambiguously on this list because they all take, on average, exactly one year to go around the ecliptic.

To understand why it is necessary to say "on average" we must clarify a further important detail about the observed motion of the planets. Whereas the Sun and Moon *always* move eastward along the ecliptic, the five planets only do this most of the time. They also occasionally stop their eastward drift, move for a short period of time *to the west*, and then stop and return to their normal eastward motion. This bizarre behavior is referred to as "retrograde" (backward) motion. Each planet retrogrades at regular, periodic intervals, but the period varies from planet to planet. Saturn does it once every 378 days, Jupiter does it every 398 days, Mars does it every 779 days, Venus every 584 days, and Mercury every 116 days. So there is no obvious correlation here between the (distance) order of the planet and their frequencies of retrograding.

Another curious feature is that the planets are not uniformly bright. A given planet (say, Mars) is sometimes brighter and sometimes dimmer than its average brightness, and (curiously) the planets Mars, Jupiter, and Saturn achieve their maximum brightness just as they retrograde.

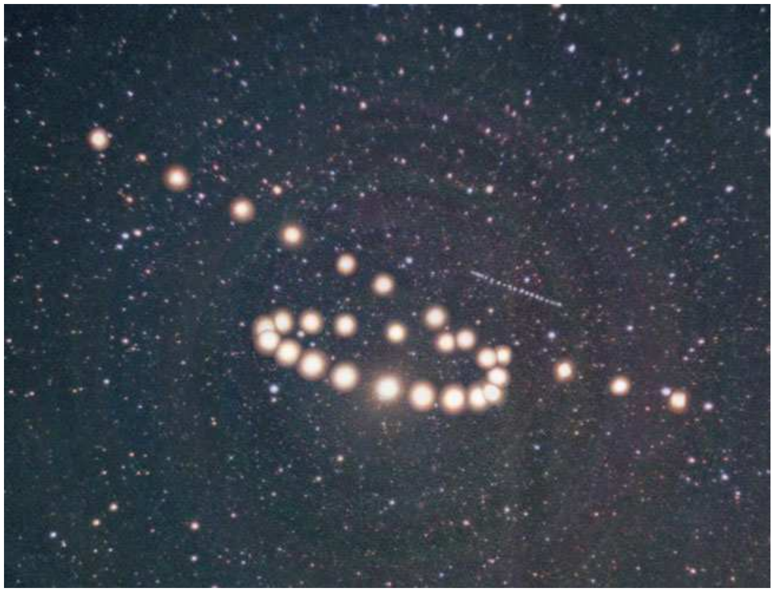

Figure 1.6: A sequence of images of Mars, stacked so that the stars line up in each frame. Mars begins on the right, moving left (to the east). But over the course of several days, Mars reverses direction, moving for a while westward, only to eventually return to is "normal" eastward drift along the ecliptic. Notice that the motion relative to the stars here is not *exactly* along the ecliptic – there is some motion in the orthogonal direction in the figure. The discussion in the text thus over-simplifies things a bit by ignoring this other aspect of the motion. One can also observe in the figure that Mars achieves maximum brightness during its retrograde motion. Finally, notice that there is some other planet in the background, which also leaves a trail through the stars.

It is a little harder to determine the brightnesses of Venus and Mercury, since both planets are always near the Sun in the sky. Think of this in terms of their motion along the ecliptic. Most of the time, Venus moves eastward along the ecliptic at a rate just a little faster than the rate of the Sun. But then, when it gets about 45° ahead of the Sun, Venus reverses direction and moves for a time *westward* along the ecliptic, until it is about 45° behind the Sun, at which point it resumes its eastward motion. It's important that the motion is "centered" on the Sun: the "normal" eastward motion of Venus along the ecliptic always has it catching up to and then overtaking the Sun, and then it passes it again in the other direction as it retrogrades, only to start over again (584 days later). Mercury does basically the same thing, only it goes back and forth faster, and doesn't get as far away from the Sun on either side.

One of our main projects for the week will be to examine some more detailed data about all of these things, and figure out how to incorporate it into a theory about how these things move. The point is: if you're a little confused and fuzzy (or just plain overwhelmed by the confusing complexity of these planetary motions), it's OK. We'll spend the week trying to get clear on all this.

1.3 Measuring the distance to the Sun and Moon

It is rather amazing that the Ancient Greeks had figured out the sizes of and distances to the Sun and Moon. (It's amazing because it seems like, and is, a pretty sophisticated kind of discovery which, one might naively guess, people would only have figured out in the last couple hundred years. But it is also amazing in a kind of opposite way: it is interesting that one could discover such geometrical facts about the heavenly bodies when one was nevertheless so wrong about such fundamental facts as whether the Sun moved around the earth or vice versa!) For that reason alone, it's worth spending a few minutes to understand how they figured these things out. But the fact that (especially) the distance to the Sun was known, becomes an important part of the story here – as we'll see in due course.

So how did the Greeks figure this stuff out? Well, to begin with (and contrary to what one sometimes hears in the context of Christopher Columbus), the Greeks knew that the earth was round and they even knew how big it was. There are several pieces of evidence for the earth's round shape that they knew. Here is the Greek astronomer Ptolemy's summary:

> "the more we advance towards the north pole, the more the southern stars are hidden and the northern stars appear. So it is clear that here the curvature of the earth covering parts uniformly in oblique directions proves its spherical form on every side. [Also], whenever we sail towards mountains or any high places from whatever angle and in whatever direction, we see their bulk little by little increasing as if they were arising from the sea, whereas before they seemed submerged because of the curvature..." (I.4)

Another piece of evidence has to do with lunar eclipses: the earth always casts a *circular* shadow, and so must itself be shaped like a ball.

Anyway, once it is established that the earth is spherical, it becomes possible to measure the size of the sphere. This was first accomplished by the Greek astronomer Eratosthenes (273-192 BC). Assuming (not arbitrarily) that the Sun is far enough away that its rays can be treated as parallel when they reach the earth, Eratosthenes arranged the experiment illustrated in Figure 1.7. The city of Syene (now Aswan, Egypt) is on the Tropic of Cancer (23.5° north latitude), which means that each year on the summer solstice, the Sun at noon would be directly overhead – as evidenced by the observable fact that a vertical pole would cast no shadow at that moment. By contrast, 500 miles to the north, in Alexandria, a vertical pole *does* cast a shadow at noon on the summer solstice. It is easy enough to measure the height of the pole (h) and the length of the shadow (s), and hence determine the angle θ according to

$$\sin(\theta) = \frac{s}{h}. \tag{1.3}$$

Eratosthenes found that $\theta = 7.2° = 0.126$ radians.

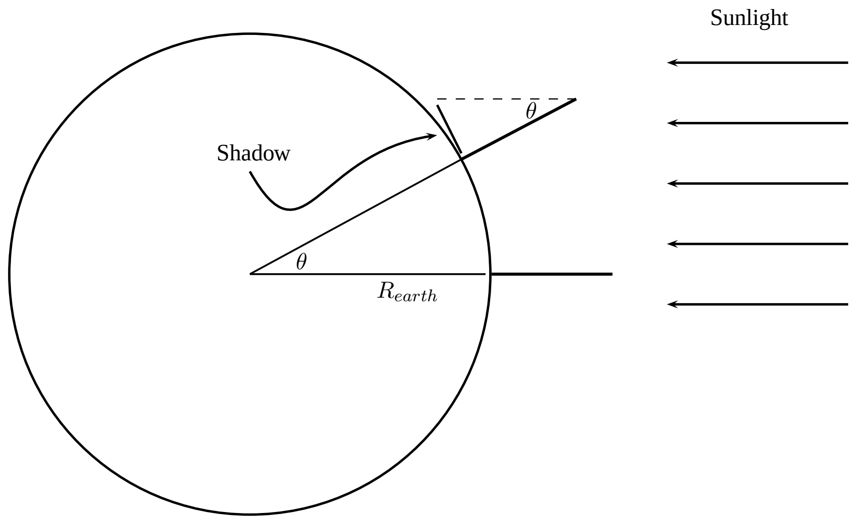

Figure 1.7: Schematic diagram of Eratosthenes' method of measuring the size of the earth. It is assumed that the Sun is far enough away that the incoming light rays can be treated as parallel. Then, one can determine the angle θ by measuring the height of a pole and the length of the shadow it casts (and using some trigonometry). And then, as is clear from the geometry in the figure, the angle θ is also the difference in latitude between the two locations. And so, since the distance (along the surface of the earth) between the two locations can also be measured, the radius of the earth R_{earth} can be calculated.

Also, as the geometry illustrated in the Figure makes clear, the angle θ which the Sun's rays make with the pole in Alexandria, is the same as the angle between Syene and Alexandria as measured from the center of the earth. And so the radius of the earth could be computed from what is essentially the definition of angle in radians:

$$\theta = \frac{D}{R_{earth}} \tag{1.4}$$

where $D = 500$ miles is the distance between Syene and Alexandria along the (curved) surface of the Earth). The result is

$$R_{earth} = \frac{D}{\theta} = 4,000\,\text{miles} = 6.4 \times 10^6\,\text{meters}. \tag{1.5}$$

Once the radius of the earth is known, the sizes and distances to the Moon and Sun can be inferred from a series of observations which relate them to the size of the earth. Let's go through these in turn.

First, the distance to the Moon can be determined by observing what is called the "parallax" of the Moon – i.e., its slightly different apparent position (relative to the background fixed stars) as seen from two different points on earth, or, equivalently, from the same point on earth at two different times (after the heavens have rotated a bit).

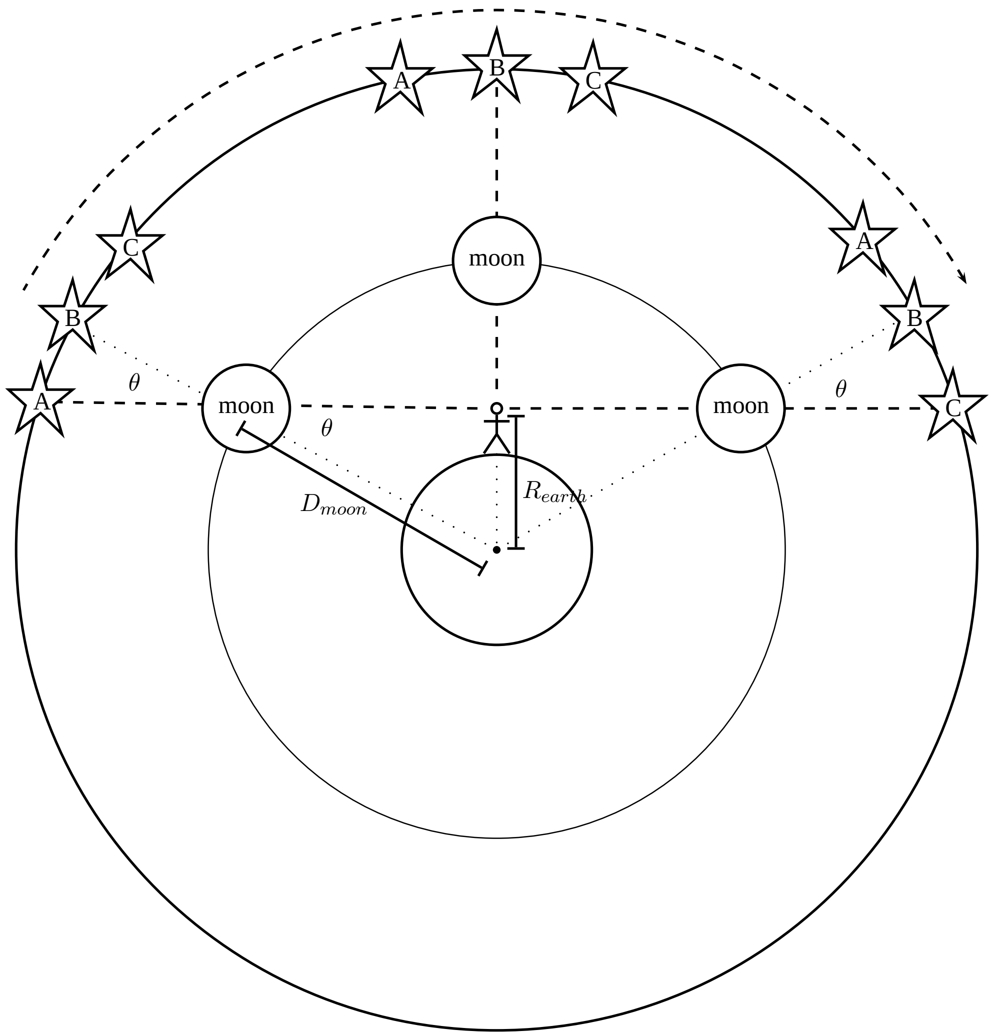

Figure 1.8: Due to the observer's location on the surface (as opposed to the center) of the earth, the Moon occupies three distinct apparent positions (relative to the fixed stars) at three times during the night. (Note that to highlight the phenomenon of parallax, the figure shows the Moon's "true position" – i.e., its apparent position as would be seen from a hypothetical observer at the center of the earth – as constant throughout the night, which in fact it isn't.) The parallax angle θ is the maximum deviation of the apparent position from the true position. And this measurable parallax angle relates the Earth's radius (R_{earth}) to the previously unknown distance to the Moon (D_{moon}) by simple trigonometry: $\sin(\theta) = R_{earth}/D_{moon}$.

Figure 1.8 shows schematically how the apparent position of the Moon will change over the course of the night due to the observer's position on the surface of, as opposed to in the center of, the earth.

Here is Ptolemy's description of the Moon's parallax:

> "since the distance from the earth's centre to the lunar sphere is not as that to the ecliptic circle which is so great that the magnitude of the earth is in the ratio of a point to it, therefore the straight line drawn from the Moon's centre to sections of the ecliptic, according to which the true courses of all the stars are conceived, necessarily does not everywhere sensibly coincide with the straight line according to which its apparent course is observed – that is, the straight line drawn from some part of the earth's surface or rather from the observer's eye to the Moon's centre. But when the Moon is directly above the observer, then only are the straight lines drawn from the earth's centre and the observer's eye to the Moon's center and the ecliptic one and the same straight line." (IV.1)

The parallax angle (θ in the figure) is the maximum deviation from the "true position" (i.e., what an observer at the center of the earth would see) due to the real observer's position on the surface of the earth. Careful observations revealed that the Moon displays a (maximum) parallax angle of just a little less than a degree: $\theta \approx 1^\circ$. The geometry of the figure – in particular the triangle formed by the Moon, the observer, and the center of the earth – then makes it clear that

$$\sin(\theta) = \frac{R_{earth}}{D_{moon}} \tag{1.6}$$

so it is possible to solve for

$$D_{moon} = \frac{R_{earth}}{\sin\theta} = 60R_{earth} = 240,000 \text{ miles} = 3.8 \times 10^8 \text{ meters} \tag{1.7}$$

You should be wondering: how is the parallax angle θ actually measured, since this is (as shown in the figure) an angle between two stars *as seen from the Moon*. The answer is: we are assuming here that (contrary to the not-to-scale figure!) the shell of fixed stars is very large compared to the circle representing the Moon's orbit. So the angle between stars A and B in the figure as seen from the Moon (i.e., the angle θ shown in the figure) will be the same as the angle between those stars as seen from here on earth. And that, of course, is readily measurable.

There is one other slight "cheat" here which is worth mentioning. In the figure, the Moon's "actual position" (namely: on top of Star B) doesn't change during the 12 hours or so between the three moments pictured. But in fact the Moon moves slowly and steadily relative to the fixed stars, along (roughly) the ecliptic. So you couldn't in fact measure the parallax of the Moon just by comparing the Moon's apparent position when it rises, and then later when it is overhead. You'd have to already have measured the Moon's *average* rate of motion along the ecliptic, so one could subtract off the part of the difference angle due to the motion of the Moon relative to the stars, leaving the parallax angle. One of the Projects at the end of the chapter steps you through this.

It is worth making a couple other comments about this method of determining the distance to the Moon.

First, the method relies on already knowing the size of the earth. In determining the size of the earth, we had to assume the Sun is very far away compared to the size of the earth. And then in measuring the Moon's parallax angle, we had to assume that the stars are very far away compared to the Moon. Is there any evidence to support these assumptions? Yes. First of all, the Moon displays an easily-noticable parallax of about one full degree. But no such parallax is observed in the Sun. (Well, it can be observed today, but the angle is so tiny the Greeks never noticed it.) So the Sun must be much further away than the Moon, and hence *very* far away compared to the size of the earth. And of course, since the shell of fixed stars is in this model the outside edge of the universe, the stars must be further away than the Sun! Or more directly: the stars also display no observable parallax. And so they, like the Sun, must be very far away compared to both the distance to the Moon or the size of the earth. Indeed, "that the Earth has the ratio of a point to the heavens" is one of the first points stressed in Ptolemy's book. He writes that "in all parts of the earth the sizes and angular distances of the stars at the same times appear everywhere equal and alike, for the observations of the same stars in the different latitudes are not found to differ in the least." (I.6)

Second, the parallax angle is in fact quite difficult to measure accurately, in part because the Moon is big and bright. A precise measurement of its position relative to the fixed stars requires some particular point on the Moon to be identified and tracked, and also requires that the background stars right next to the Moon can be seen – which is difficult because the Moon's brightness tends to overwhelm the surrounding stars. So it can be surprisingly tricky to measure the apparent position of the Moon, relative to the fixed stars, to an accuracy significantly better than a degree.

Given these sources of uncertainty, it is impressive that the parallax of the Moon can be measured at all. The implication is of course that the uncertainty on θ is pretty small compared to the reported value of θ. Actually, although Ptolemy reports a value for the parallax of the Moon and uses it to infer the distance to the Moon by essentially the argument described here, his description of the methods of determining the parallax angle (and hence that angle's uncertainty) leaves much to be desired by modern standards. In fact, he rather races through the argument presented here, and then proudly presents the reverse argument *from* the (now "known") distance to the Moon *to* the observable parallax angle. It is as if he is embarrassed to have to use observation to figure out how far away the Moon is, and so hurries through this, then lingering on what he considers more logically sound: using facts about the world to calculate appearances. We mention this here only to stress the very different approaches to science taken by the Ancient Greeks, as compared to modern *empirical* science – which of course is not at all embarrassed to base its conclusions about the world on observation.

So much for the determination of the distance to the Moon. Once this is known, it is relatively easy to determine the *size* of the Moon by measuring the angle ϕ subtended by the Moon, i.e., the angle between two opposite points on its edge, i.e., its angular diameter. The angle is easily measured to be about half a degree ≈ 0.0087 radians. Using the small angle approximation (according to which $\sin(x) = \tan(x) = x$), we then

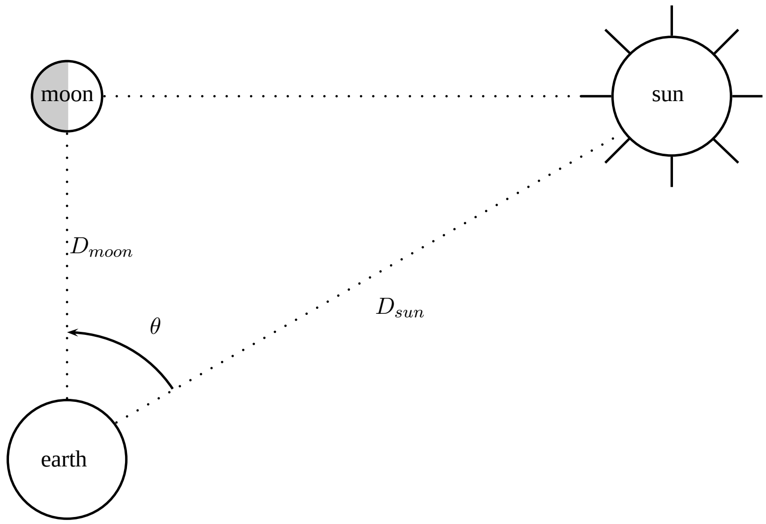

Figure 1.9: At half-moon, the earth, Moon, and Sun form a right triangle. Careful measurement of the angle between the Sun and Moon (θ in the figure) can therefore reveal via trigonometry the relative lengths of the sides of the triangle, i.e., the Sun's distance relative to the Moon's distance. This allows an absolute determination of the Sun's distance, since the Moon's distance is already known.

have

$$\phi = \frac{2R_{moon}}{D_{moon}} \tag{1.8}$$

where R_{moon} is the moon's radius. Plugging in the numbers,

$$R_{moon} = \frac{1}{2} D_{moon} \times 0.0087 \,\text{radians} = 1,000 \,\text{miles} = 1.7 \times 10^6 \,\text{meters}. \tag{1.9}$$

So the moon is about a quarter the (linear) size of the earth. (Its volume is thus the earth's volume times a quarter *cubed*, i.e., it is about a sixty-fourth as big as the earth in that sense.)

Now what about the Sun? In principle, one could determine its distance, and then its size, by following the method just outlined for the Moon: first measure its parallax angle relative to the fixed stars, and then measure its angular diameter. But as already noted, the Sun turns out to be considerably farther away than the Moon, giving it a considerably smaller parallax angle that was simply too small for the Greeks to measure. (Plus, it's *really* hard to see which stars are right next to the edge of the Sun!) So a different approach is needed.

The simplest approach, first used by Aristarchus (310 - 230 BC), is to measure the angle between the Moon and the Sun at half-moon. In order for the Moon to appear precisely half-illuminated, the angle between the Moon-earth line and between the Moon-Sun line must be precisely 90°. And so the three bodies form a right triangle. See Figure

Figure 1.10: This is a composite of several photos shows the Moon passing in front of the Sun during a solar eclipse. Note in particular that when the Moon is just over the Sun ("totality"), its disc just covers the disc of the Sun (revealing the fuzzy solar corona, which is too dim compared to the ordinarily-blinding solar disc to see except during a solar eclipse).

1.9. It is clear that

$$\cos(\theta) = \frac{D_{moon}}{D_{sun}} \tag{1.10}$$

or, equivalently, that

$$D_{sun} = \frac{D_{moon}}{\cos(\theta)}. \tag{1.11}$$

Aristarchus reported that $\theta = 87°$, which then implies that $D_{sun} = 20 D_{moon} = 1200 R_{earth}$.

It is clear, though, that since the angle is very close to 90°, the calculated distance to the Sun will be very sensitive to the measured angle. For example, using instead a value of 88° gives a distance ratio of about 30 instead of 20. In fact, accurate contemporary measurements reveal that $\theta = 89.85°$, which gives

$$D_{sun} = 400 D_{moon} = 24,000 R_{earth} = 93,000,000 \,\text{miles} = 1.5 \times 10^{11} \text{ meters}. \tag{1.12}$$

And finally, once the distance to the Sun is known, its size can be calculated from its apparent (angular) diameter. Just as with the similar calculation for the Moon, we have that the angular diameter ϕ is

$$\phi = \frac{2R_{sun}}{D_{sun}}. \tag{1.13}$$

It turns out that the angular diameter of the Sun is just equal to that of the moon, as evidenced most dramatically during a solar eclipse. See Figure 1.10. Thus, plugging in the same $\phi = .0087$ radians used above yields

$$R_{sun} = 6.9 \times 10^8 \text{ meters} \tag{1.14}$$

if one uses the correct, modern value for D_{sun}. The Greeks underestimated this value by about a factor of 20, and hence also underestimated the size of the Sun by about a factor of 20.

The actual radius of the Sun is thus about 100 times the radius of the earth. The Greeks thought it was a mere 5 times bigger. Even with their underestimate for its size, then, the Greeks (correctly) believed that the Sun was more than a hundred times bigger (in volume terms) than the earth. (In fact it is about a million times bigger.)

1.4 Ptolemy's Theory

Claudius Ptolemy (85 - 165 AD) was the most important of the Greek Astronomers, partly because he systematized and cataloged many of the things that had been done earlier by Eudoxus, Aristarchus, Eratosthenes, and Hipparchus (whose writings have largely been lost). But Ptolemy also helped develop and improve the kinds of observations we were cataloging above, and he systematically developed a theory that had been posited earlier to to explain and integrate in particular the observed motions of the planets.

The Ptolemaic theory basically starts with the two-sphere model described previously, and incorporates the Sun, Moon, and the other planets in roughly the way we've already suggested. So, according to Ptolemy's theory, the earth is at rest at the center, with a big rotating sphere of fixed stars on the outside. The Sun, Moon, and planets are placed in the region between the earth and the stars, in the order we've already mentioned. To begin with, each of these seven planetary objects is pulled around (some way or other, either mechanically or just mathematically) by the rotating sphere of stars. This accounts for the shared gross daily motion of all the heavenly bodies.

1.4.1 Epicycles

Ptolemy's major innovation was a clever scheme for accounting for the details of the planets' motions – namely, their average eastward drift along the ecliptic, punctuated (for all but the Sun and Moon) by occasional retrograde motions. But to understand why this scheme was clever, one must first appreciate what Ptolemy was trying to do. Remember here the apparent ubiquity of *circular motion* for the heavenly bodies: not only do they all move in circles each day, but the extra motion of the planets relative to the stars is (at least on average) also circular. Plus, these objects were conceived to be made of a substance (aether, the fifth element) whose natural motion was circular motion. So the problem – the assignment, if you will, which is usually attributed to Plato – was to figure out a way of explaining the detailed, observed motions of the planets *in terms of circles.*

Here is Ptolemy's own statement of the guiding principle of his work:

> "it is first necessary to assume in general that the motions of the planets in the direction contrary to the movement of the heavens are all regular and circular by nature, like the movement of the universe in the other direction. [....] But

> the cause of this irregular appearance [i.e., deviations from the just-stated first assumption!] can be accounted for by as many as two primary simple hypotheses. For if their movement is considered with respect to a circle in the plane of the ecliptic concentric with the cosmos so that our eye is the cenre, then it is necessary to suppose that they make their regular movements either along circles not concentric with the cosmos, or along concentric cricles; not with these simply, but with other circles borne up on them called epicycles. For according to either hypothesis it will appear possible for the planets seemingly to pass, in equal periods of time, through unequal arcs of the ecliptic circle which is concentric with the cosmos." (III.3)

To summarize, the idea is to account for the observed motions of the planets – including especially the fact that they *don't* just move uniformly along the ecliptic circle relative to the stars – by compounding or otherwise fiddling with circular motions. Ptolemy here mentions two devices for achieving this. The first, called the "eccentric", involves using circles whose centers are shifted away from the earth. We will explore this later, in the Projects at the end of the chapter. The second and more immediately important device is called the "epicycle" and involves letting the planet move around a *smaller, second circle* which is itself pulled around a circular orbit centered at the earth. There is also a third device, called the "equant", which we will also come back to later.

For now let's focus on the epicycles and see how these are useful in accounting for retrograde motion. The idea, to repeat, is to *compound* two circular motions for each planet. The first would account for the average easterly drift along the ecliptic, while the second would account for the occasional retrograde motion. The way it works is sketched in Figure 1.11. It should be clear (looking at the figure) how this compounding of two circular motions (one circular motion relative to another point which is itself undergoing circular motion) can give rise to precisely the sort of behavior observed for the planets. In particular, by adjusting the relative sizes of the two circles and the two speeds involved for each planet, one can match pretty well the observed motions.

A few things are worth noting. First, since the Sun and Moon never retrograde, the secondary circle ("epicycle") is only needed for the 5 other planets: Mercury, Venus, Mars, Jupiter, and Saturn. And second, although this scheme works pretty well to explain the observed gross motions of the planets along the ecliptic, it doesn't quite get all the details exactly right. So in fact Ptolemy's full theory for each planet involved all three of the devices mentioned before: not just an epicycle, but also an eccentric and an equant.

See the Projects at the end of the chapter for some more information about these extra devices; they do play an important role in understanding what was and what wasn't immediately seductive about Copernicus' heliocentric theory, when he proposed that a millenium and a half later. But for the moment we'll ignore these other two devices and work with the simple one-deferent-one-epicycle construction.

The basic idea is that there is a special point (the "deferent") which moves in uniform circular motion around the ecliptic, and then a second point (actually occupied by the planet) which moves around a second, smaller circle (the "epicycle") with uniform

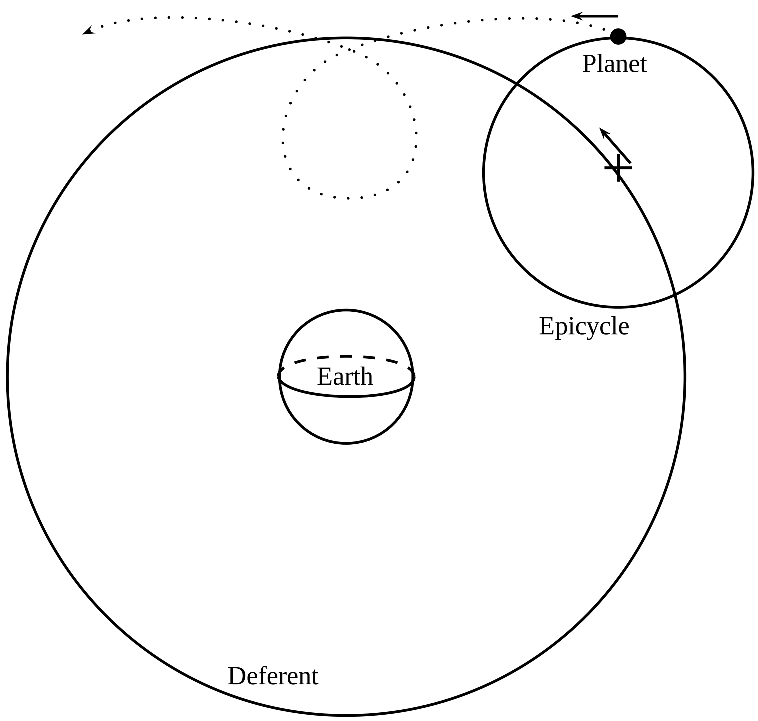

Figure 1.11: Sketch of the basic deferent-epicycle combination in Ptolemy's theory. The point marked + moves uniformly around the deferent circle, while the planet moves uniformly around the epicycle (which is centered at the + and pulled around the deferent as it moves). Both circles – the deferent and the epicycle – lie (approximately) in the plane of the ecliptic. This compounding of two circular motions gives rise to a trajectory like that sketched in the dotted line. As seen from the earth, the motion is generally counter-clockwise (which here means eastward along the ecliptic), but the occasional retrograde motion is also accounted for. Note too that the theory automatically explains the observed correlation between retrograde motion and brightness: the planet retrogrades when it is *closest* to the earth, which accounts for its increased apparent brightness.

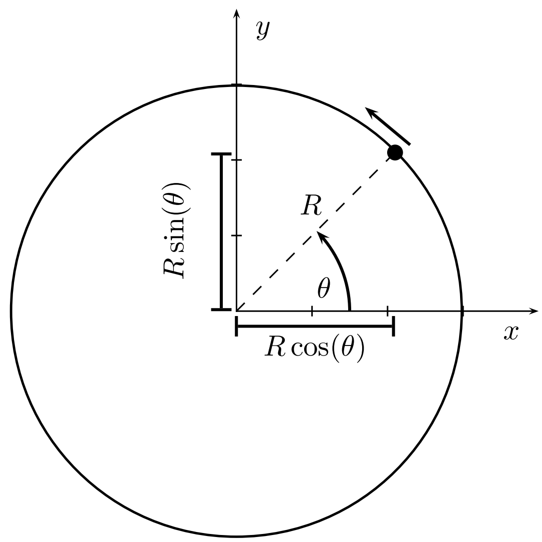

Figure 1.12: An object (the black dot) moves along a circle of radius R. Its position can thus be specified by the angle θ that the object makes with (say) the x-axis. The object's x and y coordinates are then given by $x = R\cos(\theta)$ and $y = R\sin(\theta)$. The object's angular velocity ω is given by the rate at which θ increases in time: $\omega = d\theta/dt$.

circular motion relative to the deferent point. It is worth developing some mathematical technology for dealing with all of this.

1.4.2 Angular Kinematics

Let's begin with simple uniform circular motion, described in terms of angle. For an object moving in a circle, its position can be described in terms of the angle as shown in Figure 1.12. For *uniform* circular motion the angle just increases linearly in time:

$$\theta(t) = \theta_0 + \omega t. \tag{1.15}$$

This should remind you of a corresponding equation from 1-D translational kinematics: $x(t) = x_0 + vt$. In accordance with this analogy, the quantity ω – which evidently describes the rate at which the angle θ increases – is called the *angular velocity*. Not surprisingly, it can be defined (generally, i.e., not necessarily assuming that the angular motion is uniform) as follows:

$$\omega = \frac{d\theta}{dt}. \tag{1.16}$$

This is just parallel to the familiar definition of (translational) velocity in terms of position: $v = dx/dt$.

For an object moving with constant angular velocity, its motion will be *periodic* – that is, it will repeat itself over and over again. Since a complete circuit comprises 2π radians and occurs in an amount of time we'll call T (the period of the motion), the angular velocity for uniform angular motion can also be written: $\omega = 2\pi/T$. And this is equivalent to

$$T = \frac{2\pi}{\omega}. \tag{1.17}$$

Note that this last formula only makes sense if ω is constant in time.

Following the analogy with 1-D translational kinematics, we may also define the *angular acceleration*

$$\alpha = \frac{d\omega}{dt} = \frac{d^2\theta}{dt^2}. \tag{1.18}$$

as the rate of change of the angular velocity. Since the mathematics is all completely parallel, we can immediately steal some familiar results from 1-D kinematics. For example, if an object moves with *constant* angular acceleration α, its angular coordinate will evolve in time according to:

$$\theta(t) = \theta_0 + \omega_0 t + \frac{1}{2}\alpha t^2 \tag{1.19}$$

where θ_0 and ω_0 are, respectively, the angle and angular velocity at $t = 0$.

1.4.3 Angular and Rectangular Coordinates

Let's now consider how to relate the angular coordinates of a moving object to its rectangular coordinates. Assume (again) that the object is moving in uniform circular motion with angular velocity ω and initial angle θ_0. And let's choose the origin of our coordinate system to lie at the center of the circle the object moves along. Then simple trigonometry gives

$$x(t) = R\cos[\theta(t)] = R\cos[\omega t + \theta_0] \tag{1.20}$$
$$y(t) = R\sin[\theta(t)] = R\sin[\omega t + \theta_0] \tag{1.21}$$

where R is the radius of the circle.

Probably the simplest way to think about Ptolemy's scheme for explaining retrograde motion in terms of epicycles, is in terms of *vector addition*: the position of a given planet relative to the earth is found by adding two vectors – one representing the position of the deferent point, and the other representing the position of the planet relative to the deferent point. And, of course, if we want to add these two vectors to figure out where the planet is relative to earth, the easiest way to do that is by adding (separately) the x and y components of the two vectors.

Let's work through this in detail. Suppose the deferent for a given planet has radius R_d and moves with angular velocity ω_d. Then the x, y coordinates of the deferent (as functions of time) will be

$$x_d(t) = R_d\cos(\omega_d t + \theta_0^d) \tag{1.22}$$
$$y_d(t) = R_d\sin(\omega_d t + \theta_0^d). \tag{1.23}$$

Likewise, suppose the *epicycle* for this planet has radius R_e and angular velocity ω_e. Then the x, y coordinates of the planet (relative to the deferent point!) will be

$$x_e(t) = R_e \cos(\omega_e t + \theta_0^e) \quad (1.24)$$
$$y_e(t) = R_e \sin(\omega_e t + \theta_0^e). \quad (1.25)$$

And so the x, y coordinates of the planet *relative to the earth* will be given by the vector sum:

$$x(t) = x_d(t) + x_e(t) = R_d \cos(\omega_d t + \theta_0^d) + R_e \cos(\omega_e t + \theta_0^e) \quad (1.26)$$
$$y(t) = y_d(t) + y_e(t) = R_d \sin(\omega_d t + \theta_0^d) + R_e \sin(\omega_e t + \theta_0^e) \quad (1.27)$$

Of course, the Greeks only knew how to measure the *angle* of a planet along the ecliptic. And this angle is related to the x and y coordinates as follows:

$$\tan[\theta(t)] = \frac{y(t)}{x(t)} = \frac{(R_d/R_e)\sin(\omega_d t + \theta_0^d) + \sin(\omega_e t + \theta_0^e)}{(R_d/R_e)\cos(\omega_d t + \theta_0^d) + \cos(\omega_e t + \theta_0^e)} \quad (1.28)$$

So this is, in a way, the fundamental equation of the Ptolemaic theory. It connects the observational angle $\theta(t)$ – the angular coordinate of the planet on the ecliptic at time t – with the parameters in the theory (R_d, ω_d, θ_0^d, R_e, ω_e, and θ_0^e) which give rise to the planet's detailed motion.

Note that this basic equation has been written so that only the *ratio* R_d/R_e appears. This makes it clear that, from data about the ecliptic angle $\theta(t)$, one is never going to be able to find the absolute sizes of the deferent and epicycle – one can only find the relative size of the one circle relative to the other. This makes sense, since all the angles will be the same if both circles are (say) doubled in size. (Draw a picture if that isn't clear.)

If these parameters were known (for each planet), the above equation makes it clear that it would be possible to predict exactly how the observational ecliptic angle would vary in time. Of course, in reality, it works the other way: the problem is not to figure out how the planet will appear to move relative to the stars given the detailed parameters about its deferent and epicycle; rather, the problem is to figure out how big the epicycle and deferent radii and angular velocities need to be in order to account for the observed angular position, $\theta(t)$. We'll be spending quite a bit of time this week (in the Projects) figuring out how to do this. And the numbers that come out have some interesting surprises hidden in them.

1.5 The Precession of the Equinoxes

So far we have discussed the daily rotation of the heavens as a whole, and then also the motions (more or less along the ecliptic) of the Sun, Moon, and the other 5 planets. But there is one additional motion that was first discovered by Hipparchus. Here is Ptolemy's summary:

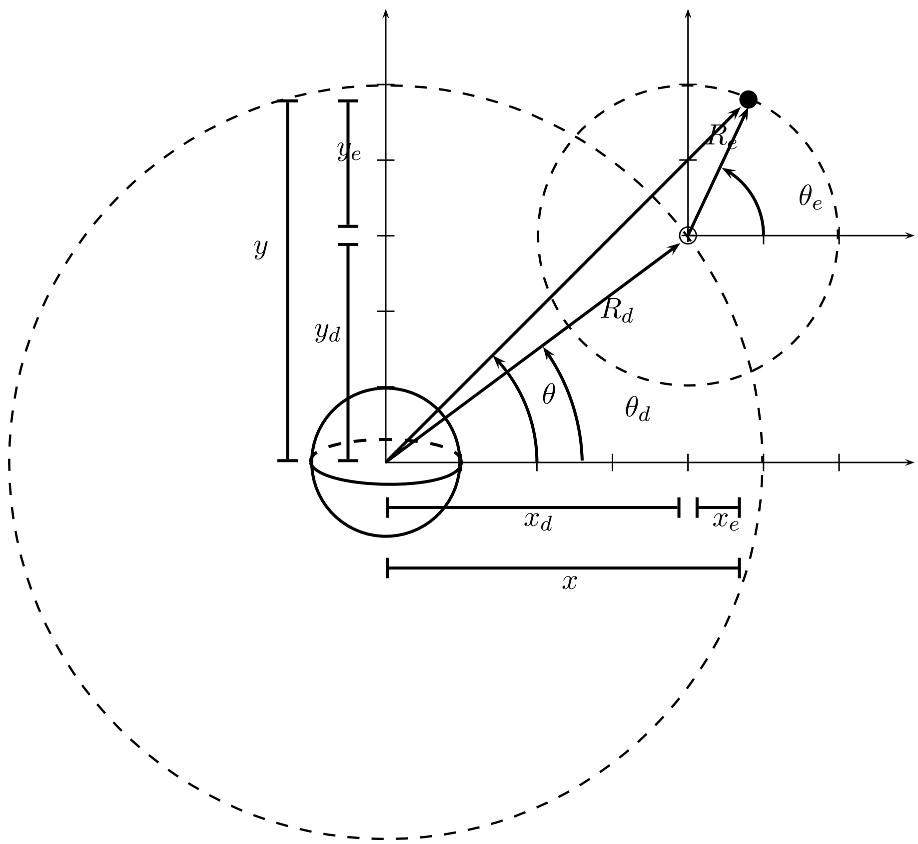

Figure 1.13: How the observable angle θ of a planet on the ecliptic relates to the radii (R_d and R_e) of the deferent and epicycle, and the angles θ_d and θ_e.

> "Hipparchus [noted that] for the apparent returns of the sun with respect to the tropics and equinoxes, the length of the year is found to be less than 365 1/4 days, but for its returns observed with respect to the fixed stars it is found to be more. And from that he conjectures that the sphere of the fixed stars also has a very slow movement, and like that of the planets is in the direction contrary to that of the prime movement which revolves the circle that passes through the poles of the equator and the ecliptic." (III.1)

Let us try to unpack what he is saying here.

There are two subtly different ways of defining a year, and they turn out not to be precisely equal. The first way – the so-called "sidereal year" – is defined as the amount of time it takes the Sun to complete a full circuit around the ecliptic and return to precisely the same spot relative to the fixed stars. The other way – the "solar year" – is defined as the amount of time between (say) two subsequent spring equinoxes. Based on what has been explained so far, one would expect these to be equal, since the ecliptic and the celestial equator have been described as *fixed* paths through the fixed stars, and the equinoxes occur when the Sun is at the intersection of these two paths. So equivalent to the claim that the sidereal and solar years are not quite equal, is the claim that either the ecliptic or the celestial equator *moves*!

And that is just right. It turns out that the ecliptic really is fixed (in the sense that, even over long periods of time, the Sun always moves along the same path, i.e., in front of the same set of fixed stars). But the celestial equator *moves* (very slowly) through the fixed stars. And since the celestial equator is just all those points on the sky that are perpindicular to the *celestial pole*, a simpler way to understand this claim is that the celestial pole moves! We described this at the beginning of the chapter by saying that there is a particular star (the "north star" or Polaris) at (or, really, near) the celestial pole, i.e., the point that all the other stars move in circles around. So the point here is that this is true today, but was not true in the past, and will not be true in the future. Instead, the celestial pole is a moving target, that moves around (of course!) a *circle*, which is concentric with the ecliptic.

This slow motion of the celestial pole means that the sidereal and solar years will be slightly different: the Sun will be at a certain point on the ecliptic on the spring equinox one year, but the next spring equinox will occur *slightly before* the Sun returns to that same particular spot on the ecliptic in the following year. That is, the solar year is slightly shorter than the sidereal year. Or put the same point again this way: the location of the Sun on the ecliptic *at the moment of the spring equinox* slowly moves (westward) along the ecliptic. And this is also equivalent to saying that the sphere of fixed stars as a whole not only rotates once per day about an axis perpendicular to the plane made by the earth's equator, but that it also "wobbles" or "precesses" such that the intersection of this axis (passing through the geographic poles of the earth) with the fixed stars moves slowly to the east around a circle $23.5°$ down from the pole of the ecliptic.

How fast is this precessional motion?

> "having sighted Spica and the brightest stars about the ecliptic .... we find their distances with respect to each other very nearly the same as those observed by Hipparchus [i.e., the arrangement of the stars themselves has not changed in the several hundred years between Hipparchus and Ptolemy] but their distances with respect to the tropic and equinoctial points to have shifted eastward very nearly 2 $1/2°$ compared to the record of Hipparchus." (VII.2)

This 2 $1/2°$ of precession in the 2 1/2 centuries between Hipparchus and Ptolemy makes for an overall precession rate of approximately one degree per century. So this is indeed a small effect, noticable only with a long span of accurate measurements. With several additional centuries of observation to work with, the number has been more recently pegged at about $1.4°$ per century. Thus, in about $26,000$ years, the north celestial pole will again be near the star Polaris (having drifted quite far – $47°$ – away from it over this period).

This subtle motion will play an interesting role in the story to come.

1.6 Arguments against heliocentrism

That pretty much wraps up what there is to say about Greek Astronomy and in particular Ptolemy's geocentric theory for the motion of the planets. It should come as no surprise that, in the next chapter, we are going to discuss the transition to the alternative heliocentric theory proposed in 1543 by Copernicus. But in understanding that transition, it will be helpful to know that Copernicus wasn't the first to propose a heliocentric theory. In all but some details, Copernicus' theory had been proposed already in Ancient Greece by Aristarchus, who is thus sometimes referred to as the Greek Copernicus!

An account of Aristarchus' proposal is given by another great Greek scientist, Archimedes, in his book *The Sand Reckoner*:

> "You King Gelon are aware the 'universe' is the name given by most astronomers to the sphere the centre of which is the centre of the earth, while its radius is equal to the straight line between the centre of the Sun and the centre of the earth. This is the common account as you have heard from astronomers. But Aristarchus has brought out a book consisting of certain hypotheses, wherein it appears, as a consequence of the assumptions made, that the universe is many times greater than the 'universe' just mentioned. His hypotheses are that the fixed stars and the Sun remain unmoved, that the earth revolves about the Sun on the circumference of a circle, the Sun lying in the middle of the orbit, and that the sphere of fixed stars, situated about the same centre as the Sun, is so great that the circle in which he supposes the earth to revolve bears such a proportion to the distance of the fixed stars as the centre of the sphere bears to its surface."

That is, in this heliocentric theory, it is not the sphere of fixed stars but the *earth* which (gasp!) rotates around once per day. Likewise, the yearly motion of the Sun around the ecliptic is accounted for by a yearly orbit of the earth around the Sun. It is sometimes speculated that Aristarchus proposed the heliocentric system not only because he saw it was an alternative (and arguably simpler) way of accounting for the observed apparent motion of the heavens, but also because he knew that (as discussed earlier) the Sun was significantly larger than the earth – and somehow it seemed more plausible for the smaller object to orbit around the bigger object, than vice versa.

Why does the heliocentric theory require the universe to be "many times greater" than implied by the geocentric theory? Recall that the only direct observational evidence bearing on the size of the sphere of fixed stars was the lack of observed parallax.

We discussed above how the parallax of the Moon could be used to measure the distance to the Moon. In principle, one could use the same effect to measure the distance to the stars, if only their parallax could be detected. Or, since no parallax was detected, this placed a lower limit on the distance to the stars. Assuming that, using the whole diameter of the earth as a baseline, the stars' parallax is less than (say) a hundredth of a degree (i.e., 0.000175 radians, which would definitely prevent it from being noticed by the Greeks), one can infer that the distance to the stars must satisfy

$$D_{stars} > \frac{2R_{earth}}{.000175\,\text{radians}} = 50,000,000\,\text{miles}. \tag{1.29}$$

Remember that (due to not-very-accurate measurement of the angle between Sun and Moon at half-moon) the Greeks thought it was only about $5,000,000$ miles to the Sun. So it was conceivable to them that the outer edge of the whole universe – the sphere of fixed stars – was only about one order of magnitude farther out than the Sun. This maybe seemed about right, as it left just about the right amount of room for the shells of Mars, Jupiter, and Saturn.

But if the earth orbits the Sun (instead of vice versa), the baseline for parallax increases dramatically! Instead of just being able to look at the stars from opposite sides of the earth, one can look at the stars from opposite sides of the earth's orbit – which the Greeks thought to have a diameter of some $10,000,000$ miles. And with *that* baseline, the condition on the size of the shell of fixed stars (required to account for their lack of observable parallax) becomes

$$D_{stars} > \frac{10,000,000\,\text{miles}}{.000175,\text{radians}} = 57,000,000,000\,\text{miles}. \tag{1.30}$$

This is really an incredible distance, about a thousand times bigger than the earlier geocentric estimate (because they thought the distance to the Sun was about a thousand earth radii). To the Greeks, it seemed impossible that the universe could be so vast, so much bigger than the biggest other things known. And hence it seemed to them impossible that the earth could orbit the Sun. So this was one major reason why, despite being proposed much earlier, the heliocentric worldview proposed by Copernicus was not taken seriously by the Greeks.

Another reason has to do with the several *motions* of the earth required by the heliocentric system. If the earth is spinning around on its axis once per day, then the surface of the earth near the equator must be moving at an incredibly large speed:

$$v = \frac{2\pi R_{earth}}{24\,\text{hours}} \approx 1,000\,\text{miles/hour}. \tag{1.31}$$

The Greeks thought: wouldn't we *notice* this? Shouldn't we have to hold on for dear life just to keep up with the earth as it moves? And, for example, wouldn't this imply a constant westerly wind of about this same speed? And if you dropped a rock, wouldn't it fall straight down (as rocks are known to do) *while the earth raced out from under it to the east* such that the rock would hit the ground a mile or more to the west of where it was dropped? As Ptolemy summarizes,

> "never would a cloud be seen to move toward the east nor anything else that flew or was thrown into the air. For the earth would always outstrip them in its eastward motion, so that all other bodies would seem to be left behind and to move towards the west." (I.7)

Yet none of these fantastic implications are in accordance with actual experience. So the earth cannot possibly be rotating in the way suggested by the geocentric model.

And note that all of these same objections can be made again – and with even more force – in relation to the alleged yearly motion of the earth around the Sun, which evidently requires the earth as a whole to be moving with a speed

$$v = \frac{2\pi D_{sun}}{1\,\text{year}} \approx 20\,\text{miles/second}. \tag{1.32}$$

Thus, for example, a ball thrown high up in the air such that it stays up for 5 seconds should evidently land 100 miles away. Or more precisely: *we'd* be 100 miles away when the ball came back down to the same place it had been thrown from. And so forth.

Another argument against the heliocentric system makes (even more explicit) reference to the Aristotelian physics ideas we began with: even if the earth were somehow displaced from the center of the universe, it would simply return to the center. That, after all, is what the laws of physics sketched in the beginning of this chapter require. In order to perform the sorts of circular motions attributed to it in the heliocentric model, the earth would have to be made of aether, not earth – which is clearly preposterous.

It is easy to laugh at this sort of objection. It's maybe less easy to take it seriously and answer it clearly. So it's something to puzzle over for yourself before we cover the arguments of Copernicus and his followers in the following chapter.

Questions for Thought and Discussion:

1. What other observations can you think of that are consistent with – and might have been taken as evidence for – the Greek cosmology sketched in Figure 1.3? Can you think of any observations that are definitely contrary to this picture?

2. Orient yourself spatially, i.e., figure out which way is north. Now indicate with a sweep of the arm the path that the Sun will take across the sky today. Now indicate (again with your arm) how this path changes over the course of the year.

3. Describe how the daily trajectory of the Sun across the sky would vary over the course of the year if you were in Ecuador (which, of course, is on the equator). How about at the north pole? What is the significance of the Tropic of Cancer and the Arctic Circle (in the northern hemisphere, and correspondingly the Tropic of Capricorn and Antarctic Circle in the southern)? What are the latitudes of these "special" points, and why are those numbers significant?

4. Suppose it is the Winter Solstice, so the Sun's arc across the sky during the day is low and short. What can you say about the path taken by the Moon across the sky – and/or about the duration of time between the Moon's rising and setting – *that night*? What if in addition you are told that it happens to be full Moon? In general, how does the path/duration of the Moon – around full Moon – relate to the path/duration of the Sun, over the course of the year?

5. It was mentioned in passing in the text that the Moon can only exhibit the full range of observed phases (new, crescent, half, gibbous, full) if the Moon is closer to the earth than the Sun. Explain why. For example, suppose to the contrary that the Moon and Sun both orbited earth in circles, but the Sun's orbit was smaller/closer than the Moon's. What range of lunar phases would then be observed?

6. It was mentioned in passing in the text that, unlike the Moon, the stars do not exhibit any parallax (that would have been noticable to the Greeks). Explain, by reference to Figure 1.8 or some diagram of your own, what it would mean for the stars to display parallax. (Parallax for the Moon or Sun means a change in the apparent position of that body *relative to the background fixed stars.* How could the fixed stars appear to move relative to themselves?)

7. You should understand how the deferent-epicycle device sketched in Figure 1.11 can account for the retrograde motion of a planet. But look more carefully at Figure 1.6. What aspects of the actual motion of Mars *cannot* be accounted for by the deferent-epicycle device as shown in Figure 1.11? How could this be fixed?

8. As you probably know, lunar and solar eclipses are somewhat rare. They certainly don't happen as often as every month. What does this imply about the deferent circles for the Sun and Moon in Ptolemy's system? In particular: does the deferent circle for the Sun lie in the plane of the ecliptic? Does the Moon's?

9. Can observers at different locations on earth disagree about whether a lunar eclipse is "total" (i.e., whether all of the Moon enters the earth's shadow cone)? Can observers at different locations disagree about whether a solar eclipse is total (i.e., whether the Sun is completely covered by the Moon)?

10. Argue using vector addition that the apparent motion of a planet is the same if (a) it moves on an epicycle of radius R_e and angular velocity $\omega_e = 0$, and (b) it moves around the deferent circle directly, but with the center of the deferent circle displaced by a distance R_e from the position of the earth.

11. In the Ptolemaic theory discussed in this chapter, the moon, Sun, and 5 other planets all orbit around the earth. We discussed explicitly how the distances to the moon and Sun could be measured. What about the distances to the other planets? Could the Greeks have measured these? Why or why not? What other sorts of arguments could be used (within the Ptolemaic system) to estimate or put bounds on the size of the universe?

Projects:

1.1 Reproduce Eratosthenes' measurement of the size of the earth, perhaps by coordinating with another class in a different part of the country.

1.2 Your teacher will provide some data for the apparent angular position of the moon, along the ecliptic, over the course of several days. Do a linear curve-fit to find the average angular velocity of the moon. Is it what you expect? Look at the residuals of your fit and explain qualitatively what gives rise to the obvious feature. (The data include one unrealistic feature, which is that exact positions for the Moon are given throughout the course of several days, even when the Moon is below the horizon!) Suppose the data were taken from near the Equator when the Moon's position in the stars was close to one of the Equinox points (i.e., when the moon was at the same location relative to the stars that the Sun occupies at one of the Equinoxes). What is the distance to the Moon?

1.3 Here is another way to determine the relative distances to the Sun and moon. (Ptolemy discusses the method and credits it to Aristarchus.) During a lunar eclipse, the Moon passes through the shadow cast by the earth. Because the Sun is bigger than the Earth, the region of complete shadow "behind" the Earth is shaped not like a cylinder, but like a *cone*. From a knowledge of the size and distance of the Moon – and from observing the size of the Earth's shadow at the position of the Moon during an eclipse – the "slant" of the cone can be determined, and one can hence work out the distance to the Sun.

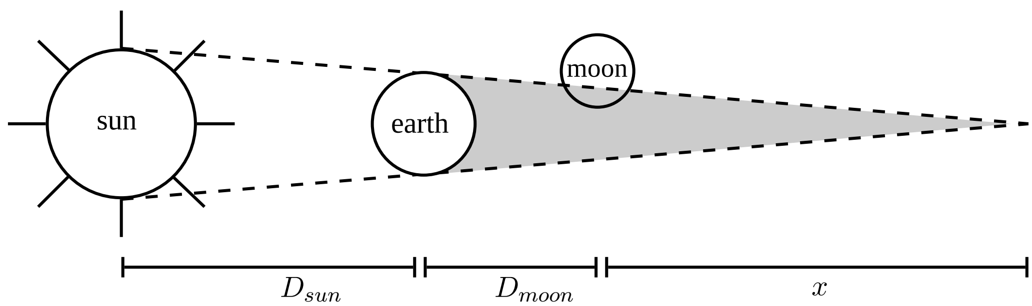

Figure 1.14: During a lunar eclipse, the moon passes into the conical shadow cast behind the earth.

Figure 1.14 shows the relevant geometry, and Figure 1.15 shows the relevant kind of observation. Use Figure 1.15 to estimate the relative size of the shadow compared to the Moon (e.g., is the shadow twice as wide as the moon, or 2.5 times as wide, or what?). Be as precise as you can. Then use the already-known relationship between the Moon's radius and the Earth's radius to determine how the size of the shadow (at the Moon's distance) compares to the size of the Earth. Knowing also already the distance to the moon allows one to then calculate the angle that the edge of the shadow slants at in the Figure. And this now known slant angle can then be related to the size and distance to the Sun. Finally, combine this relation with Equation 1.13 to yield an expression for the distance to the Sun. There are some tricky aspects to this, which should warrant further thought and discussion. In particular, it is worth considering whether this method is more or less reliable than the method described in the main text.

1.4 Your teacher will provide some data for the angular position (along the ecliptic) of the Sun, over the course of several years. Use Excel to make a graph of the angle vs time. Then use Excel to compute the angular velocity vs time. What is the average value of the Sun's angular velocity? What is the period of its motion (relative to the stars)? Are these numbers related the way they should be? Does the Sun

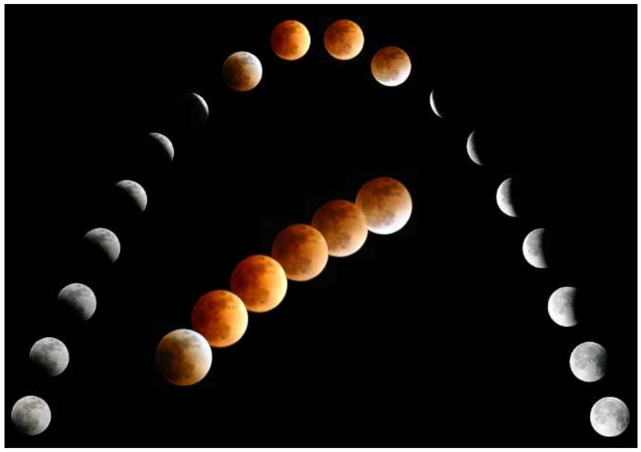

Figure 1.15: A mosaic of pictures showing the moon passing through the earth's shadow during a lunar eclipse. By looking at one of the shots of the moon where it is approximately half-shadowed, it is clear that (a) the shadow has a round edge (helping establish the round shape of the earth), and (b) that the size of the shadow is bigger than – one might guess roughly twice as big as – the size of the moon itself. Photo by Anthony Ayiomamitis, http://antwrp.gsfc.nasa.gov/apod/ap040506.html

move with constant angular velocity? Describe qualitatively any deviations from constancy.

1.5 Your teacher will provide some data for the angular position (along the ecliptic) of a planet such as Mars, over the course of several years. Use Excel to make a graph of the angle vs time. Then use Excel to compute and graph the angular velocity vs time, and the angular acceleration vs time. Be able to explain in words any interesting features of the graphs, and how the graphs relate to one another.

1.6 Putting the Sun in a circular orbit around the earth with constant angular velocity accounts for its motion pretty well. But it is not exact. You maybe noticed in Project 1.4 that the angular velocity is not exactly constant, but is rather sometimes a bit faster and sometimes a bit slower than usual. One way to account for this behavior in Ptolemy's system is by having the Sun move in a circular orbit with uniform angular velocity relative to the center of the circle – but displacing the center of that circle somewhat from the earth. This construction is referred to as an "eccentric" circle, and was in fact used by Ptolemy not only for the Sun, but also for the moon and the other 5 planets. Let's see how it works for the Sun.

Use the triangle in the figure to write an expression for ϕ (the observable angular position of P on the ecliptic) in terms of θ. Then use the fact that θ should increase according to $\theta(t) = \theta_0 + \omega t$ (with ω a constant) to write an expression for $\phi(t)$. It should depend on c/R, and ω. Finally, do a curve-fit with the actual data for the Sun to find the best values for these parameters. How good is the fit? Examine

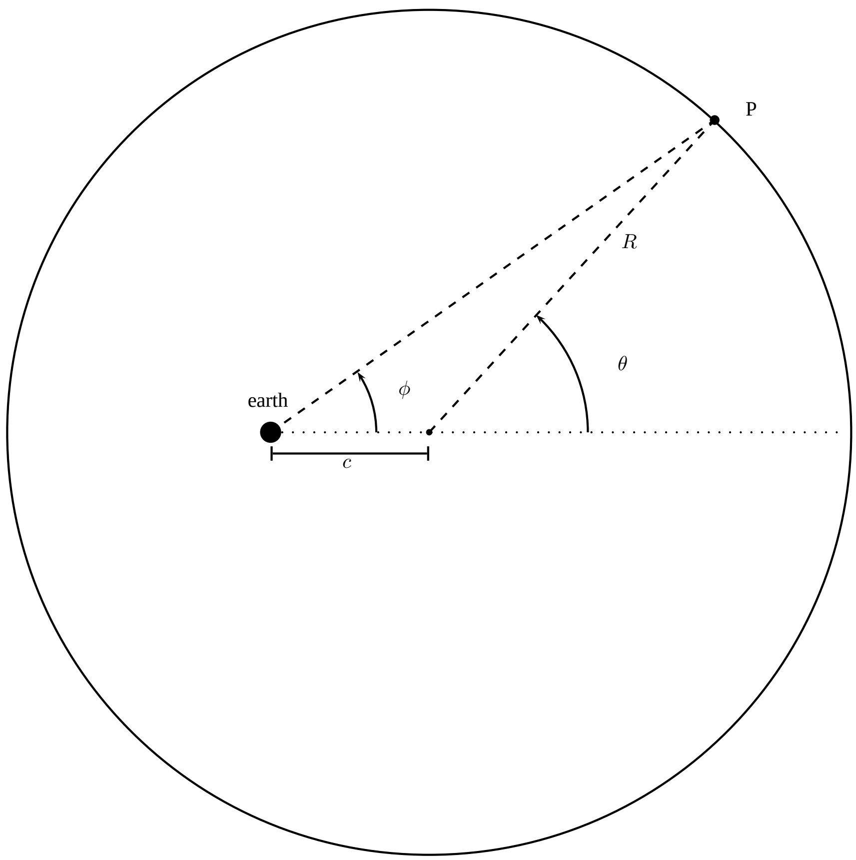

Figure 1.16: An eccentric circle. The point P (which represents a planet, here the Sun) moves around the circle with uniformly-increasing θ. But since the center of the circle doesn't coincide with the earth, the observed angle of P relative to the stars will be ϕ, which will increase in a not-quite-uniform way.

the residuals. Are there still systematic errors?

1.7 Following up again with the Sun: hopefully you found that an eccentric orbit can reproduce quite accurately the data for the Sun's progression along the ecliptic to within an accuracy of about a tenth of a degree. Let's now explore how well you can do with the final scheme developed by Ptolemy for accounting for "anomalous" motion of the planets: the "equant." Here the motion of the Sun is around a circle centered at the earth, but the angular velocity is constant with respect to a point called the equant which is off to the side by some distance. Draw a simple diagram to help you work out how the observed ecliptic angle of the Sun should vary in time in this scheme, and then do another curve-fit. You should find that you can do just about as well with the equant as with the eccentric (though the exact nature of the remaining small errors will be slightly different with the two devices). According to Ptolemy, the Sun has just one "anomaly" and it is hence a matter of choice whether one corrects this using an epicycle, an eccentric, or an equant. The 5 planets, however, turn out to have *three* anomalies each – so the fully detailed account of their motion required, in Ptolemy's system, all three devices for each planet! This turns out to play an important role in Copernicus' arguments for a heliocentric system.

1.8 Your teacher will provide some data for the angular position (along the ecliptic) of a planet, over the course of several years. Perform a curve-fit to find the values of ω_d, ω_e, and R_d/R_e – from Equation 1.28 – that provide the best fit to the data. (The data will be pre-arranged so that, at t=0, the planet will be in the middle of its retrograde cycle – so you can automatically set $\theta_0^d = 0$ and $\theta_0^e = \pi$.) Note also the time between subsequent retrogradings of your planet, the so-called synodic period. Be as accurate as possible and get at least two significant figures of precision for each parameter. Students should be assigned to different planets, so the class as a whole can compile and discuss the results for all the planets.