Contents

CHAPTER ONE

Classical Accounts of Space and Time

The Birth of Physics

The intellectual tradition that produced modern theoretical physics begins in ancient Greece. Babylonian and Egyptian astronomers compiled voluminous and accurate data about the visible positions of the sun and the planets and created mathematical models that could predict phenomena such as eclipses. But the Greek nature-philosophers introduced a novel strand of speculative theorizing into this observational enterprise. Thales, Anaxagoras, and Democritus, for example, each offered conjectures about the ultimate structure of matter: that material objects are all derived from water; are mixtures of earth, air, fire, and water; or are composed of an infinite variety of differently shaped atoms. The observable behavior of familiar objects was then explained in terms of this material constitution. According to Democritus, sweet things are composed of smooth, rounded atoms; sour things are composed of angular atoms; and so on. The notion that the perceptible properties and behaviors of large objects should be accounted for by the structure and nature of their imperceptible parts underlies physics to this day.

Aristotle provided this speculative enterprise with its name. The term "physics" derives from the Aristotelian text *Physike Akroasis*: Lectures on Nature. In Greek, *physis* refers to the nature of a thing, and Aristotle defined the nature of an object as an internal source of motion and rest that belongs to an object primarily and properly and nonaccidentally (*Physics* 192^{2}20–23). Thus, for Aristotle, the nature of an object is revealed by how the object moves, and stops moving, when left entirely to its own devices. Release a rock in midair, without pushing it in any direction, and (apparently) of its own accord it starts to move downward. A bubble of air in a tank of water spontaneously rises. The rock and the air can be forced to do other things, but only under the constraint of an outside agent. Their inborn predilections to motion and rest are not attributable to outside agents and so must arise from the very nature of the thing itself.

Aristotle's definition of the "nature" of an object does not readily connect physics with the project of explaining sweetness or sourness. The emphasis instead is on change in general and on locomotion in particular. Aristotle believed that different sorts of material have different natural motions, so to describe these natural motions, he needed a way to describe and categorize motion in general. He started with the most intuitive descriptions available. The element earth's natural motion is to fall—that is, to move downward. Water also strives to move downward but with less initiative than earth: a stone will sink though water, demonstrating its overpowering natural tendency to descend. Fire naturally rises, as anyone who has watched a bonfire can attest, as does air, but with less vigor.

It is all well and good to say that a stone naturally falls, or moves downward, but what exactly does "downward" mean? It is here that Aristotle leaves common opinion and begins theoretical postulation. To move downward, according to Aristotle, is to move *toward a particular place*. The natural motion of earth, on this view, is goal-directed: the stone wants to get someplace in particular, and its spontaneous motion always takes it closer to this ultimate objective. The special place that the stone strives to reach, according to Aristotle, is the center of the universe. Aristotle conceived of the whole material cosmos as forming a sphere, whose outer surface contains the fixed stars. The celestial sphere has a unique center. The "down" direction at any place in the universe is the direction toward that central point, and an unobstructed piece of earth will naturally move down in a straight line, toward the center, until it reaches that target. If it should manage to make it all the way to the center, the piece of earth will, of its own accord, stop moving.

Similarly, "up" is the direction in space that points directly away from the center. Fire and air naturally move upward in straight lines as far as they are able, with fire displacing air if they are in competition. According to Aristotle, if the sublunary sphere (the part of the universe below the orbit of the moon) were left entirely unmolested, all of the earth, air, fire, and water would naturally segregate out into four concentric spheres: pure earth at the center, surrounded successively by concentric spherical shells of water, then air, then fire. This provides a very rough approximation to the world as Aristotle believed it to be: a spherical rocky earth covered largely with oceans, surrounded by air.

The moon and sun and planets do not fit into this scheme, so Aristotle invented a fifth sort of substance, a quintessence, called *aether*. Unlike earth, air, fire, and water, aether does not naturally engage in straight-line motion toward some goal: its natural motion is uniform circular motion around the center. This motion is most perfectly realized by the sphere of fixed stars, which spins (as far as Aristotle knew) with perfect regularity, making one complete rotation in about 23 hours and 56 minutes (a *sidereal day*). The rest of the superlunary objects—the moon, and sun, and planets—are not so regular: as they are carried about by the sphere of fixed stars, they also execute their own more complicated periodic motions. Having identified uniform circular motion as the natural state for aether, Aristotle set a problem for astronomers of succeeding generations: account for the apparent motions of the sun, moon, and planets by some compound effect of different uniform circular motions. This basic constraint on astronomical theory remained in place until Kepler proposed his first two laws of planetary motion in 1609.

Unfortunately, even a comically inadequate sketch of the history of physics and astronomy is beyond our scope. But Aristotle's innovation, his focus on natural locomotion as the primary subject of physics, shapes the field to this day. Our first order of business is to understand what exactly "locomotion" is.

The term "locomotion" wears its meaning on its sleeve: it is not just any change but change of place (*locus*). And place, for Aristotle, is location in a spatial universe with a very special shape: a sphere. Because it is a sphere, Aristotle's universe contains a geometrically privileged center, and Aristotle makes reference to that center in characterizing the natural motions of different sorts of matter. "Upward," "downward," and "uniform circular motion" all are defined in terms of the center of the universe. If Aristotle's universe did not have a spherical shape, his physics could not have been formulated.

The importance of an account of space in the formulation of physics cannot be overstated. If physics is first and foremost about motion, and motion is change of place, then (it seems) there must be *places* that material objects can successively occupy. An object rests when it occupies the same place over time, like Aristotle's stone at the center of the universe. It is tempting to say that without some sort of space in which things move, physics cannot even get off the ground. Aristotle adopts the concept of space, and the correlative concept of motion, that we all intuitively employ. He realizes that his physics requires this space to have some particular structure—a target goal that falling objects are seeking—and postulates a physical geometry that provides this structure. The resulting finite spherical universe is foreign to us today but would have felt quite familiar to any ancient Greek.

In short, space is the arena of motion, and so an account of space will play a central role in any scientific theory of motion. Abandoning Aristotle's spherical universe entails abandoning his basic physical principles and rethinking the form that the laws of physics can take. This task was undertaken by Isaac Newton.

Newton's First Law and Absolute Space

If we were to axiomatize Aristotle's physics, there would be different axioms for different sorts of matter: "Earth, if unimpeded, will move in a straight line toward the center of the universe" and "Aether, if unimpeded, will move in uniform circular motion about the center of the universe." Newton did present his physics as a set of axioms, which he denominated Laws of Motion. A tremendous amount of theory is packed into these laws, and it is only a slight exaggeration to say that everything we need to know about Newtonian physics is implicit in his First Law of Motion:

> Law I: Every body perseveres in its state either of rest or of uniform motion in a straight line, except insofar as it is compelled to change its state by impressed forces.[1]

[1] My translation.

This single law smashes the Aristotelian universe to smithereens.

First, Newton's law governs *every* body: stones and planets alike. Newton obliterates the distinction between astronomy and terrestrial physics, postulating a single set of principles that explains the behavior of both. We have become so accustomed to thinking of physics as possessing this sort of universality that it takes an effort to appreciate how momentous this shift is. One of the crowning moments in the argumentative structure of the *Principia* occurs when Newton calculates that the force that maintains the moon in orbit about the earth is precisely the same force that causes an apple to fall from a tree. Newton postulates a commonality of physical structure where the tradition preceding him had seen fundamental diversity.

More importantly, Newton does not ascribe any particular natural motion to a body, as Aristotle did. Rather, the Law of Inertia attributes an innate tendency of every body to *maintain* its state of motion, whatever that might be. There is no place in the universe that any body is inherently "directed toward," as a stone is directed toward the center of the universe in Aristotle's account. Newton's theory does not require that space have a special central point.

The arena of motion for Newton is rather an entity he calls *absolute space*. Motion, for Newton, is change of location in this space. The role of absolute space in Newton's theory is so deep and pervasive that it seems impossible to make sense of anything he writes without accepting its existence. We will consider various properties of absolute space, leaving the most controversial claims for last.

First, Newton assumes that absolute space possesses the geometrical structure of three-dimensional Euclidean space. We will designate this structure E^3. E^3, unlike Aristotle's physical universe, is infinite in all directions and so has no geometrical center. Figures in E^3 obey the axioms of Euclidean geometry: for example, the sum of the interior angles of any triangle equals two right angles.

It will be useful in the following discussion to distinguish several different sorts of geometrical structure, which form a hierarchy. Each level corresponds to one of the three instruments used in Euclidean geometry: the pencil, the straightedge, and the compass. What sort of geometrical structure must a space have in order for each of these instruments to operate?

The most basic, fundamental level of geometrical structure in a space is called its *topology*. The topology of a space determines facts about *continuity*. For example, when we use a pencil in making a Euclidean diagram, we are supposed to draw *continuous* lines in the space: if we were to occasionally lift the pencil and then drop it down again elsewhere while drawing what should be a single line, we will not get a single, connected, continuous line. But in order for there to be a distinction between a single line in a space and a pair of disconnected lines, the points in the space must have some geometrical organization. This level of organization is the topology of the space.

Topology is sometimes called "rubber-sheet geometry," and the name is properly evocative. Suppose some figures are drawn on a rubber sheet, and then the sheet is stretched without tearing or pasting. Many of the properties of the figures will be changed under such deformations: straight lines can be bent into curved lines; nearby points can be pulled apart so they are more distant; a triangle can be deformed smoothly into a circle; and so on. But some features of the figures will remain unchanged: if two lines intersect before the deformation, they intersect afterward; if one point is in the interior of a closed figure and another outside before the deformation, they will remain so after; and so on. The deformations are not allowed to "tear" or "paste" the space, and topology provides the level of geometrical structure that defines what counts as "tearing" and "pasting." Tearing separates some continuous lines into discontinuous pieces, and pasting joins discontinuous lines into continuous wholes. If a space did not have a topology, then there could be no distinction between drawing a single continuous curve and drawing several disconnected curves, so Euclidean constructions could not even start.[2]

[2] In modern mathematics, the topology of a space is specified in terms of the *open set* structure of the space, and continuity is defined in terms of the open sets. I believe that this account of continuity and hence topology is not the most perspicuous way to describe the intrinsic geometrical structure of space-time, and have developed an alternative (see Maudlin [2010] for an overview). This is not the place to fight that battle.

The second instrument of Euclidean geometry is the straightedge (N.B.: not the *ruler*; the straightedge has no marked scale). With a straightedge and a pencil, we can draw not just continuous lines but straight lines. The first two postulates of Euclidean geometry concern the use of the straightedge and hence make implicit claims about the structure of straight lines in Euclidean space. In particular, the first two postulates state:

1. It is possible to draw a straight line from any point to any point.
2. It is possible to extend any finite straight line continuously in a straight line.[3]

In order for these postulates to obtain, there must first of all be a distinction in the space between straight lines and other lines. This distinction, which is not determined by the topology, is provided by the *affine structure* of the space. In Euclidean space, the affine structure ensures that every pair of points are the endpoints of exactly one straight line and every finite straight line can be continued indefinitely in either direction. We can describe spaces that do not have this sort of affine structure: a pair of points might determine no straight line, or more than one, or there might be a limit to how far a straight line can be produced. So Euclid's first two postulates, which describe the uses to which a straightedge can be put, already restrict the affine structure of the space he is describing.

The affine structure of a space does not determine any facts about the *lengths* of lines or the *distances* between points. This requires yet another level of geometrical form, called the *metrical structure* of the space. The compass indicates metrical structure in a space: a circle is the locus of points all equidistant from a given center. Euclid's third postulate asserts that a complete, continuous, closed circle can be drawn with any given center and radius. Again, we could imagine spaces in which this does not hold.

The hierarchical form of these three levels of structure can be illustrated by three different sorts of transformation that can be carried out on figures in Euclidean space. A *topological transformation* carries continuous lines into continuous lines. An *affine*

[3] My translation.

transformation must further map straight lines onto straight lines. A "uniform stretching" of the space qualifies as an affine transformation even though it changes distances between points and deforms circles into ellipses. An *isometry* is a mapping from a space onto itself that preserves distances, so circles are carried into circles.[4] Figure 1 illustrates the three sorts of mapping.

Every isometry is an affine transformation and every affine transformation is a topological transformation, but not conversely.

Modern geometry introduces another level of structure, situated between the topology and the affine structure. This is the *differentiable structure*, which distinguishes smooth continuous curves from curves with corners or sharp bends. A mapping that preserves the differentiable structure is called a *diffeomorphism* and maps smooth curves into smooth curves. While a topological transformation can map a triangle onto a circle, a diffeomorphism cannot, since a circle is smooth and a triangle has corners. The topological transformation depicted in figure 1 is a diffeomorphism: notice that the three corners of the triangle are still identifiable.

In most discussions of Euclidean geometry, the lion's share of attention goes to the Fifth Postulate. This postulate concerns the existence and properties of parallel lines. The original discovery of non-Euclidean geometries arose from attempts to prove the Fifth Postulate from the other four. Eventually, it was shown that both the Fifth Postulate and its denial are consistent with the rest of Euclid's Postulates, so there can be spaces in which straightedges and compasses behave as Euclid requires, but in which geometric figures do not have the properties that Euclid demonstrates. For example, in some non-Euclidean spaces the interior angles of a triangle sum to more than two right angles, and in others they sum to less. The Fifth Postulate plays no essential role in the formulation of Newton's physics: Newtonian mechanics could obtain in a space that contains no parallel lines at all. The existence of an affine structure and a metrical structure, on the other hand,

[4] While an affine transformation merely must map straight lines to straight lines, an isometry must do more than map circles to circles: the size of the circles must also be unchanged. A scale transformation, which uniformly shrinks or expands all figures, is not an isometry even though it takes circles to circles.

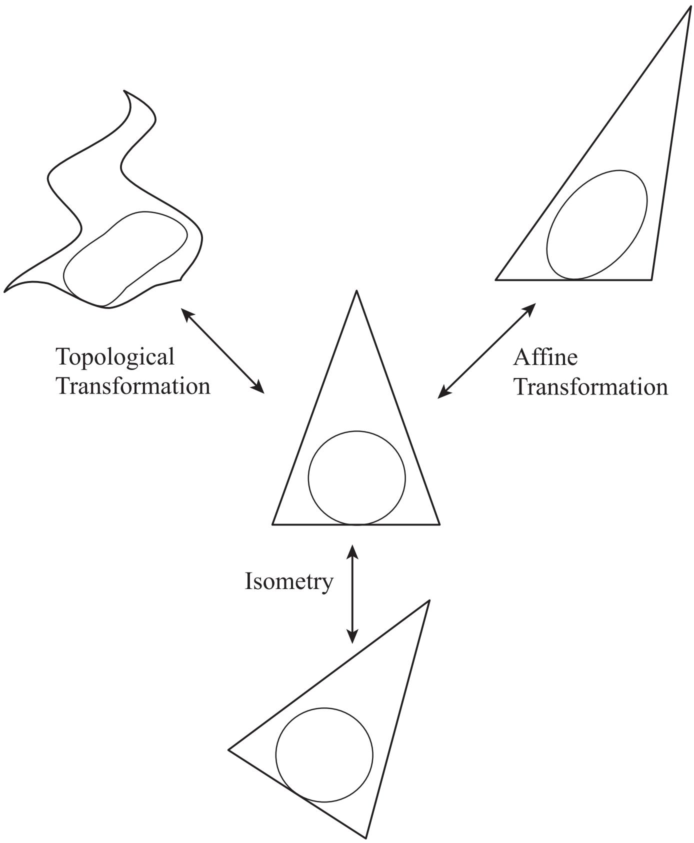

Fig. 1

is absolutely essential to make sense of Newton's Laws. But before we can make contact with those laws, we need to bring time into the picture.

Absolute Time and the Persistence of Absolute Space

Newton believed in the existence of a spatial arena with the geometrical structure of E^3. He believed that this infinite three-dimensional space exists at every moment of time. And he also believed something much more subtle and controversial, namely, that *identically the same points* of space persist through time.

We are trying to understand what must be postulated if the First Law is to make sense, and the First Law asserts that a body with no forces applied to it remains at rest if it is at rest, and continues moving uniformly in a straight line if it is moving. But what is it for a body to be *resting* or to *remain at rest*? If the individual points of space persist through time, then we have a precise account: a body is at absolute rest when it occupies the same points of absolute space over a period of time. The account of absolute uniform motion in a straight line is similar but more complicated. First: if the points of absolute space persist through time, then any moving body has a trajectory in absolute space, namely the set of points in absolute space that it occupies over a given period. And if absolute space has an affine structure, then such a spatial trajectory either forms a straight line in space or it does not. Thus, to make sense of "uniform motion in a straight line," the points of space not only must persist through time and have a topology (so it makes sense to characterize a trajectory of a body as a continuous line), but they also must have an affine structure so the spatial trajectory can be characterized as straight or curved.

But these conditions alone do not define "uniform motion in a straight line," since "uniform" has not been explained. A drag racer, unlike an Indy racer, runs on a straight track, so its motion is "in a straight line." Still, the motion is not *uniform*: the drag racer accelerates, constantly moving faster and faster. This sort of motion is called *linear acceleration*. (The Stanford Linear Accelerator Center has a straight tube about two miles long down which particles are accelerated, unlike the Large Hadron Collider, which accelerates particles around a closed loop.) In order for the motion to be *uniform*, it must cover the same distance in the same time. So Newton's First Law presupposes that there is a fact about *how far* a body moves and a fact about *how much time* it takes for it to complete the motion. The first fact requires a metric on the space, so that the spatial trajectory of a body can be ascribed a length. And the second fact requires something altogether new: a metrical structure on time. Newton's First Law of Motion presupposes not only absolute space but also absolute time.

The geometrical structure of time, according to Newton (and to common sense), is simpler than that of space. Newtonian time is one-dimensional: there is a single, ordered sequence of instants that forms the totality of history. That collection of instants has a topology, which is determined by their time order. It makes no clear sense to ask whether this "time line" is straight or curved, so the notion of an affine structure does not arise. But there is a temporal metric: between any two instants a certain quantity of absolute time passes, and these quantities can be compared with each other in terms of size. If a certain amount of time passes between instant 1 and instant 2, and a certain amount of time passes between instant 2 and instant 3, there is a fact about whether these intervals are the same size or different, and a fact about what the exact ratio between the intervals is.

With all of this structure in place, we can define "uniform motion": a uniform motion is a motion that covers the same amount of space in the same time. A uniform motion need not be straight: uniform circular motion, for example, can keep a constant speed even as it continuously changes direction. So Newton needed all of these characteristics in order for the First Law to make a precise statement: when no force is put on an object, if it is at absolute rest it will remain at absolute rest, and if it is moving it will continue moving in a straight spatial trajectory, covering equal distances in equal times.

There is one last feature that Newton ascribes to absolute time: it, unlike space, has a *direction*. Newton does not make any explicit remarks about this, and it is not immediately relevant to understanding his laws of motion. But it is a perfectly natural thing to say that time passes from the past to the future, and it is worthwhile remarking here because we will return later to questions that surround the direction of time.

Indeed, Newton does not explicitly discuss the geometrical structure of space or time at all. He always *uses* E^3 as his account of space, and he always presumes in his proofs that there is a definite metric for the passage of time. It would not have occurred to him that there could be any alternative. What we have seen in the foregoing analysis is which features of his account are required in order to state the First Law: space must have a topology, an affine structure, and a metric; time must be one-dimensional with a topology and a metric; and, most importantly, the individual parts of space must persist through time. Given all this, there is a fact about whether a body remains in the same region of space through time, a fact about the spatial trajectory of a moving object, and a fact about how quickly a moving body covers different parts of that trajectory. Without this much structure, it is unclear how to make any sense of Newton's First Law. But if we were to deny that space is E^3, ascribing it instead some other affine structure and metric, the law would still make perfect sense.

The Metaphysics of Absolute Space and Time

While Newton does not make explicit note of the geometrical structure he ascribes to absolute space and time, he does provide a very clear discussion both of their metaphysical status and of the reasons he thinks we must accept their existence. The basic issue is obvious. Take, for example, an object at absolute rest at some time that is not subject to any external forces. According to the First Law, the object will remain at absolute rest—that is, it will remain located at the same place in absolute space. But Newton is perfectly aware that these persisting parts of absolute space cannot be perceived by the senses. No observation can reveal whether a body remains in the same region of absolute space or constantly moves from one part to another. It would seem, then, that even if Newton's First Law is true, and even if we could ascertain that there are no forces on an object, no observation could verify the law. And, more seriously, if we cannot perceive absolute space and *a fortiori* cannot perceive absolute motion, it is not obvious how a theory that treats of such absolute motion could make any predictions about observable fact at all.

What we can observe, Newton asserts, are the *relative* positions of bodies with respect to each other. Similarly, we cannot directly observe the passage of absolute time, but we can observe changes in the *relative* positions of bodies. In the Scholium that follows his definitions of novel terms, Newton carefully distinguishes observable quantities from the absolute entities that he postulates:

> Hitherto I have laid down the definitions of such words as are less known, and explained the sense in which I would have them to be understood in the following discourse. I do not define time, space, place, and motion, as being well known to all. Only I must observe, that the common people conceive those quantities under no other notions but from the relation they bear to sensible objects. And thence arise certain prejudices, for the removing of which it will be convenient to distinguish them into absolute and relative, true and apparent, mathematical and common.
>
> I. Absolute, true, and mathematical time, of itself, and from its own nature, flows equably without relation to anything external, and by another name is called duration: relative, apparent, and common time, is some sensible and external (whether accurate or unequable) measure of duration by the means of motion, which is commonly used instead of true time; such as an hour, a day, a month, a year.
>
> II. Absolute space, in its own nature, without relation to anything external, remains always similar and immovable. Relative space is some movable dimension or measure of the absolute spaces; which our senses determine by its position to bodies; and which is commonly taken for immovable space; such is the dimension of a subterraneous, an aerial, or celestial space, determined by its position in respect of the earth. Absolute and relative spaces are the same in figure and magnitude; but they do not remain always numerically the same. For if the earth, for instance, moves, a space of our air, which relatively and in respect of the earth, remains always the same, will at one time be one part of the absolute space into which the air passes; at another time it will be another part of the same, and so, absolutely understood, it will be continually changed.

> III. Place is the part of space which a body takes up, and is according to the space either absolute or relative. . . .
>
> IV. Absolute motion is the translation of a body from one absolute place into another; and relative motion the translation from one relative place into another. Thus in a ship under sail, the relative place of a body is that part of the ship which the body possesses; or that part of the cavity which the body fills, and which therefore moves together with the ship: and relative rest is the continuance of the body in the same part of the ship, or of its cavity. But real, absolute rest is the continuance of the body in the same part of that immovable space, in which the ship itself, its cavity, and all that it contains, is moved. Wherefore, if the earth is really at rest, the body, which relatively rests in the ship, will really and absolutely move with the same velocity that the ship has on the earth. But if the earth also moves, the true and absolute motion of the body will arise, partly from the true motion of the earth, in immovable space, partly from the relative motion of the ship on the earth. . . .[5]

Newton distinguishes the "absolute, true, and mathematical" notions of space, time, place, and motion from their "relative, apparent, and common" counterparts. The crux of the problem is that while Newton's Laws of Motion are framed in terms of the absolute notions, these do not fall under our immediate observation. When we try to observe the motion of an object, all we can directly see is its relative motion: the change in its position with respect to other visible objects, with the rate of change being measured by the visible motion of clocks or other instruments for telling time. But if the absolute motion of an object is imperceptible because absolute space and time are imperceptible, how can the postulation of such entities have any relevance to empirical science?

Newton devotes the rest of the Scholium to answering this question:

[5] Newton (1934), vol. 1, pp. 6–7.

> But because the parts of space cannot be seen, or distinguished from one another by our senses, therefore in their stead we use sensible measures of them. For from the positions and distances of things from any body considered as immovable, we define all places; and then with respect to such places, we estimate all motions, considering bodies as transferred from some of those places into others. And so, instead of absolute places and motions, we use relative ones; and that without any inconvenience in common affairs; but in philosophical disquisitions, we ought to abstract from our senses, and consider things themselves, distinct from what are only sensible measures of them. For it may be that there is no body really at rest, to which the places and motions of others may be referred.
>
> But we may distinguish rest and motion, absolute and relative, one from the other, by their properties, causes, and effects. . . .[6]

Newton produces powerful *empirical* evidence for the existence of absolute motion (and hence absolute space and time) using considerations of the causes of motion. For this argument, we will need to consider his Second Law.

But before turning to the Second Law, we should pause to reflect how deeply intuitive Newton's account of absolute space and time is, even though absolute space and time are not directly observable. It sounds as if Newton is postulating some weird, ghostly, unfamiliar entities, but most people conceive of the physical world in terms of absolute space and time. For example, craftsmen and scientists continually try to improve the design of timepieces, to produce clocks that are ever more accurate and precise. But what is it for a clock to be "accurate"? What we want is for the successive ticks of the clock to occur *at equal intervals of time*, or for the second hand of a watch to sweep out its circle *at a constant rate*. But "equal" or "constant" with respect to what? With respect to the passage of time itself, that is, with respect to absolute time. Our natural, intuitive view is that a certain amount

[6] Ibid., p. 8.

of absolute time elapses between the successive ticks of a clock, and the better and more accurate the clock is, the more similar these intervals are to one another. Swiss watchmakers, and designers of atomic clocks, are trying to get their devices to accurately measure something, and that something is not any sort of relative, observable time. Physics treats every observable physical motion—the rotation of the earth, the motion of the earth around the sun, and so on—as subject to disturbances and hence not automatically uniform. But the nonuniformity is not defined with respect to any observable motion. Clock design reflects this commitment: disturbing factors are eliminated or compensated for. This practice implicitly assumes some measure of time itself that provides the standard of uniformity.

Similarly, our everyday understanding of the world conceives of it in terms of absolute space. No one is puzzled upon hearing, for example, that the orbit of the earth is an ellipse with the sun at one focus. Any picture of the solar system in a science book will draw the orbits of the planets. But what, exactly, is this supposed to be a picture of? At any given moment, the earth is in some one place. The "orbit" is somehow a collection of all the places the earth occupies over the course of a year. But that implies that the places the earth occupies at different times are all parts of one common, three-dimensional space: absolute space. We can, with effort and careful thought, come to comprehend how the world could exist without any absolute space or absolute motion. But when we do so we not only reject Newton's theory, we reject common sense as well.

Newton did not appeal to common sense to justify his belief in absolute space and time: he appealed to experiment. Newton tried to prove the existence of absolute motion in the laboratory rather than by conceptual analysis. This is our next topic.

Index