Newton, and the World Perceived

First-Order Thought Experiments and the Rationalization of the Sensible

Roger N. Shepard
Stanford Univesity

Note: This is a draft of Lecture 1 of seven lectures with the general title: *Mind and World: From Newton, Einstein, & Darwin to Principles of Mind* to be given by Roger N. Shepard as the William James Lectures at Harvard Univerisity during October and November 1994, and then to be published in book form. The first few of the following pages pertain to the set of lectures as a whole and include titles of all seven lectures. Drafts of some of the subsequent lectures may follow, separately. (Some of the material to be covered in Lectures 4-6 has been summarized in Santa Fe Institue Reprint 93-11-073, "Perceptual-Cognitive Universals as Reflections of the World.") Comments and corrections on any of these will be greatly appreciated!

Until June 29, 1994

Roger N. Shepard
Santa Fe Institute
1660 Old Pecos Trail, Suite A
Santa Fe, NM 87505

e-mail: roger@santafe.edu
phone: 505-984-8800
fax: 505-982-0565

After July 1, 1994*

Roger N. Shepard
Department of Psychology, Bldg. 420
Stanford University
Stanford, CA 94305-2130

e-mail: roger@psych.stanford.edu
phone: 415-725-2445
fax: 415-725-5699

(*However I will be away for meetings in Europe from July 2 until July 23)

Acknowledgments: This draft, still subject to revision, has already benefited from discussions with many colleagues, including, at the Santa Fe Institute: John Casti, Rajarshi Das, April Gifford, Atlee Jackson, James Hartle, Piet Hut, Stuart Kaufman, Seth Lloyd, and Bruce Sawhill; and, elsewhere, Johnathan Cobb, Leda Cosmides, Tambi Faigao, Geoffrey Miller, Wilfred Stone, and John Tooby.

MIND AND WORLD

From Newton, Einstein, & Darwin to Principles of Mind

(The William James Lectures, Harvard University, Autumn 1994)

Roger N. Shepard

Mind and world ... have evolved together,
and in consequence are something of a mutual fit.

-- William James (1893)

... at the very origin ... Each atom ... must have had
an aboriginal atom of consciousness linked with it.

-- William James (1890)

O Nature, and O soul of man!
how far beyond all utterance are your linked analogies!
not the smallest atom stirs or lives in matter,
but has its cunning duplicate in mind.

-- Herman Melville (1850)

Titles of the seven William James Lectures

MIND AND WORLD

From Newton, Einstein, & Darwin to Principles of Mind

Roger N. Shepard

I. PRINCIPLES OF THE WORLD AS EMERGENTS IN MIND

Lecture 1. Newton, and the World Perceived
First-Order Thought Experiments
and the Rationalization of the Sensible

Lecture 2. Einstein, and the World Beyond
Second-Order Thought Experiments
and the Rationalization of the Transcendent

Lecture 3. Darwin, and the World Within
Evolutionary Epistemology, Consistency,
and the Rationalization of the Rationalizer

II. PRINCIPLES OF MIND AS EMERGENTS IN THE WORLD

Lecture 4. Principles of Perception
Internal Representation of Things Present
and of the Manners of their Presentation

Lecture 5. Principles of Generalization
Internal Representation of Things Hidden
and of the Probabilities of their Manifestation

Lecture 6. Principles of Transformation
Internal Representation of Things Possible
and of the Paths to their Realization

III. THE NATURE AND RELATION OF MIND AND WORLD

Lecture 7. Unresolved Philosophical Issues
Necessary versus Arbitrary Principles
Conscious Experience versus Physical Reality

PART I.

PRINCIPLES OF THE WORLD AS EMERGENTS IN MIND

... new knowledge merely comes from thinking ...
Not new sensations ... but new conceptions,
are the indispensable conditions of advance.

-- William James (1890)

By always thinking unto them.

– Isaac Newton
(when asked how he made his discoveries)

In a certain sense ... pure thought can grasp reality,
as the ancients dreamed.

– Albert Einstein (1933)

... the whole of my pleasure was derived
from what passed in my mind.

– Charles Darwin
(on his studies as a naturalist
during the voyage of the Beagle)

1. Newton, and the World Perceived

First-Order Thought Experiments and the Rationalization of the Sensible

Contents of Lecture 1 — Page

1. Newton, and the World Perceived

Illustrations for Lecture 1 — Page

1. Newton, and the World Perceived

First-Order Thought Experiments and the Rationalization of the Sensible

> The elementary laws of mechanics, physics, and chemistry... [have] to be sought under and in spite of the most rebellious appearances ...
>
> --William James (1890)

> Our business is with the causes of sensible effects.
>
> --Isaac Newton

> It required an extraordinary genius to unravel laws of nature from phenomena that are always before our eyes.
>
> --Joseph-Louis Lagrange (commenting on Newton)

The invitation to return to Harvard to give the William James Lectures is an honor I could not refuse. The thesis I am going to present in these lectures began to take form in my mind during my two previous two-year periods at Harvard. The first (from 1956 to 58) was as a lowly postdoctoral associate with George Miller in the basement of Memorial Hall (where I shared an office with Ulric Neisser, who was to define the field Cognitive Psychology with his 1967 book of that name). The second (from 1966 to 68) was in the doubly elevated position of professor of psychology and occupant, in the newly completed William James Hall, of the lofty ninth floor (which I shared with Professor S. Smith Stevens, the 20th century's foremost psychophysicist). Moreover, my thinking leading up to this thesis has been influenced by several previous William James Lecturers -- including the two I was privileged to hear and to talk with during my two periods here, J. Robert Oppenheimer and Edwin Land.

Still, it is daunting to follow William James Lecturers of the eminences of Oppenheimer and Land; of John Dewey, Bertrand Russell, and Wolfgang Köhler, who loomed large in my undergraduate education; of Robert Thorndike and B. F. Skinner, who figured prominently in my graduate studies; of Karl Popper and Donald Campbell, who contributed to my later thinking about evolutionary epistomology; and of Herbert Simon and Allen Newell, who opened my eyes to the possibility of gaining an understanding of things in the world (including minds) not through direct observation of those things but through computer simulation of them.

It is also daunting to attempt lectures named for America's pre-eminent philosopher-psychologist, William James, who, throughout the cognitively dark ages of American behaviorism, remained a towering if distant beacon for those of us who sought to understand the inner life of the mind. Nor does it help much that I can trace my own Ph.D. advisor ancestory directly back (through Carl Hovland, Clark Hull, Joseph Jastrow, and G. Stanley Hall) to William James himself; for, American psychologists who can claim similarly direct intellectual descent from James must now be legion.

To counteract any rising expectations induced by this recalling of the lofty statures of William James and of previous James Lecturers, let me share with you an incident from the end of my own first year as a Harvard Professor. Appearing at my ninth floor office one day, Professors Stevens and Herrnstein proposed to offer me the title of Director of the Psychological Laboratories. It seemed a great honor, but I thought it only prudent to ask exactly what the title entailed. By way of reply, they escorted me up one flight to the tenth floor, which, alone, remained unfinished, without any interior walls or furnishings excepting the restroom fixtures. It was not, however, that this unfinished floor was to be made into some grand office/laboratory suite for the new Director. Instead, they told me that because the floor was expected to remain unoccupied for some years, the water in the toilet traps might eventually evaporate, risking the escape of sewer gases into the building. My only duty as Director of the Psychological Laboratories, they assured me, would be to protect Willam James Hall from such an ill wind by periodically climbing the one flight of stairs and flushing the tenth floor toilets. That too was an honor I could not refuse.

It was during this period on the ninth floor with Stevens, that I first began to articulate to myself the central idea that has guided my thinking and research ever since. During the more than ten years since we had both worked in the basement of Memorial Hall, Stevens had been continuing to find that for every sensory magnitude tested, the relation between the amount or strength of the stimulus as subjectively experienced and as physically measured is describable by a function having the same mathematical form (a power law), but requiring a different constant (exponent) for each sensory continuum -- whether visual length of line or brightness of light, auditory loudness of tone or duration of noise, tactile thickness of block or roughness of sandpaper, kinesthetic force of grip or weight of object, or gustatory sweetness or saltiness of solution.

Here, evidently, was a very general psychological regularity. But should it be accepted as a universal principle of mind; or might it merely be the consequence of an historical accident in the evolution of human or mammalian sensory transducers? To answer such a question, it seemed to me that we needed to understand why the empirically obtained psychophysical law had that particular form; and why, for each sensory continuum, it had that particular exponent.

The empirically found psychophysical relation might be nonaccidental if the transformation from sensory inputs to internal representations had been evolutionarily shaped just so that internal representations provide for the most simple and general guidance of behavior with respect to their corresponding external objects. In his 1890 Principles of Psychology, William

James had observed, "What appeals to our attention far more than the absolute ... quantity of a given sensation is its ratio to whatever other sensations we may have at the same time." Perhaps Stevens's power law could be understood as a consequence of our having evolved in a world in which it is generally the ratios of sensory inputs, and not the absolute values of the inputs themselves, that have an invariant significance in the external world.

Similarly, the exponent for a given sensory dimension might reflect the kind of relation in the external world that corresponds to ratios on that dimension. Thus, to understand Stevens's finding that the exponent is close to unity for "extensive" attributes such as length or distance, we might note that as children we learn that the phrase "twice as long" applies to a stick that matches two equal shorter sticks laid end-to-end, or applies to a distance that requires twice as many steps to traverse; and to understand his finding that the exponent is close to one-half for "intensive" attributes such as apparent strengths of lights, sounds, or odors, which fall off (like gravity) with the inverse square of distance from the source, we might conjecture (as Richard Warren suggested for loudness) that the phrase "half as intense" is appropriate for an equivalent source twice as far away.

My thinking along this line was reinforced at just this time by attending Land's William James Lectures. Supporting his arguments with compelling visual demonstrations, Land presented his ratio-based "retinex" theory for how we perceive the color of a surface as invariant despite the wide variations in the spectral composition of the light that the surface scatters back to our eyes under different conditions of over-all illumination.

In the ensuing years, I have increasingly pursued the possibility that the mind is not just a patchwork of ad hoc heuristics and neuroanatomical hacks evolved to deal with this or that accidental circumstance in the world. After all, some biologically significant features of the world are not accidental but universal. Beyond the pervasive invariance of ratios, for example, space (on the biologically relevant scale) is three-dimensional and Euclidean, time has a unique forward direction, and objects belong to distinct natural kinds. In these lectures I explore the possibility that evolutionary accommodations to such universals of the world may have given rise to corresponding universals of mind.

Organization of these Lectures

Guided by George Miller's well known universal constraint on human cognition, the "magical number seven, plus or minus two," I have prepared just seven lectures. The first three, grouped under the heading "Principles of the World as Emergents in Mind," focus on how universal principles of the world have been discovered by thinking about the world, as epitomized, respectively, by Newton, Einstein, and Darwin; and, also, on what the possibility of such discovery through thought tells us about the mind. The next three lectures, grouped under the heading "Principles of Mind as Emergents in the World," focus (not very modestly) on one psychologist's 40-year search for similarly universal principles governing the three fundamental mental faculties of, respectively, perception, generalization, and transformation; and, also, on how these principles may have arisen as evolutionary accommodations to universal features of the world. The final lecture, under the separate heading "Physics, Necessity, and Consciousness," considers what appear to me to be the most fundamental (if largely philosophical) of the still unanswered questions about mind and world: Are the fundamental principles governing the world arbitrary or necessary? And how can the subjective feel of conscious experience be reconciled with the objective world described by physical science? Throughout, I will have many occasions to note how much of what I have to say was foreshadowed over 100 years ago in William James's 1890 Principles of Psychology.

Science According to James

Many still subscribe to the Baconian idea that scientific knowledge is achieved solely through the gradual accumulation of objective facts by means of careful observation, description, and measurement -- augmented when possible by judicious experimental manipulation. But it is not the mere accumulation of observational facts, no matter how assiduously pursued, that produced Newton's mechanics, Maxwell's electrodynamics, Einstein's relativity, Bohr's, Heisenberg's, and Schrödinger's quantum mechanics, or Darwin's theory of evolution. These required, as we have it from Einstein himself, "free creations of the human intellect" guided by an intuitive "feeling for the order lying behind the appearance" and a desire to reveal that order as "a logical system." In his Principles of Psychology, William James expressed the idea in a way that will serve my purposes well -- both in his already quoted remark that "new knowledge merely comes from thinking" not from "new sensations," and, more fully, when he wrote:

> The only cohesions which experience in the literal sense of the word produces in our mind are ... the proximate laws of nature, and habitudes of concrete things, that heat melts ice, that salt preserves meat, that fish die out of water, and the like. ... The 'scientific' truths ... arise in the mind in no such passive associative way ... Even those experiences which are used to prove a scientific truth are for the most part artificial experiences of the laboratory gained after the truth itself has been conjectured. Instead of experiences engendering the 'inner relations,' the inner relations are what engender the experiences here.

The idea has in fact a history that can be traced back, much further, through Locke: "... as to all general knowledge, we must search and find it only in our own minds;" through Descartes: "... we ought to investigate what we can ... intuit or deduce with certainty ... For knowledge can be attained in no other way;" through Galileo: "You cannot teach a man anything. You can only help him discover it within himself;" -- back at least 24 centuries to the Socratic dialogues.

Thought Experiments and the World

The passage I just quoted from James raises the noteworthy possibility that the greatest advances in our scientific understanding of the world may have arisen not from experiments performed on the external world but from operations performed in the mind. With the insignificant substitution of "experiments" for James's "experiences" and "scientific insights" for his "inner relations," the essential portion of this passage becomes: "experiments ... are for the most part [carried out] after the truth itself has been conjectured. Instead of experiments engendering the scientific insights, the scientific insights are what engender the experiments." Primary among the mental operations yielding such scientific insights, I suggest, are the thought experiments or Gedankenexperimente, so named and explicitly discussed by the late 19th century Austrian physicist Ernst Mach; later made famous by Mach's early 20th century successor Albert Einstein; but practiced long before by Newton, Galileo, and the ancient Greeks.

In my first two lectures, then, I take up the challenge of considering how thought experiments may have led to each of what (again in compliance with Miller's "Magical Number") I regard as The Seven Wonders of Physical Science:

1. The theory of space (Lecture 1, p. 14),
2. The birth of physics (Lecture 1, p. 17),
3. Newton's mechanics (Lecture 1, p. 25),
4. Thermodynamics (Lecture 1, p. 31),
5. Electrodynamics (Lecture 2, p. 8),
6. Einstein's relativity (Lecture 2, p. 13),
7. Quantum mechanics (Lecture 2, p. 31).

The Eighth Wonder of Science -- Darwin's theory of evolution -- being a theory of biological rather than physical science and one with special significance for the origin of mind, I leave for separate consideration in my third lecture. I attempt to treat each of these "wonders" of science in a relatively self-contained way. (A reader of these lectures who prefers to skip one or more of the seven major developments in physics can therefore do so without losing the thrust of my general argument.)

The principal subject of these lectures is not, however, the history of science but the nature of the mind and its relation to the world. Accordingly, the crucial issue for me is not whether the thought experiments I am about to recount were originally performed in exactly the ways I describe. (In most cases, we will never know.) For me, the crucial issue is whether the mind is such that these thought experiments may have been so performed and may have led to such scientific advances (or could still be so performed and could still lead to such advances in other minds not yet apprised of those advances).

Undoubtedly, actual experiments have historically played an important role in the advancement of science. Yet, there are reasons to suspect that most, if not all, actual experiments were carried out only after they had been tried, at least in part, mentally. Admittedly, many of these imagined experiments were then carried out physically precisely because imagination alone did not suffice to yield a conclusive outcome. But this does not in itself rule out the possibility that a more incisive thought experiment would have led to a more decisive outcome. More telling is the fact that many of the most influential thought experiments in modern physics (as well as in philosophy) have been essential precisely because the corresponding actual experiment were physically (or morally) precluded. An acutal experiment would have required traveling at the speed of light, following the motions of a single electron, or interchanging the brains in two person's bodies. In any case, an enquiry into which of the laws of physics could in principle can be discovered through thought alone might tell us something both about the nature of the world and about the nature of the mind.

Birth of Mathematics and the Theory of Space

According even to such an empiricist as the great ninteenth century scientist Hermann von Helmholtz, mathematical reasoning is "the conscious logical activity of the mind in its purest and most perfect form." Ultimately, I shall suggest, it may be through the search for mathematical simplicity, invariance, and symmetry that despite "rebellious appearances," we are able to progress toward "a logical system" as described by Einstein, with, in James's words, "later views correcting earlier ones, until at last the harmony of a consistent system is reached." Perhaps it is principally in this way that we are able to attain new knowledge "merely by thinking."

Mathematical deduction is said to have originated with Thales of Miletus in the first half of the 6th century BC. Later in that same century it was zealously promulgated by Pythagoras, whose proof that the square of the hypotenuse of a right triangle equals the sum of the squares of its two other sides and whose doctrine that all things are reducible to numbers foreshadowed Descartes's reduction, 22 centuries later, of geometry to algebra (a reduction that was later to facilitate both Newton's formulation of a universal mechanics and the conception of the higher-dimensional spaces required for relativity theory and quantum mechanics).

Euclid

But it was with the emergence of Euclid's Elements in the beginning of the 3rd century BC, that, as Einstein remarked, "for the first time, the world witnessed the miracle of a logical system which proceeded from step to step with such precision that every single one of its propositions was absolutely indubitable." Perhaps the only gap in Euclid's Elements was the elegant general treatment, soon thereafter furnished by Apollonius, of the conic sections -- the curves that were to prove indispensable in Kepler's and Newton's mathematical characterizations of planetary motions.

The history of the idea that space itself has a physical existence, independent of any physical matter it may contain, extends at least from Democritus (by the beginning of the 4th century BC) to present day physicists, who, following Einstein, attribute even to empty space definite physical properties such as curvature and dimensionality. Indeed, one can say, with Harvard mathematician George Mackey, "Geometry is, in a sense, the simplest and best understood branch of physics."

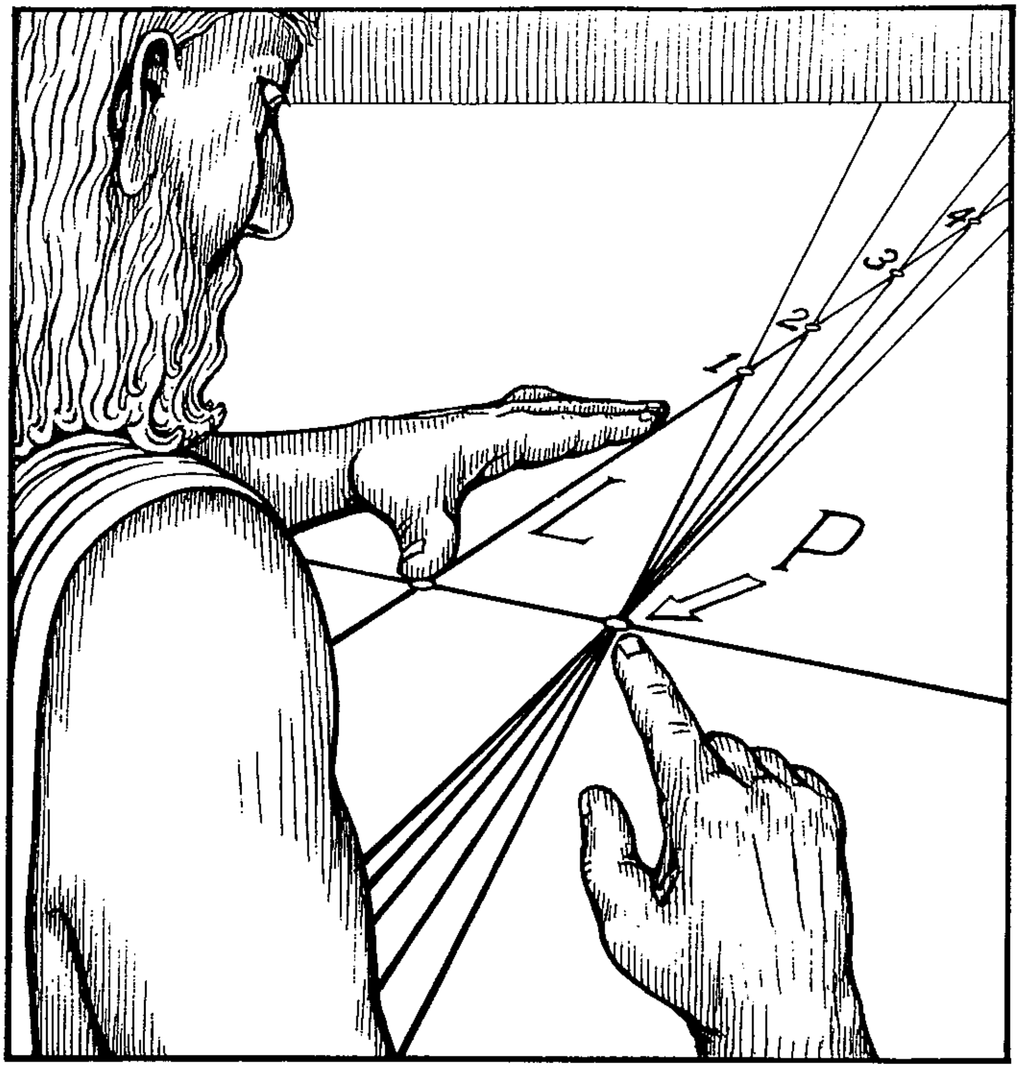
Accordingly, as a theory of space, Euclid's Elements can be regarded as the only theory prior to Newton's mechanics that was taken to apply to the physical world universally -- including both heaven and earth.

To ensure the truth of such a theory, it seemed sufficient to select only postulates and rules of inference whose validities were self evident -- for which disconfirming empirical evidence could not even be imagined. By this criterion, the Euclidean theory of space appeared to be largely secure. What conceivable observation or experimental outcome could induce us to reject, for example, Euclid's rule of inference that "Things that are equal to the same thing are equal to each other," or Euclid's postulate that "All right angles are equal"?

The system was seen to have only one possible Achilles' heel: Postulate 5, the notorious Parallel Axiom (Figure 1.1), "Through a given point P not on a line L, there is only one line in the plane of P and L that does not meet L." Only in this postulate was the reassuring ring of self-evidence somehow muted. Perhaps, as I have tried to illustrate, Euclid satisfied himself of the validity of Postulate 5 by trying in thought the physically impossible experiment of following out until they meet, successive pairs of straight lines whose interior angles with a crossing line more and more closely approximate two right angles.

Beyond Euclidean geometry

Euclid's thought experiment seemed almost enough, until Lobachevsky, Bolyai, and Gauss, performing (as I conjecture) more adventuresome thought experiments some 21 centuries later, independently discovered the existence of spaces, and hence possible worlds, in which lines, though making less than right angles with a crossing line need never meet each other. Following the advent of general relativity, the possibility has even arisen that (unless enough "dark matter" is found to approximate gravitational closure) our own universe could be such a world. So, although no physical observation or experiment had produced any evidence against Euclid's fifth postulate, the vague uneasiness that many geometers intuitively felt about that postulate seems to have been a valid, if weak and largely ignored, signal concerning what might be true about the world.



Euclid
imagines following lines to their intersection

(Figure 1.1)

Birth of Physics

Physics proper, traditionally taken to be the study of material objects and their forces and motions in space, may also have started with Democritus around the beginning of the 4th century BC. Democritus anticipated the modern atomic theory (and the principle of conservation of matter) by performing the thought experiment of dividing a material substance into smaller and smaller pieces until constituent particles are reached that can not be further divided (or divided without destroying the essential character of the substance). For Democritus, it was self evident that the final, irreducible constitutents of a material substance must be finite in size. For, within a dimensionless mathematical point, there would be no room for the structure that would surely be needed to support sensible properties such as hardness or softness or solidity or fluidity (to say nothing of color or taste).

Archimedes

In the third century BC, long before Galilo's and Newton's discoveries of universal laws of dynamics, governing the motion of material objects, Archimedes formulated universal laws of statics, governing the balance-of-forces conditions under which objects do not move. Archimedes is generally deemed the greatest mathematician and physicist prior to Newton. Indeed, the methods that Archimedes devised to compute the areas and volumes of various geometrical shapes (and to approximate the value of π to any desired accuracy) anticipated the calculus that was to be invented so much later by Newton. The day the bathing Archimedes discovered his principle of hydrostatics and ran through the streets in his "birthdy suit" shouting "Eureka! Eureka", ought, as in the assessment of Harvard's Alfred North Whitehead, "to be celebrated as the birthday of mathematical physics."

That discovery arose, I suggest, from a thought experiment (Figure 1.2). True, the discovery reportedly occurred when Archimedes, physically lowering himself into his bath, noted the actual rise in water level. But it was by imagining the measurement of the volume of water displaced by immersion of an irregular object that he solved, according to the familiar story, the problem of determining whether the goldsmith had fraudulently admixed silver with gold in making the king's crown. The density of the metal could be computed simply as the ratio of the already determined weight of the crown to the volume of water it displaced. In one intuitive flash (or should we say splash), Archimedes had arrived at two fundamental Archimedean principles: the principle that two physical objects cannot occupy the same place at the same time, and the principle of conservation of hydrostatic volume.

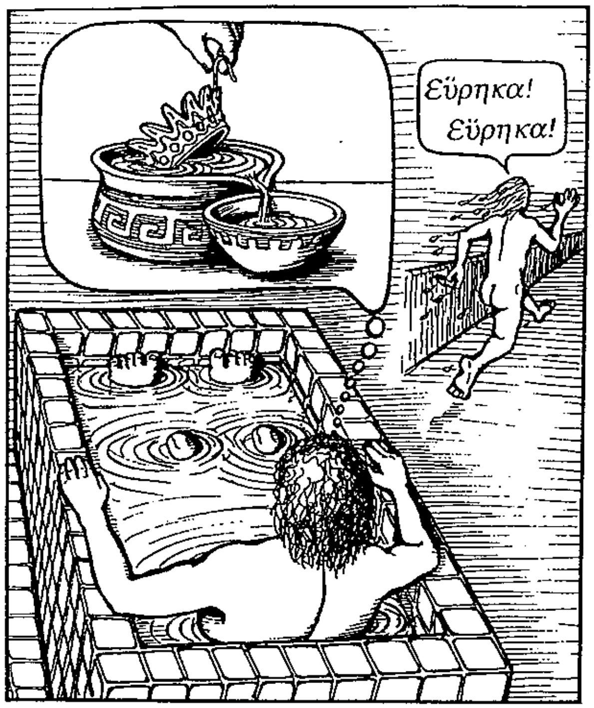
Aristotle

The attempt to formulate principles governing the motions of objects began a bit earlier with Aristotle, in the middle of the 4th century BC. Unlike Euclid's geometry or Archimede's statics, however, Aristotle's theory of motion was not put forward as universal. While celestial bodies exemplified, for Aristotle, a Platonic order of eternally uniform motion in perfectly circular orbits, the natural condition of earthly objects seemed to be that of remaining at rest on the earth. To lift and support an object required a continuing effort and, immediately upon being released, the object returned to the ground. True, a sudden push could send an object skidding over a horizontal surface. But, as Aristotle undoubtedly concluded from a simple thought experiment (Figure 1.3), the object soon returned to its natural state of rest. The external push imparted, according to Aristotle, an "impetus" that, being an accidental rather than essential property of the object, quickly decayed.

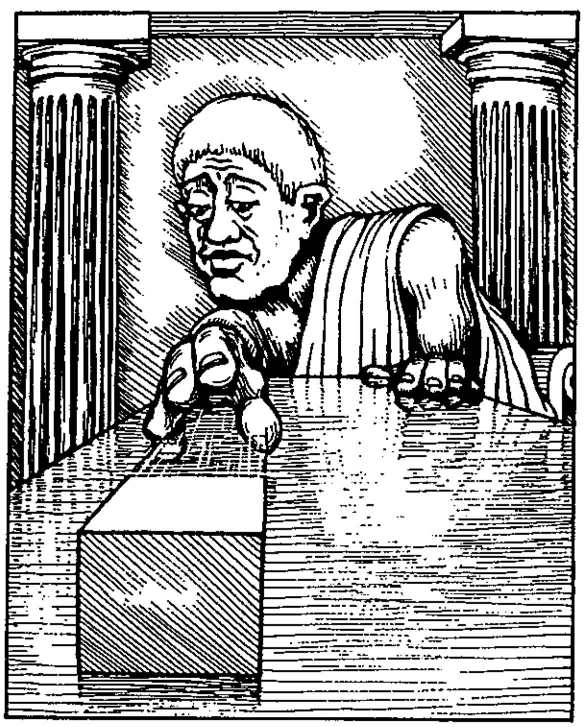
To subsume the widely different celestial and terrestrial behaviors under the same, universal laws of motion could not even be conceived without, as James said, disengaging such laws from under experience "by ignoring conditions which are always present." The ever-present conditions of the terrestrial environment include, in particular, surface friction, air resistance, and the earth's gravitation. These conspire to bring to rest objects that might otherwise -- if hurled out into the vacuum of extraterrestrial space -- go on forever.

Aristotle also held that of two objects released from the same height, the heavier one will drop faster. Because the conclusion is incorrect (except for the usually negligible effects of air resistance), we can guess that Aristotle did not take the trouble to perform the actual experiment. Quite likely he stopped with either or both of two misleading thought experiments: He may have imagined dropping a block of stone and a feather. Or he may have imagined simply hefting a large and a small block (Figure 1.4). In the latter case, feeling the palpably greater downward "striving" of the larger block in his "mind's arms" (so to speak), Aristotle would understandably have concluded that on release of the two blocks, the greater striving would surely actualize itself in a more precipitate descent.

Granted, Aristotle could have avoided his particular mistake by carrying out an actual experiment. What is more significant for my thesis is that he also could have avoided it simply by trying his thought experiment on imparting horizontal "impetus" with each of his previously imagined blocks, the heavy and the light one. He would then have realized that the heavier object would require a more forceful push than the lighter object to send each sliding at the same speed. Aristotle might



Archimedes
imagines displacing an equal volume
(Figure 1.2)



Aristotle
imagines imparting impetus to an object
(Figure 1.3)

Aristotle
imagines hefting unequal weights
(Figure 1.4)

then have asked himself: "Could it be that the greater downward striving felt in hefting the heavier block is exactly the greater force needed to get that heavier block moving earthward at the very same speed as the lighter object?" But, for that insight -- later to be formalized in Newton's second law of motion and in Einstein's principle of equivalence of gravitational and inertial mass -- we had to wait another 19 centuries, for the Italian physicist and astronomer, Galileo Galilei.

Copernicus

Before progress could be made toward the formulation of truly universal laws of motion, however, the celestial and earthly realms kept asunder by Aristotle, would have to be brought within one encompassing frame. The Polish astronomer Nicolaus Copernicus took the most momentous step in this direction when, early in the 16th century, he performed in thought an experiment that could never be carried out in reality (Figure 1.5). Mentally escaping from our island earth, Copernicus pictured how the planets might be seen to move in a simpler way if observed from a vantage point on the sun. Although taken in imagination only, this step was an incomparably greater step "for a man" as well as "for mankind" than the actual step that astronaut Neil Armstrong took, over four centuries later, onto the moon.

The shift to a heliocentric organization did not in itself yield more accurate predictions than the highly complex geocentric systems of Ptolemy and, later, of Tycho Brahe, with their epicycles upon epicycles. (Upon having the Ptolemaic system explained to him, King Alfonso X, "The Wise," had reportedly remarked, "If the Lord almighty had consulted me before embarking upon Creation, I should have recommended something simpler.") But the shift did permit "something simpler." Moreover, it was essential for Johannes Kepler's early 17th century discovery of his three simpler (and more accurate) laws of planetary motion and, hence, for Newton's formulation of the still simpler universal law of gravitation.

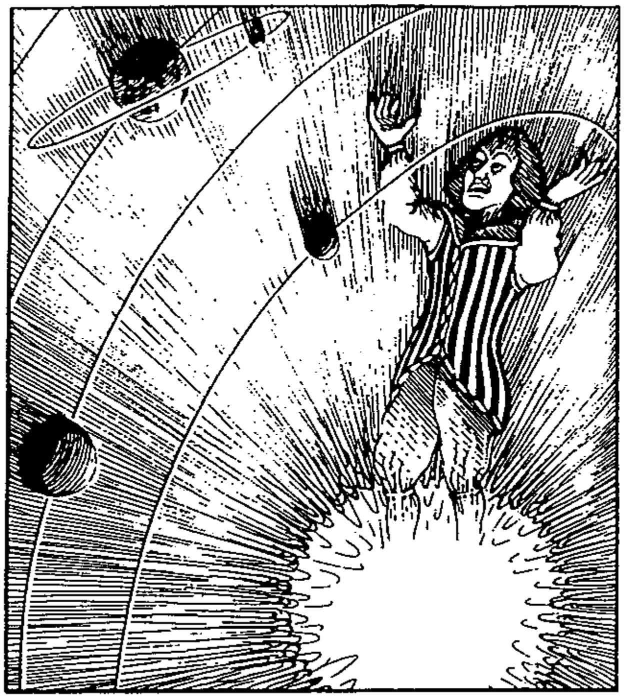
Galileo

The shift was not immediately accepted, however. Despite Galileo's discovery that at least one other body beyond the earth, namely, Jupiter, was incontrovertibly orbited by other bodies (Jupiter's moons), the chuch authorities disdained to look through Galileo's newly perfected telescope. As if the immensity of space and the eternity of time were but created for the enormity of man, they were willing, literally, to "take leave of their senses" in order to preserve Aristotle's certification of our earthly abode as the center of all heavenly circulation. Galileo might well have cried out, with St. Thomas, "The kingdom of God is spread before you and you see it not!" Toward the end of his life, as we all know, Galileo was forced by the Roman Inquisition to recant his belief in the heliocentric system. He was aware of the fate of his countryman Giordano Bruno, who, also having followed Copernicus in his mental journey to the sun but refusing to recant, was burned for his imaginings, not by solar flames but by earthly ones -- at the stake.

Two of Newton's three laws of motion -- the law of inertia and the force law (that force equals mass times acceleration, F = ma) -- can be traced back to Galileo (who died in the year of Newton's birth). I suggest that Galileo arrived at these laws through thought experiments. With regard to inertia, Galileo had only to extrapolate from his intuitive knowledge that a given force will send the same object farther and farther on surfaces offering less and less frictional resistance -- say, on rough ground, marble floor, and smooth ice. He could then surmise that in the absence of all resistance, the object would never come to rest. Thus, through thought alone, Aristotle's concept of an "impetus" that spontaneously decays can be judged inferior to Galileo's concept of an inertia that endures unchanged unless or until counteracted by some other (for example, frictional) force.

In his Two New Sciences, Galileo illustrated the equivalence of intertial frames, by noting that from what one could observe inside the closed cabin of a ship, one could not tell whether or how rapidly the vessel itself is moving. To the question of whether he had actually performed an experiment to verify that relative to a shipboard observer, a ball drops from the mast in the same way whether the ship is moving or stationary, Galileo's reply was ready: "I do not need it, as without any experience I can affirm that it is so, because it cannot be otherwise."

With regard to the force law, Galileo, supposing that the full strength of one person would impart a certain velocity to a given object such as a heavy cart, could immediately deduce from symmetry that exactly two such persons would be required to impart that same velocity to two identical carts -- even if the carts were attached side-by-side to consitute an object of just twice the mass of the original object. Imagining all this taking place on the deck of a ship that was itself already moving in the same direction, Galileo could conclude that whatever a cart's current velocity relative to a fixed frame external to the ship, a specified increment or decrement in that relative velocity would require a force that is proportional to the mass of the object and that is independent of its current relative velocity. From such a conclusion one can deduce Newton's second law of motion, F = ma.



Copernicus
imagines viewing the planets from the sun
(Figure 1.5)

Galileo
imagines dropping three identical bricks
(Figure 1.6)

The most definitive of Galileo's thought experiments, however, is the one he used to refute Aristotle's claim that objects fall at rates proportional to their weights. Galileo imagined releasing, from the leaning tower of Pisa, a large iron ball and a small one that were connected together by a short length of chain. For this experiment, Aristotle's theory leads to a striking inconsistency: The fall of the heavier ball should be held back by its connection to the more slowly descending lighter ball; at the same time the descent of the heavier ball should be accelerated because the addition of the smaller ball adds weight to the combined object!

Before I had read Galileo's thought experiment, I myself carried out a very similar thought experiment while puzzling (early one morning before arising) over Einstein's principle of equivalence of inertial and gravitational mass. I include a description of my thought experiment because it illustrates, particularly clearly, the use of symmetry. (By "symmetry," I do not mean merely bilateral symmetry but the more general, mathematical symmetry that holds whenever there is invariance under some transformation. The transformation need not be reflection about a vertical axis; it could be, as here, a permutation of two or more objects.) I describe my thought experiment as it might have been performed by Galileo, in recognition of the possibility that it might in fact have been performed by him and in recognition of his nearly 400-year priority! Certainly, Galileo could well have imagined himself at the top of the tower of Pisa with three identical bricks (Figure 1.6). By symmetry, he could only suppose that these three mutually interchangeable bricks, on being individually hefted, would manifest exactly the same downward force and, on being released together over the edge, would reach the ground at exactly the same instant.

Now comes the crucial step. The following thought could suddenly have struck Galileo (as it did me): It would take no more than a piece of string or dab of glue to make two of these three bricks into the equivalent of a single brick. The resulting larger brick, on hefting, would surely be felt to exert a double downward force, exactly as imagined by Aristotle. But hold on! Surely the two bricks, which (by symmetry) would have fallen side-by-side with the third brick, would not now -- merely by the addition of a virtually weightless bit of string or glue --- fall, as Aristotle asked us to believe, twice as rapidly! Possibly it is only because I thought of it myself, but this single thought experiment gave me my own most direct and compelling experience of the possibility that fundamental facts about the physical world can be discovered by thought alone.

Kepler

While Galileo was formulating laws of motion of terrestrial objects, the contemporary but shorter-lived German astronomer Johannes Kepler was working out his three laws of motion of the planets around the sun. These are:

1. Each planet moves in an elliptical orbit (with the sun at one focus of the ellipse).

2. A planet moves more swiftly around the end of the ellipse closest to the sun, in just such a way that an imaginary radial line segment extending from the sun to the planet sweeps out equal areas in space during equal intervals of time.

3. More distant planets move more slowly, in just such a way that the square of the time to complete each revolution about the sun is proportional to the cube of the planet's average distance from the sun.

Kepler arrived at these three laws by a trial-and-error process of trying to fit simple mathematical equations to the measurements of planetary positions that his employer, the Danish astronomer Tycho Brahe had been systematically collecting with unprecedented precision. Kepler's laws thus provided no more than a concise quantitative description of the astronmical data, not an explanation for why those data were well approximated by just such mathematical equations. For such an explanation, we had to wait for Newton's discovery of the universal law of gravitation. But, given that Kepler already had both data and the mathematical equations that fit the data, why did he himself fail to discover that universal law?

Kepler seemed to have had the essential idea required to account for the faster motion of planets closer to the sun that is at the basis of both his second and third laws, for he wrote: "There exists only one moving soul in the center of all the orbits; that is the sun which drives the planets the more vigorously the closer the planet is ..." Indeed, he came very near to getting the law of gravitational attraction itself when he later said, more specifically, "this cause of motion diminishes in proportion to distance just as the light of the sun diminishes in proportion to distance from the sun ..." If only he had thought this analogy through just a bit more carefully, he might have reached the correct conclusion that the sunlight diminishes not exactly "in proportion to distance" but in proportion to the square of distance. For, if the total amount of light radiating from the sun remains undiminished as it passes through succesively larger spherical shells imagined around the sun (conserving, as we would now say, the total energy of this radiation), the purely geometrical fact that the areas of these spherical surfaces increase as the square of their radial distances from the sun implies that the amount of light falling on a patch of fixed size must decrease with the inverse square of its distance. By following his own line of reasoning, Kepler might thus have been led to the conjecture that the same is true for the gravitational force the sun exerts on a planet of fixed size (that is, mass).

To his credit, Kepler was willing to give up Platonic circularity of planetary cycles and epicycles in order to account for astronomical observations in terms of a few simple mathematical equations. But he seems to have been distracted from a search for one universal law of gravitation underlying his three separate laws of planetary motion by notions that were still of Platonic and even theological origin. He felt that the spacing of the successive planets' orbits must somehow correspond to the five regular Platonic solids, and that the sun, the planets, and the fixed stars must be direct manifestations, respectively, of "God, the Son, and the Holy Ghost." Kepler's metaphysical preconceptions together with the perceptual salience of the sun in the terrestrial sky, conspired to support Kepler's assignment of a unique status to the sun. If he had instead adoped Bruno's surmise that our sun is but one of countless others (the fixed stars), Kepler might have been more ready to entertain the possibility that not only our sun but every material body, including the earth, and all the objects upon it, exert their own individual gravitational attractions.

In short, the very great advance that Kepler did achieve came through his search for mathematical laws to explain facts of the world. That he did not attain an even greater advance may be traceable to his mistaken ideas about which facts of the world (such as the regularities in motions of the planets) are determined by universal laws, and which facts (such as the particular number and spacing of the planets) are the results of historical accidents. Kepler's failure to discover the fundamental laws of celestial mechanics despite having very complete and precise data illustrates Einstein's point that the "axiomatic basis of theoretical physics cannot be abstracted from experience but must be freely invented." Thus it was that the discovery of the first truly universal laws of motion remained for the more circumspect Newton, who cautioned "Physics, beware of metaphysics."

Newtonian Mechanics

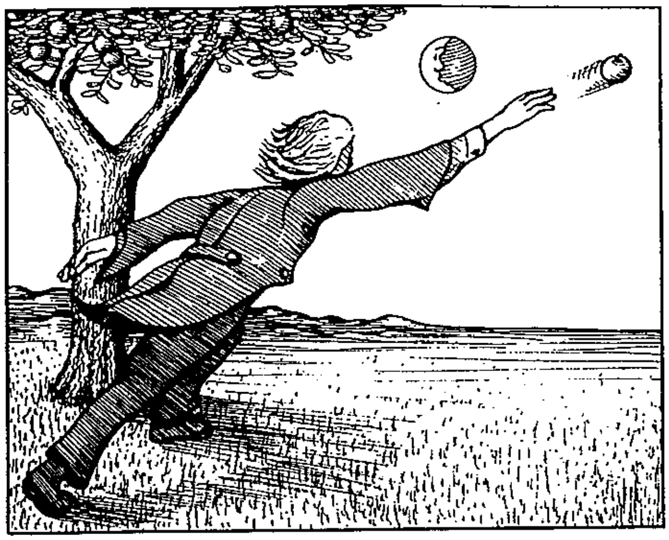
Newton

We all know the story that, some 20 years before finishing his 1687 Principia, Isaac Newton had already had an inkling of his universal law of gravitation. Sitting under an apple tree, gazing at the moon on a mild autumn evening at Woolsthorpe, so the story goes, Newton had been (metaphorically if not literally) struck by a falling apple (Figure 1.7). Whether or not this story, promoted on occasion by Newton himself, is historically accurate, there is reason to believe that at some time before writing the Principia, Newton performed the thought experiment of throwing a terrestrial object (such as an apple) with greater and greater force. He already knew that the trajectories of such hurled objects always curve downward to return to earth, but at greater distances for more forceful throws. He also knew that the earth is a sphere of finite size. Newton could well have imagined that with a throw of sufficient force, the downward curve would carry the apple not to earth but around it (Figure 1.8). In just such a way, might not the moon (originally hurled, as Newton supposed, from the figurative hand of God) be falling around the earth, and the earth falling around the sun? The essential insight, unlike its full mathematical formulation, did not have to wait for Newton's invention of the calculus or his decision as to the specific, inverse-square form of gravitational attraction.

As already noted, the possibility that that the sun attracts each of the planets with a force that decreases with distance had been surmised by Kepler. Moreover, Newton's rival, Robert Hooke, had even proposed that the sun's attraction falls off specifically with the inverse square of distance. It was Newton, however, who evidently first conceived of gravitation as a force operating universally between all bodies. The attraction that the sun exerted on each planet was exactly reciprocated by the attraction that each planet exerted back on the sun. Indeed, a mutual attraction similarly operated between every pair of objects in the solar system -- whether planets, comets, moons, or apples.

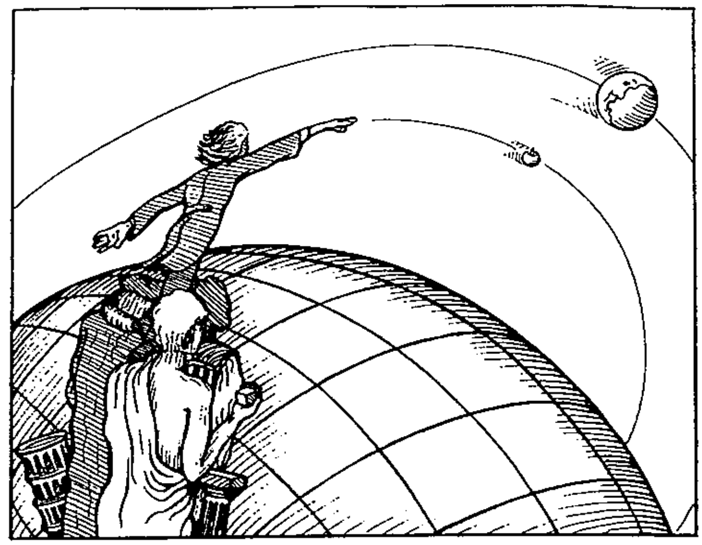
This reciprocity was directly entailed by Newton's third and perhaps most original law of motion -- that for every action there is an equal and opposite reaction. To arrive at this third law, Newton had only to imagine himself arched over with his feet braced on one apple cart while he pushed a second, identical apple cart with his hands. Simply from the symmetry of the situation, again, the result should be independent of which cart the push was exerted through his hands and which through his feet. (Still more symmetrically, he could have pushed with his right hand and foot on one cart and his left hand and foot on the other.) Barring the upsetting of an apple cart, the two carts would surely move equally, but in opposite directions.

By extension, Newton could also have considered that if he found himself stranded, oarless, in the middle of a pond with a boatload of apples, he could still make



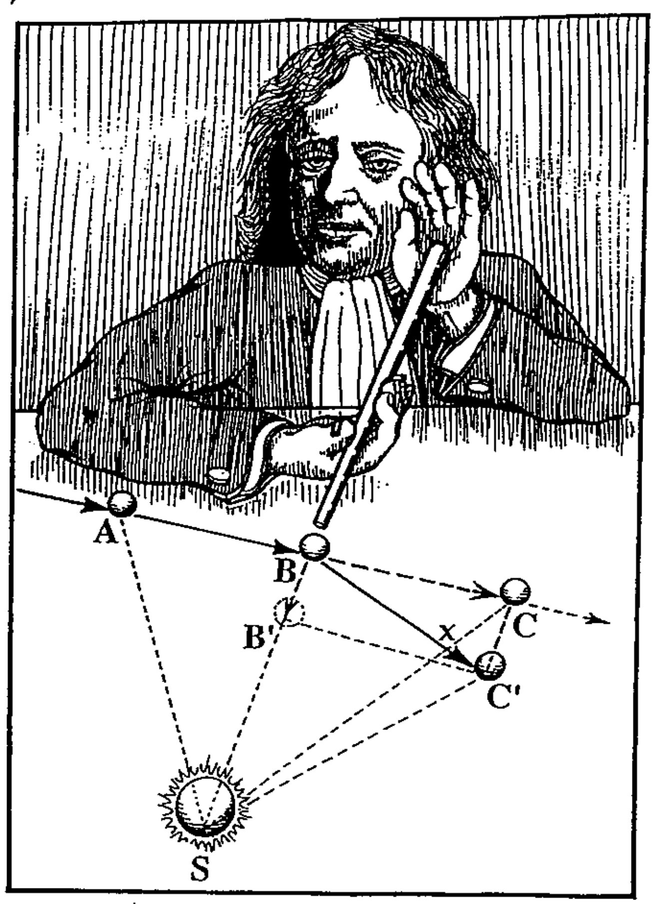
Newton
imagines hurling an apple with great force

(Figure 1.7)



Standing on the shoulders of giants,
he imagines the apple falling around the earth

(Figure 1.8)



Newton
imagines simulating a gravitational push

(Figure 1.9)

his way to shore by forcefully pitching the apples to the rear. In essentially the same way, the rocket carrying Neil Armstrong and Edwin Aldrin to the moon made its way across empty space by the rearward ejection of the even smaller molecules of gas at even greater velocities. Indeed, the principle of action and reaction should have been discoverable in antiquity, when devices were already built in which (as in Hero's aeolipile of the 1st century AD) steam escaping from nozzles bent in one direction around the perimeter of a spherical chamber supported by two horizontal bearings caused the chamber to spin in the opposite direction.

Derivation of Kepler's laws

Newton was able to show that each of Kepler's three laws of planetary motion were derivable from the simpler, more intuitive, and universal laws that he proposed to govern the motions of all bodies. I sketch the argument only for Kepler's second law, the conservation of swept area (which is mathematically equivalent to what we would now call the conservation of angular momentum). This purely geometrical argument, which did not require Newton's newly invented calculus, was based on a simple thought experiment (Figure 1.9).

According to Newton's first law (the law of inertia), if there were no gravitational force, a planet would continue to move at constant velocity along a straight line, reaching equally spaced positions, A, B, C, at equally spaced times. Imaginary lines drawn from the these successive positions A, B, C to the center of the sun, S, would form two triangles having the same area, because their bases are the equal intervals (A-B and B-C) along the line of motion and their altitudes are just the distance of that line from the sun, S. Newton then imagined approximating the sun's continuous gravitational pull by giving the planet a discrete (if not discreet) little push toward the sun at the end of each of the equal temporal intervals. The push at position B, would have caused a stationary planet to move directly toward the sun, from B to B', during the next time interval. The actual motion of the planet under consideration would, however, be a combination of this gravitationally induced motion B -> B' and the inertial continuation B -> C of its preceding motion A -> B. By the parallelogram rule, the composition of these two component motions would be the resultant motion B -> C'. Simply from the geometry, Newton could then show that the triangle BC'S, for this gravitationally deflected motion, preserves the area of BCS and, hence, of ABS, in accordance with Kepler's law.

(Briefly: Any change in area from the old triangle BCS to the new triangle BC'S is obtained by subtracting the area of triangle BCC' and adding the area of triangle

SCC'. The irrelevant overlap, CC'X, of these two triangles, being first subtracted and then added, cancels. But the two triangles BCC' and SCC' have the same base, C-C', and, because C-C' is parallel to B'B', the same altitude. Their areas are therefore equal, yielding the claimed equality in area of triangles BCS and BC'S.)

Kepler's first and third laws can be approximated by iteration of this same geometrical construction, if the gravitational pushes (appropriately reduced at greater distances from the sun) are administered at sufficiently small time intervals. The resulting trajectory then follows one of the conic sections of Apollonius (ellipse, parabola, or hyperbola); but the only conic section for which a planet would remain bound to the sun in an enduring planetary orbit is the ellipse posited by Kepler. By a further thought experiment, like Galileo's of dropping two iron balls, one can immediately see that the velocity of a planet around its elliptical orbit should not depend on its mass; for, the velocity of that mass surely would not change simply because two halves of the object were detached from each other, making two planets each of half the original mass. Given that the planet's mass is irrelevant, the only variable left to determine its velocity around an elliptical orbit of fixed shape but different size, is its distance from the sun. (Newton's laws require that this velocity be proportional to the square root of this distance which, in turn, entails Kepler's third law, that the square of the orbital period be proportional to distance cubed.)

Rational basis of Newtonian mechanics

Initially, the specific inverse-square form of the law of gravitation, from which Newton was able to derive Kepler's laws, may have come to Newton as a lucky guess (whether by Hooke or by crook). Even without reference to precise astronomical data, however, several purely theoretical considerations converge in support of an inverse square law. Mathematical and numerical analyses show that only when distance is raised to the second power (in the denominator) do we realize the following three conditions: First, if a spherical shell of fixed mass per unit area were to surround a massive body, such as the sun, the total gravitational force on that shell would be independent of its radius, r. (As I have already observed, this is a kind of conservation of gravity with distance, analogous to the conservation of the total energy of sunlight falling on such a shell, regardless of its radius.) Second, the gravitational field around a spherically symmetric body would be unchanged if the mass of that body were wholly concentrated in a single point at its center (a result that Newton regarded as crucial, and for which he had to develop the integral calculus, delaying publication of the Principia for close to twenty years). Third, the resulting orbit of a planetary orbit is stable in a way that it is not for powers different from 2. We can therefore say not only that Newton's entire system follows from just his three laws of motion and his one law of gravitation. We can also say that these first three laws are discoverable solely through thinking, and that the fourth has unique, optimally stable properties that are discoverable through thinking supplemented by mathematical analysis -- which is another kind of thinking.

Physical implications of Newtonian Mechanics

Of course, Newton's laws explained much more than Kepler's laws of planetary motion. They explained the laws that Galileo had discovered for projectiles and falling bodies on earth, they explained the terrestrial tides. They explained the highly eccentric (potentially nonelliptical) orbits of comets. And they explained the subtle interactions among planets, moons, and comets. Indeed, it was by virtue of the universality and reciprocity of Newton's gravitation that the small observed perturbations in planetary motions, which were at first thought to be violations of Newton's laws, were one by one shown to be entailed by those laws (thanks to the prodigeous analytical efforts of the likes of Joseph-Louis Lagrange and Pierre Simon de Laplace). The crowning triumph, requiring nearly a year of hand calculations (independently carried out by Adams in England and Leverrier in France in the mid 19th century), was the deduction, from unexplained perturbations of Uranus, of exactly where astronomers should aim their telescopes to spot an eighth planet, now called Neptune -- but previously unknown and still invisible to the unaided eye. (Again, what may be inferrable through thinking are the laws of nature -- not accidental facts, such as that there are more than seven planets in our particular solar system.)

Psychological implications of Newtonian Mechanics

In the Preface to the Principia, Newton expressed a desire to "derive the rest of the phaenomena of Nature by the same kind of reasoning from mechanical principles." And, in a letter to Oldenberg, he explicitly included, among such "phaenomena," "that puzzling problem, 'By what means the muscles are contracted and dilated to cause animal motion'." Indeed, it was on the basis of Newton's Principia, which Laplace characterized as standing "pre-emenently at the head of all human productions," that Laplace made his oft-quoted claim that for "an intelligence that could ... embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom, ... nothing would be uncertain and the future, as the past, would be present to its eyes." Laplace further noted that "The human mind offers, in the perfection which it has been able to give to astronomy, a feeble idea of this intelligence." Perhaps the most significant legacy of Newtonian mechanics for the science of psychology, then, has been the idea (foreshadowed by Descartes and Thomas Hobbes) that if the whole universe is nothing but a gigantic, deterministic clockwork, then we, too, must be but parts of that clockwork. Such an idea has encouraged the quest for mathematical principles of mind, including those I will be presenting later in my lectures, here. As we shall now see, however, a claim that Newton's laws govern the motions of all material bodies cannot be accepted as a claim that those laws, by themselves, explain all orderly phenomena that we experience, whether mental or physical.

Thermodynamics and Statistical Mechanics

Unexplained phenomena

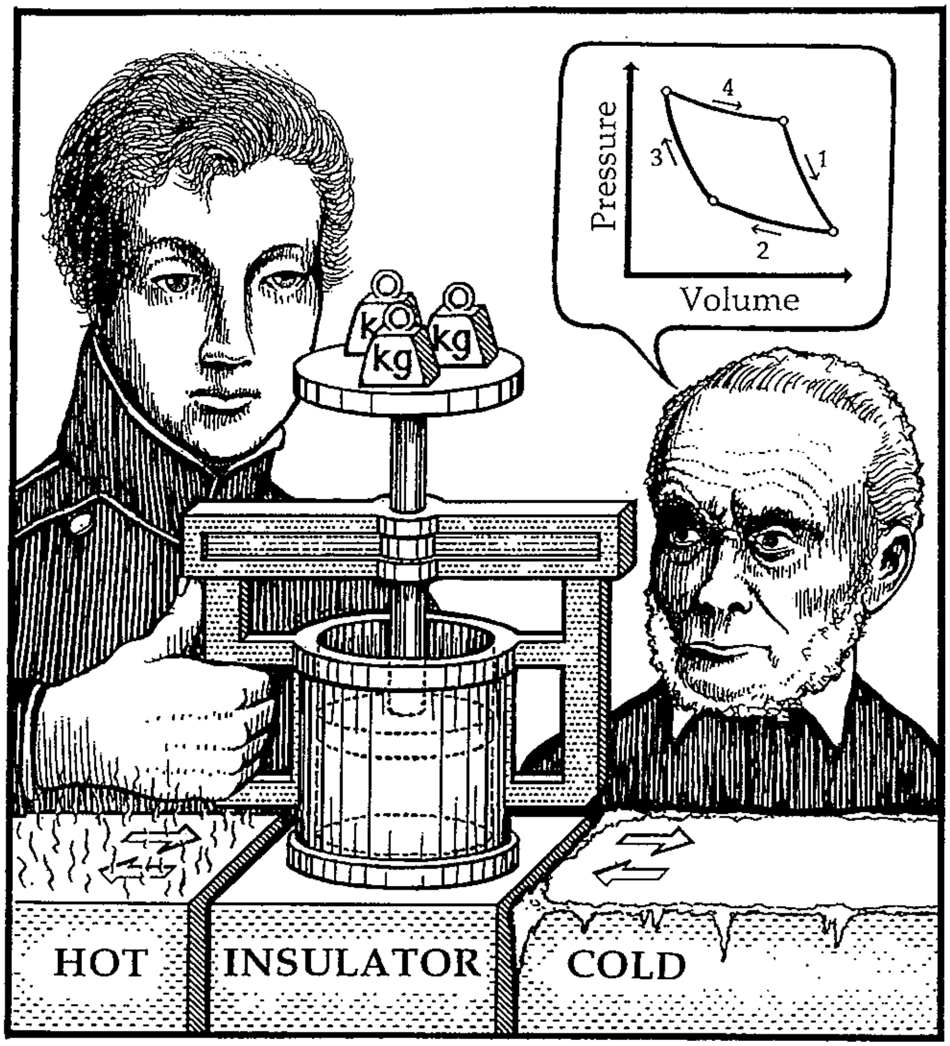
For 150 years, natural philosophers were so enamored of the universality and predictive power of Newtonian mechanics, that they blithely took it as the "theory of everything" that eludes us still. They had come under the thrall of the "kingdom of God" that was spread before them, but they saw it as yet through a glass darkly. Oblivious to "rebellious appearances," they continued in their contemplation of a universe in which, for example, all movements arose through force-driven acceleration, a, in accordance with Newton's second law ($F = ma$). They failed to notice that because acceleration has the units of the inverse square of time, t, and because t^2 is equivalent to $(-t)^2$, a Newtonian universe (like a system of orbiting planets) must look the same run forward or backward. But, even as they sipped their coffee by the fire, "the moving finger wrote and, having writ, moved on." The wood in the fireplace turned to ashes, and stayed that way. The coffee in each cup grew cold, and stayed that way. The milk and the coffee mingled, from white and black to a uniform tan, and stayed that way. And, eventually, the philosophers themselves, one by one, grew cold and still, and stayed that way. Run backward, such a world would have looked very different. But they saw it not.

Carnot

It was well into the 19th century that scientists finally turned their attention to irreversible phenomena and the new laws such phenomena evidently implied. Joseph Fourier's 1822 law governing the conduction (or "motion") of heat within a material body depended linearly on t and, so, differed from Newton's laws of the motion of material bodies themselves in being irreducibly asymmetric with respect to time. The steam engine was now beginning to power the industrial revolution through the conversion of heat into mechanical motion. The reverse conversion of mechanical motion into heat had long been known; on braking, the forward progress of a carriage is converted into the heating of the brake pads. But no one had yet elucidated the relation between mechanical motion and heat, or the degree of mutual convertability from the one to the other. Indeed, even when the first such analysis, by an obscure 23-year old French artillery engineer and physicist, Sadi Carnot, was finally published in 1824 it was virtually ignored for another 20 years (until well after Carnot's untimely death at age 36).

Recognizing that general principles are only achieved through abstraction, Carnot set his goal: "if the art of producing motive power from heat were to be elevated to the stature of a science, the whole phenomenon must be studied from the most general point of view, without reference to any particular engine, machine, or operating fluid." Accordingly, Carnot performed his thought experiments on a highly idealized "heat engine" that abstracted the theoretically essential properties of heat, temperature, pressure, and volume of the "elastic fluid" away from the irrelevant complexities of the fire, smoke, steam, valves, cranks, and wheels of an actual steam engine (Figure 1.10). Central to Carnot's heat engine was a cylinder and tight-fitting but frictionless piston enclosing a fixed charge of the "elastic fluid" or gas. He imagined the walls of the cylinder and the piston to be perfect heat insulators, but the bottom of the cylinder to be a perfect heat conductor. The cylinder could then be placed on the flat upper surface of any of three bodies: a hot body (acting as a large-capacity heat source at temperature T_{hot}), a cold body (serving as a large-capacity heat sink at temperature T_{cold}), or a perfect insulator.

If that cylinder were moved from the cold to the hot heat reservoir, the gas in the cylinder would increase in temperature from T_{cold} to T_{hot}, causing the gas to expand, performing mechanical work. Moreover, this work could be quantified simply as the product of the height to which the piston was raised by the expanding gas and the weight that the piston carried upward. Carnot realized, however, that the relation between heat and work would be meaningful only if the conversion process



Carnot and Clausius
imagine the operation of a heat engine

(Figure 1.10)

could be continued indefinitely. He therefore sought a cycle of operations that, like that in an actual steam engine, returned the system to the same internal state at the end of each cycle. But Carnot also knew that whenever a warmer body comes into thermal contact with a cooler body, heat flows irreversibly from the warmer to the cooler body (in accordance with Fourier's law), reducing the temperature difference necessary for the production of further mechanical work. This led Carnot to the construction of an ingenious cycle, now known as the Carnot cycle, which consists of four operations in which thermal contact occurs only between bodies that are already at exactly the same temperature:

1. The cylinder is moved from the hot reservoir to the insulator and the piston is unlocked, allowing the gas (initially at temperature T_{hot}) to expand without loss or gain of heat (that is, adiabatically) until its temperature has fallen to T_{cold}.
2. The cylinder is placed on the cold reservoir and the piston is forced back down, sufficiently slowly that the compressing gas remains in thermal equilibrium with the cold reservoir at its constant temperature T_{cold} (that is, isothermally).
3. The cylinder is returned to the insulator and the gas is further compressed, without exchange of heat (that is, adiabatically), until the temperature of the gas rises to its initial value, T_{hot}.
4. The cylinder is returned to the hot reservoir (still at that temperature, T_{hot}) and the gas is allowed to expand, without change of temperature (that is, isothermally), back to its state at the beginning of the cycle.

Because any of the four expansions or compressions could be reversed (with corresponding reversal between input and output of mechanical work), each of the four operations, and hence the entire cycle, is reversible. In the direction of operation just described, the output of mechanical work during expansion exceeds the required input of mechanical work during compression. Hence the net mechanical output could be used externally at the expense of the flow of heat from the hotter to the colder reservoir (much as a water turbine performs mechanical work at the expense of the flow of water from a higher to a lower reservoir). In the opposite direction of operation, the input of mechanical work for compression exceeds the output during expansion. Hence the heat engine acts as a refrigerator, "pumping" heat from the colder back to the hotter reservoir at the expense of an external supply of mechanical energy (much as the the turbine could pump the water from the lower back to the higher reservoir through the input of mechanical work to turn the turbine in the opposite direction).

At any given moment during the cycle, the state of the gas in the cylinder can be abstractly represented as a single point in a two-dimensional space whose horizontal and vertical coordinates are, respectively, the volume and the pressure of the gas at that moment. During a complete Carnot cycle, the state of the gas traverses the four sides of a closed rhombic figure in this pressure-volume space (see Figure 1.10). One direction of traversal corresponds to standard operation as a heat engine, in which case the area enclosed in this rhombic figure represents the net mechanical work output to the external world. The other direction of traversal corresponds to operation as a heat pump, in which case that same area represents the net mechanical work that must be supplied from the external world.

Defining the efficiency of the engine as the ratio of the net mechanical work (in or out) during each complete cycle to the amount of heat transferred between the two reservoirs, Carnot proved his fundamental theorem: No heat engine working between two given temperature (T_{hot} and T_{cold}) can have an efficiency greater than that of such a reversible engine working between those temperatures. High efficiency requires a large temperature difference. (If the two temperatures are expressed in terms of Kelvin's subsequently developed scale of absolute temperature, the efficiency is computed simply by subtracting the ratio of the two temparatures T_{cold}/T_{hot} from 1.)

Carnot's theorem provided the theoretical basis for assessing the efficiencies of all sorts of subsequently developed engines (of internal combustion, refrigeration, air conditioning, and so on) as well as the original steam engine. In fact, however, the point at which the industrial revolution really "picked up steam" (so to speak) was when the Scottish engineer James Watt, after calculating that the existing Newcomen engine wasted three-fourths of the heat supplied to it, invented the condensing steam engine through a spontaneous thought experiment. Strolling during the enforced idleness of "a fine Sabbath afternoon" in May 1765, Watt later wrote, "the idea came into my mind that ... if a connection were made between the cylinder and an exhausting vessel [the steam] would rush into it and might then be condensed without cooling the cylinder. ... I had not walked further than the Golf house when the whole thing was arranged in my mind." Ultimately, what Carnot began turned out, however, to have implications extending far beyond the realm of human inventions, to naturally occurring processes throught the universe.

Clausius

It was the German physicist Rudolph Clausius who, building on Carnot's thought experiments with ideal heat engines, first enunciated (in 1850) the celebrated Second Law of Thermodynamics: "It is impossible for a self-acting machine, unaided by any external agency, to convey heat from one body to another at a higher temperature" -- although the law was also discovered, independently, and published in the next year by the English physicist William Thomson (Lord Kelvin). The Second Law clearly applies to heat engines of the specific type described by Carnot. If one such reversible engine is converting a flow of heat from the hot to the cold reservoir into a net output of mechanical work, a second such engine would require all of the mechanical output of the first to pump the same quantity of heat from the cold back to the hot reservoir. No further energy would be available to transfer additional heat from the cold to the hot reservoir.

Clausius sought, however, to achieve Carnot's originally stated goal of studying the conversion between heat and mechanical work "from the most general point of view, without reference to any particular engine." Using the abstract pressure/volume state space, he was able to show that a reversible cycle need not be restricted to Carnot's four operations -- those corresponding to the four sides of the rhombic figure in that space. Theoretically, given an ordered series of sufficiently many heat reservoirs having sufficiently closely spaced temperatures, a reversible cycle could be constructed to approximate, arbitrarily closely, any continuous closed curve in that space. For, any such curve could be approximated by a zig-zag path corresponding to a system that is always working between reservoirs of adjacent temperatures through small transformations that are alternately adiabatic and isothermal and, hence, reversible.

By an adroit extension of Carnot's reasoning, Claussius established that all reversible paths of transformation from any one state of the sytem, A, to any other state, B, are thermodynamically equivalent. If, for each successive transformation along a reversible path from A to B, we simply add a certain ratio (the positive or negative quantity of heat transferred, say, to the slightly cooler reservoir divided by the temperature at which the transfer occurred), the resulting sum must be the same regardless of which reversible path we take. Because this sum thus depends only on the starting and ending states (A and B) and not on the (reversible) path taken between them, the assignment of the value zero to a reference state (taken to be the state in which the gas contains no heat) immediately determines a unique value for each other state -- namely, the sum over any reversible path from the reference state to that other state. The unique thermodynamic value thus conferred on each possible state was named, by Clausius, the entropy of the state. It is a measure of the extent to which the energy in the system is in the form of heat that is uniformly distributed and hence unavailable for the performance of mechanical work. The second law of thermodynamics could now be restated in a more general form: Within any closed system, entropy can only increase. This, then, is the basis of such pervasive and familiar facts of irreversibility as that left to themselves, hot coffee always cools and iced tea warms.

Boltzmann

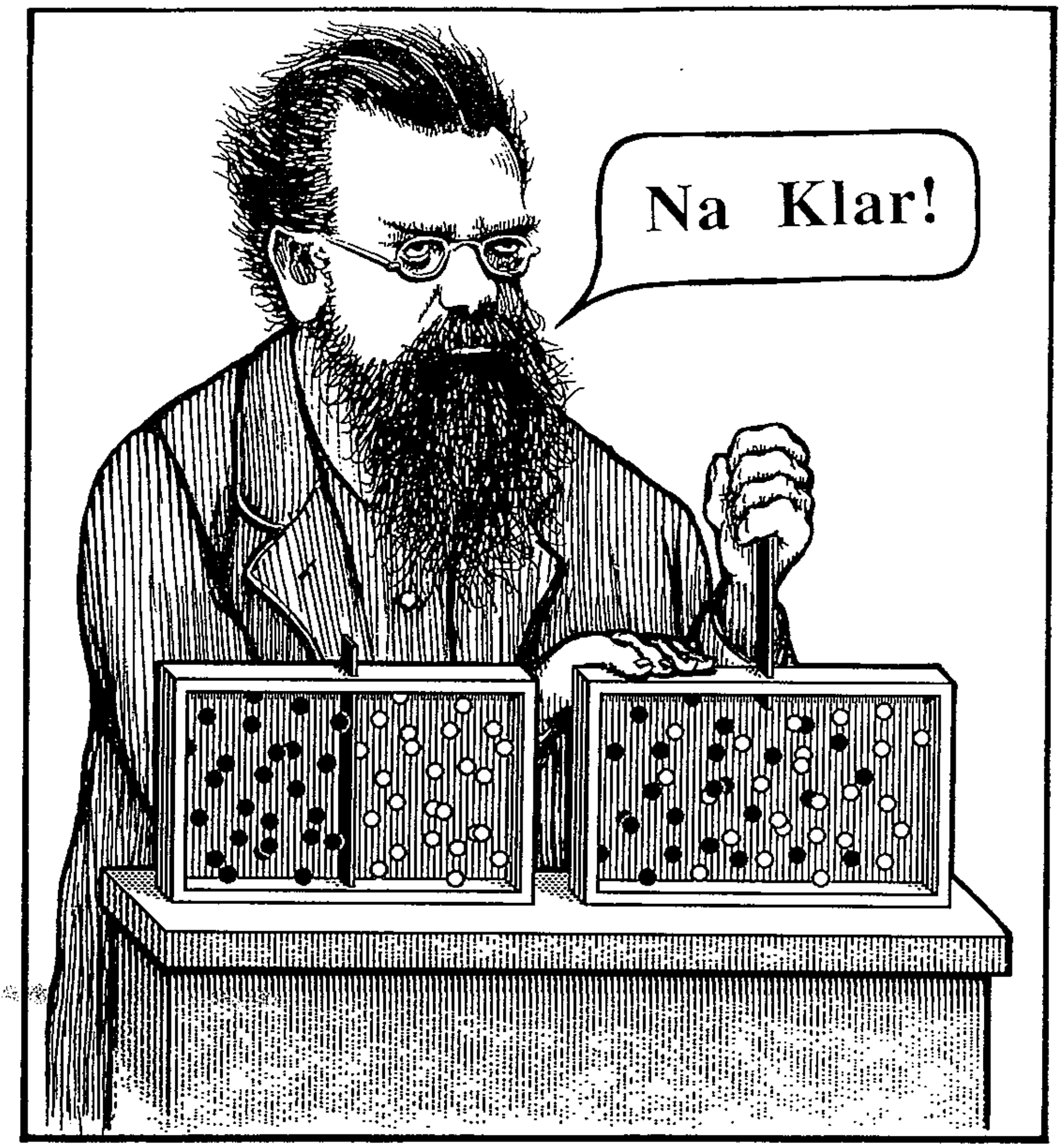
But the second law of thermodynamics raised a vexing problem for the Newtonian world picture. In accordance with Democritus's atomic theory of matter, Robert Hooke (inventor of the compound microscope, as well as Newton's rival) proposed that a body's heat is "nothing else but a very brisk and vehement agitation" of its constituent particles. Similarly, in notes written toward the end of his short life, Carnot himself, accepting the vibratory theory of light (to be considered in my next lecture) and noting that sunlight warms material bodies, embraced the kinetic theory of heat and anticipated Helmholtz's law of conservation of energy when he wrote, "Heat is simply ... a movement among the particles of bodies. Wherever there [appears to be] a destruction of motive power, there is at the same time production of heat ... motive power is in quantity invariable ... it produces sometimes one sort of motion, sometimes another, but it is never annihilated." With the ensuing formal development of the kinetic theory by Joule, Kelvin, and Clausius, however, it seemed that if Newton's laws of motion are universal, they should apply not just to each macroscopic body but to each of the microscopic constituents of each body as well. Science thus confronted -- and not for the last time -- an apparently direct clash between two theories of supposed universal validity. How could the macroscopically observed irreversibility prescribed by a universal second law of thermodynamics be reconciled with the microscopic reversibility prescribed by a universal Newtonian mechanics?

It was the Austrian physicist Ludwig Boltzmann who, in the second half of the 19th century, pointed the way to a reconciliation and, in so doing, founded statistical mechanics. The crucial insight came, once again, from a simple thought experiment (Figure 1.11). Boltzmann imagined a container with a central partition separating the fast moving particles of a hot gas on one side from the slow moving particles of a cool gas on the other. Upon withdrawl of the partition, Boltzmann saw, all particles (fast or slow) that had just then been moving toward the partition would, by Newton's laws, now cross over to the other side. The resulting intermixing of slow and fast particles throughout the container would be macroscopically experienced as an equalization of temperature. The same process would similarly lead to the mixing of particles of (miscible) fluids that were intially separated not by speed but by kind, such as black ones (for coffee) and white ones (for milk).

Bolzmann showed that a purely mechanical definition of entropy simply as the logarithm of the number of micromechanical states of the system that correspond to a particular macroscopically identifiable condition (such as the right side of the container being hot and the left side cold) agrees with Clausius's thermodynamic definition of entropy, which was based entirely on macroscopic observables. At the same time, Boltzmann's definition greatly generalized the concept of entropy by subsuming phenomena of mixing and disorder in which there need not be any disequilibrium of temperature or transfer of heat as such.

According to Newton's strictly deterministic laws, however, if the motion of each particle were exactly reversed, the particles should resort themselves back into their corresponding original sides of the container (hot or cold, black or white). For, there is nothing in the Newtonian laws of motion to say that the motion of any one particle should be more likely in the one direction than in the opposite direction. Why, then, do systems that are left to themselves always evolve only in the one direction -- toward homogenization and equalization? Boltzmann's mathematical analysis showed that for any reasonably large number of particles, their possible positions and velocities consistent with a macroscopically identifiable separation (between hot and cold, or black and white), though astronomically numerous, are as nothing compared to the hyper-astronomically vaster number of possible positions and velocities consistent with the absence of any macroscopically identifiable separation.

An intuitive understanding of how apparent macroscopic irreversibility can arise from microscopic reversibility can be gained from the urn model that Paul and Tatyana Ehrenfest introduced in 1907 to clarify Boltzmann's original thought experiment. Instead of invisible particles moving about in a container with a removable partition, the Ehrenfests imagined an even number (2N) of balls numbered 1 to 2N and distributed in some way between two urns, U_1 and U_2. At each successive time step, a number between 1 and 2N is selected at random and the ball inscribed with that number is transferred from whichever urn it occupies to the other urn. Clearly, such a process is symmetric with respect to time, and could be



Boltzmann
imagines withdrawing a barrier
(Figure 1.11)

continued indefinitely. There is, however, a striking asymmetry between the unbalanced state of the system in which the 2N balls are all in one of the two urns and the balanced state in which they are evenly distributed with N balls in each urn. In the unbalanced case (starting with all balls in U_1), no matter what number is drawn next, the ball bearing that number will necessarily be in U_1 and, hence, will necessarily be transferred to U_2, leading toward a more even balance between the two urns. Granted, at the next time step, a further move toward equalization is no longer absolutely certain. It remains overwhelmingly probable, however -- failing to occur only if the next random draw happens to come up again with the very same number out of the 2N possibilities (thus requiring the transfer of the one ball now in U_2 back to U_1), but such a draw has the very small probability of 1/(2N). Clearly, for large N the distribution of balls will continue to move, with overwhelming probability, toward the equalized distribution (of approximately N balls in each urn). Indeed, a recent mathematical analysis (by Gunnar Blom) shows that the expected number of time steps to reach this equalized distribution approximates $N \log(2 \sqrt{N})$. For a total of one hundred balls (2N = 100), this works out to only 147 time steps (which at the rate of molecular motions would pass in a twinkling).

In the balanced case (beginning with N balls in each urn), only half the numbers that might be drawn would lead toward more balls in U_1; the other half of the numbers would lead toward more balls in U_2. In this case, there is no expectation of a movement toward the distribution in which all 2N balls are in U_1. Indeed, by symmetry, one can conclude that this will remain true, on average, for succeeding time steps. For the same total of one hundred balls in this case, the number of time steps that we should expect to wait before all 2N balls happen to land together in urn U_1 is, by striking contrast, of the order of 9×10^{87}, that is, 9 followed by a string of 87 digits. Indeed, it has been estimated that for molecules numbering 100 million (or 10^8, which is still much smaller than the roughly 10^{25} molecules in a cup of coffee), the percent of the time that a departure of as little as a tenth of a percent from an exactly even distribution between the two halves of the container -- let alone a complete amassing of all molecules to one side -- is as small as 10^{-21}. We can safely conclude that even if the molecules of milk in coffee resorted themselves between the two sides of the cup at a trillion times a second, we should not expect to find a spontaneous accumulation of all the milk on one side of the cup during the entire history of the earth.

Physical implications of the Second Law

The general conclusion to be drawn from the mathematical analyses based on Boltzmann's thought experiment is that any isolated system that we observe changing in an irreversible way must have branched off from an ancestral system from which it inherited its initially improbable, low entropy state of unequal temperature or incomplete mixing. When we do encounter a cup of coffee in which all the milk is on one side, we immediately infer the intervention of an external agency that, for example, has just added the milk on that side.

Nor is the Second Law restricted to the pure processes of thermal equilibration and fluid mixing so far described. It governs all irreversible, that is, dissipative processes. A brick dropped into a pond gives rise to an expanding ring of waves whose energy is dissipated around the perimeter of the pond in, say, small agitations of grains of sand and molecules of air (partly manifested as heat and sound), at different times depending on the distance of each part of the irregular perimeter from the point at which the brick was dropped. Newtonian mechanics would not be violated if molecular fluctuations at these different peripheral points and times all around the edge just happened to produce, in reverse, a perfectly circular wave converging to form a big splash in the middle of the pond. Nor would Newtonian mechanics be violated if a subset of the unimaginably vast number of photons randomly streaking through interstellar space happened similarly to converge on a burnt out star in just such a way as to fission its previously fused atomic nuclei and, thus, to bring the star back to glorious life. But the probabilites in either case say "Don't hold your breath." The anisotropy of time that is evident everywhere around us is accordingly to be found not in Newton's time-symmetric laws of motion but, ultimately, in the time-asymmetric low-entropy initial condition of the universe.

Despite Boltzmann's revelation of the merely statistical basis of the second law of thermodynamics and despite two circumstances (to be considered in my next lecture) in which it seemed for a while that the second law might be violated, that law has overcome all challenges. Few would now take issue with the assessment by Einstein: "It is the only physical theory of universal content concerning which I am convinced that ... it will never be overthrown;" or by Sir Arthur Eddington: "The law that entropy always increases ... holds, I think, the supreme position among the laws of Nature. ... if your theory is found to be against the Second Law of thermodynamics I can give you no hope; there is nothing for it but to collapse in deepest humiliation." Although the second law is only statistical, when powered by sufficiently large numbers, statistics rule. This particular statistical law has, moreover, extraordinarily far-reaching and profound implications -- not least of which is the cosmological fact that all branch systems throughout the observable universe, including all living beings, manifest the same direction of "time's arrow."

Psychological implications of the Second Law

The second law of thermodynamics thus underlies the fact that we ourselves, being such branch systems, have a fundamental orientation with respect to time, remembering only the past and planning only for the future. This connection between the second law and the functioning of minds can be brought out by performing thought experiments on an idealized universe consisting only of bodies of different temperatures moving about in space. Neglecting the inevitable (entropy-increasing) loss of heat through thermal radiation, we can suppose that the temperature of each body remains relatively constant until the body encounters some other body, whereupon the two bodies establish and maintain thermal contact until their temperatures have come to equilibrium at a predictable intermediate value. Having converged to the same temperature, the two bodies then part company and each continues on its separate way until it encounters another body, with which it similarly engages in thermal intercourse.

Every such interaction has the same space-time structure. This can be abstractly diagramed as a short line segment (representing the conjoined existence of the two bodies during their period of connection), with a pair of lines, representing the two separate bodies, branching off at each end of the short segment via a Y-shaped junction, representing (at one end of the segment) the joining together of the two bodies and (at the other end) their moving apart. These two Y junctions always have distinctly different significances, however. The two lines connected at one end of the segment (the earlier end) correspond to bodies at different temperatures, while the two lines connected at the other end (the later end) correspond to bodies at the same temperature.

The same abstract structure would still apply if the intercourse between the bodies were the exchange of bodily fluids (rather than heat) -- for example, if each body consisted of a membrane encapsulating a mixture of two miscible fluids (such as coffee and milk) in some particular proportion. After coupling, complete intermixing of fluids, and then disengagement, each body (which originally had its own distinct proportions of the two fluids) would now contain the two fluids in exactly the same proportions as the other body.

The asymmetry of either type of interaction, thermal equilibration or fluid mixing, has a consistent temporal orientation throughout the universe, permitting a global assignment of a direction for "time's arrow." With the conventional assignment, the universe becomes intelligible. The state of a body is correlated with the states of only those bodies with which it has already interacted in the past -- never with the states of bodies with which it will interact in the future. Scanning such a universe in the usual "forward" direction, we find that causes preceed effects, we can anticipate the consequences of actions, and events can be explained. Run backward, the same universe presents only an unending series of unanticipatable and inexplicable miracles. Differences, instead of being reduced (as specified by the Second Law), continually emerge from nowhere. While two identical bodies embrace, one becomes hot while the other grows cold, one fills with coffee, the other with milk, and nothing in the present or the (now reversed) past furnishes any basis for predicting which body does either, or for explaining why.

That we ourselves remember the past but not the future has a similar thermodynamic and, ultimately, statistical mechanical basis. A memory is just the physical trace left in one branch system as a result of a preceding interaction with another branch system. As such, a memory is analogous to the low-entropy footprint in the normally high-entropy sand that led the shipwrecked Robinson Crusoe to infer that there must already be another person on the island who recently passed that way -- not that another person would subsequently pass that way.

That we plan for the future but not for the past, can similarly be understood in terms of the asymmetry of the interactions between branch systems in our imagined universe of freely moving bodies. If endowed with a homeostatic desire to attain a particular temperature, such a body might seek contact with other bodies in order to gain or to shed heat. In the usual forward direction of time, this body could predict the outcome of thermal intercourse with another body having a higher or lower temperature than its own. No such planning or prediction is possible under time reversal: any other body that is approached from the future side of the interaction is always found to have a temperature exactly equal to its own -- a fact that affords no basis for anticipating the very different temperature of its own body on the other (formerly, the past) side of the interaction.

Culmination of Newtonian Mechanics and Beyond

Boltzmann's reconciliation of thermodynamics with Newton's laws of motion brings us to the culmination of Newtonian mechanics and to the end of my first lecture. Many have exclaimed "Lucky Newton!" echoing Lagrange's sentiment that "we cannot find more than once the system of the world to discover." Yet, in retrospect it seems that Newton's laws could hardly be other than they are, and that they could have be discovered by any one of Newton's predecessors -- including Kepler, Galileo, Copernicus, Archimedes, and even Aristotle. These thinkers would only have had to do two things: First, they would have had to think through, sufficiently fully, the implications of the intuitive knowledge that each of them already possessed about the behaviors of familiar physical objects, such as bricks, apples, and carts. (Emerson might as well have been referring to Lagrange's view of Newton when he wrote: "Men say, Where did he get this? and think there was something divine in his life. But no; they have myriads of facts just as good, would they only get a lamp to ransack their attics withal.") Second, these earlier thinkers would have had to adopt the bold hypothesis that the laws they discovered through their thought experiments on familiar terrestrial objects apply universally to all objects -- including the immense celestial bodies beyond as well as the submicroscopic particles within.

But, as Einstein once observed, "The history of scientific ... discovery teaches us that the human race is poor in independent thinking and creative imagination." This is why, within the range of abilities so far attained by our race, "It required an extraordinary genius," as expressed by Lagrange, "to unravel laws of nature from phenomena which are always before our eyes;" by Paul Valéry, "to notice that the moon is falling, when everyone sees that it doesn't fall;" or by William James, to experience "the flash of similarity between an apple and the moon." In addition, it took Boltzmann's insightful extension of Newton's laws to the hidden realm of Democritus's atoms to show how the quite different and apparently irreversible macroscopic phenomena of heat conduction and fluid mixing were already entailed by Newton's laws, provided only that the universe began in a low entropy state.

In my next lecture, I turn to other phenomena that although also familiar to every person since long before Newton, had been entirely neglected in the construction of Newtonian and statistical mechanics. Lagrange's premise turns out to have been mistaken; Newton's was not the only "system of the world to discover." To Newton's apples, as we shall see, subsequent physicists would find themselves "adding oranges" and perhaps (to borrow from John Updike) even "dividing by grapefruit." Nor, with the advent of quantum mechanics, would the oranges be just "clockwork oranges."