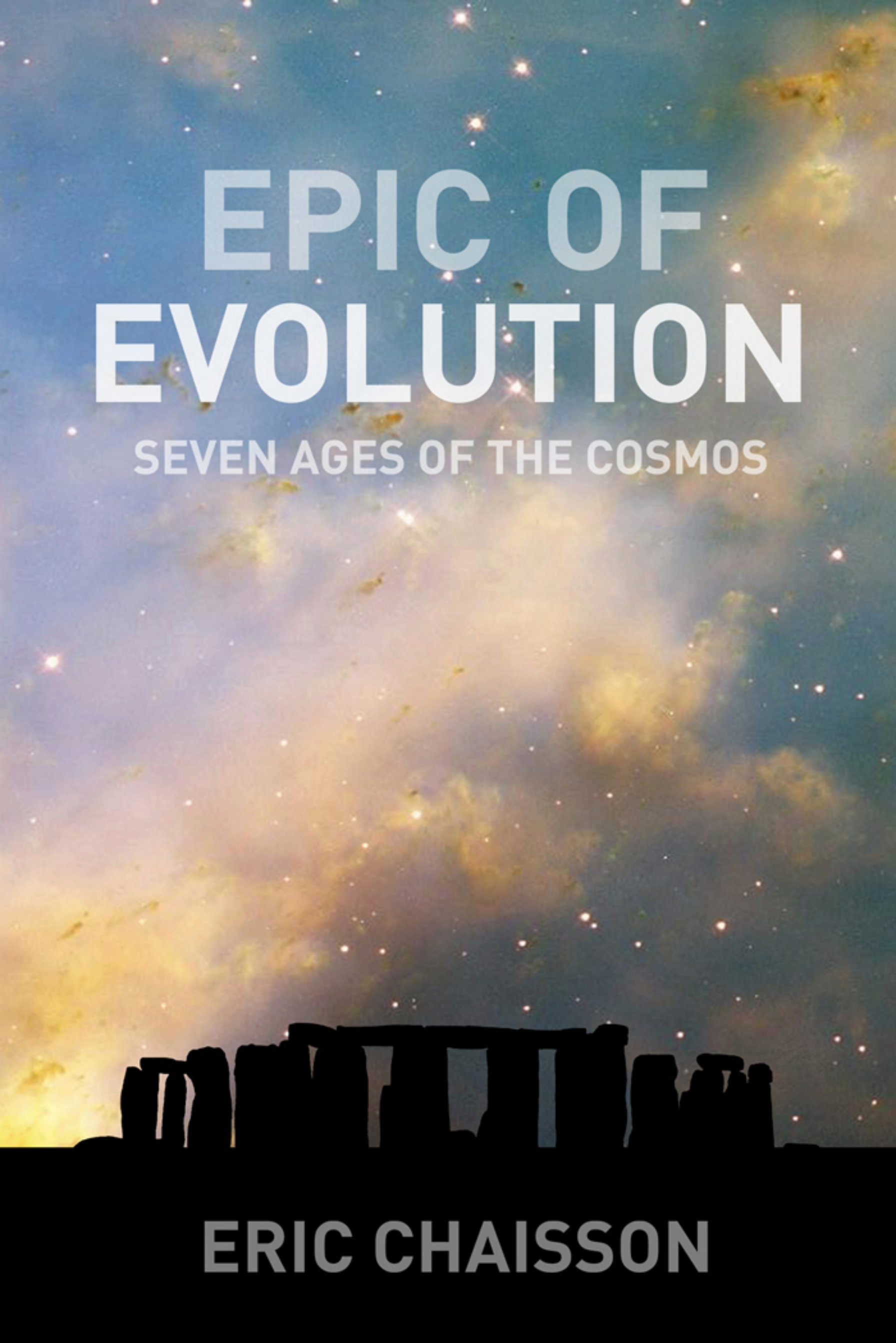

EPIC OF
EVOLUTION
SEVEN AGES OF THE COSMOS
ERIC CHAISSON

EPIC OF EVOLUTION

Other Books by Eric Chaisson

Cosmic Dawn: The Origins of Matter and Life

The Invisible Universe: Probing the Frontiers of Astrophysics (with G. B. Field)

The Life Era: Cosmic Selection and Conscious Evolution

Relatively Speaking: Black Holes, Relativity, and the Fate of the Universe

Universe: An Evolutionary Approach to Astronomy

Astronomy Today (with S. M. McMillan)

The Hubble Wars: Astrophysics Meets Astropolitics in the Two-billion-dollar Struggle Over the Hubble Space Telescope

Astronomy: A Beginner's Guide to the Universe (with S. M. McMillan)

The Thirteenth Labor: Improving Science Education (coedited with T.-C. Kim)

Cosmic Evolution: The Rise of Complexity in Nature

EPIC OF EVOLUTION

SEVEN AGES OF THE COSMOS

ERIC CHAISSON

Illustrated by Lola Judith Chaisson

Columbia University Press

NEW YORK

Columbia University Press
Publishers Since 1893
New York Chichester, West Sussex

Library of Congress Cataloging-in-Publication Data
Chaisson, Eric.
Epic of evolution : seven ages of the cosmos / Eric Chaisson ; illustrated by Lola Judith Chaisson.
p. cm.
Includes bibliographical references and index.
ISBN 0–231–13560–2 (cloth : alk. paper)
1. Cosmology. 2. Life—Origin. 3. Matter—Constitution. I. Title.

QB981.C414 2005
523.1—dc22

2005045452

Columbia University Press books are printed on permanent and durable acid-free paper.
Printed in the United States of America
c 10 9 8 7 6 5 4 3 2 1

To our parents

∞

Contents

Preface

"Everything flows and nothing stays."

—Heraclitus, a Greek philosopher of twenty-five centuries ago

WHEN CONSCIOUSNESS DAWNED among the ancestors of our civilization, men and women perceived two things. They noted themselves, and they noted their environment. They wondered who they were and whence they came. They longed for an understanding of the starry points of light in the nighttime sky, of the surrounding plants and animals, of the air, land, and sea. They contemplated their origin and their destiny.

Thousands of years ago, all these basic queries were treated as secondary, for the primary concern seemed well in hand: Earth was presumed to be the stable hub of the Universe. After all, the Sun, Moon, and stars all appear to revolve around our planet. It was natural to conclude, not knowing otherwise, that home and selves were special. This centrality led to a feeling of security or at least contentment—a belief that the origin, maintenance, and fate of the Universe were governed by something more than natural, something supernatural.

The ancients thought deeply and well, but not much more. Logic was paramount; empiricism less so. Their efforts nonetheless produced such notable endeavors as myth, religion, and philosophy.

Eventually, yet only a few hundred years ago, the idea of Earth's centrality and the reliance on supernatural beings were shattered. During the Renaissance, humans began to inquire more critically about themselves and the Universe. They realized that thinking about Nature was no longer sufficient. Looking at it was also necessary. Experiments became a central part of the process of inquiry. To be effective, ideas had to be tested experimentally, either to refine them if data favored them or to reject them if they did not. The "scientific method" was born—probably the most powerful technique ever conceived for the advancement of factual information. Modern science had arrived.

Today, all natural scientists throughout the world employ the scientific method. Normally it works like this: First, gather some data by observing an object or event, then propose an idea to explain the data, and finally test the idea by experimenting with Nature. Those ideas that pass the tests are selected, accumulated, and conveyed, while those that don't are discarded—a little like the evolutionary events described in this book. In that way, by means of a selective editing or pruning of ideas, scientists discriminate between sense and nonsense. We gain an ever-better approximation of reality. Not that science claims to reveal the truth—whatever that is—just to gain an increasingly accurate model of Nature.

Despite an emphasis on objectivity, some subjectivity does affect the modern scientific enterprise, for this is work done by human beings with strong emotions and personal values. Yet, with the test of time and repeated observations, objectivity eventually emerges and then dominates, enabling us to reach conclusions free of the biased viewpoint of any one scientist, institution, or culture. As a rational investigative approach used to formulate descriptions of natural phenomena, the scientific method is designed to yield a reasonably objective consensus on the nature, contents, and workings of the Universe.

People today still query along the same lines as did the ancients. We ask the same fundamental questions: Who are we? Where did we come from? What is the origin of all things? But our attempts to answer them are now aided by the intricate tools of modern technology: astronomical telescopes to improve our vision of the macroscopic realm of stars and galaxies; biological microscopes to display up close the minute world of cells and molecules; particle accelerators to probe the subatomic domain of nuclei and quarks; robotic spacecraft to gather facts unavailable from our vantage point on Earth; powerful computers to keep pace with the prodigious flow of new data, tentative ideas, and experimental tests.

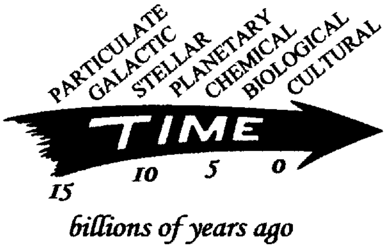
We live in an age of technology—a time of rapid intellectual advancement unprecedented in history. And even though technology threatens to overwhelm us—perhaps even replace us—that same technology now provides us with a remarkable, yet still growing, understanding of ourselves and our richly endowed Universe.

Of all the scientific achievements since Renaissance times, one discovery stands out most boldly: Our planet seems neither central nor special. Use of the scientific method has demonstrated that as living creatures, we inhabit no unique place in the cosmos. Research, especially within the past few decades, strongly implies that we live on an ordinary rock called Earth, one planet orbiting an average star called the Sun, one star in the suburbs of a much larger swarm of stars called the Milky Way, one galaxy among billions of others spread throughout an observable abyss called the Universe.

Now, at the beginning of a new millennium, modern science is helping us construct a truly big picture. We are coming to appreciate how all objects—from quark to quasar, from microbe to mind—are interrelated. We are attempting to decipher the scenario of *cosmic evolution*: a grand synthesis of many varied changes in the assembly and composition of radiation, matter, and life throughout the history of the Universe. These are the changes, operating across almost incomprehensible domains of space and nearly inconceivable durations of time, that have given rise to our galaxy, our star, our planet, and ourselves.

To be sure, change is ubiquitous in Nature. Some of that change is subtle, such as when our Sun shines daily or Earth's continents drift slowly. Other change is more dramatic, such as when massive stars explode catastrophically as supernovae or when landmasses fault suddenly as quakes and volcanoes. Regardless of whether Nature is examined macroscopically with a telescope, microscopically with an accelerator, or mesoscopically with our own eyes, we see change. Thus, we give this process of universal change a more elegant name—cosmic evolution, which includes all aspects of evolution: particulate, galactic, stellar, planetary, chemical, biological, and cultural.

Emerging now is a unified worldview of the cosmos, including ourselves as sentient beings, based upon the time-honored concept of change. Change—to make different the form, nature, and content of something—has been the hallmark in the origin, evolution, and fate of all things, animate or inanimate. From galaxies to snowflakes, from stars and planets to life itself, we are beginning to identify an underlying pattern penetrating the fabric of all the natural sciences—a sweepingly encompassing view along the “arrow of time” of the formation, structure, and function of all objects in our multitudinous Universe.



. . . a sweepingly encompassing view along the “arrow of time” . . .

Heraclitus of old Greece had it correct: Everything flows; nothing is permanent except change. It’s perhaps the best observation anyone ever made, minus the devilish details. Today, some twenty-five centuries later, scientific researchers are steadily discovering many of those details—and the results are both insightful and unifying, even awesome. We now have a reasonably good understanding not only of how countless stars were born and have died to create the matter composing our world but also how life has come to exist as a natural consequence of the evolution of matter. We can reliably trace a thread of knowledge linking the evolution of primal energy into elementary particles, the evolution of those particles into atoms, in turn of those atoms into stars and galaxies, the evolution of stars into heavy elements, and of those elements into the molecular building blocks of life, and furthermore the evolution of those molecules into life itself, of advanced life-forms into intelligence, and of intelligent life into cultured and technological civilization.

∞

To answer the fundamental questions of who we are and whence we came, we need to probe far back into the past—beyond our birth dates some tens of years ago, beyond Renaissance times centuries ago, beyond the onset of civilization some ten thousand years ago, beyond our ancestral hominids who emerged from the forests several million years ago, even beyond the time when multicellular life began to flourish on our planet about a billion years ago, some million millennia before now.

To appreciate our deep origins in a cosmic-evolutionary setting, we must broaden our horizons, expand our minds, and visualize what it was like long, long ago. Go back, for instance, five billion years, when there was no life on Earth, indeed no planet Earth itself. Nor were there a Sun, a Moon, or a Solar System. These objects were only then forming out of a giant, swirling gas cloud near one edge of a vast galaxy of older stars that had already existed in one form or another for a long time before that.

Modern science now combines a wide variety of curricula—physics, astronomy, geology, chemistry, biology, anthropology, among others—in an interdisciplinary attempt to address the two most fundamental issues of all: the origin of matter and the origin of life. If we can decipher the scenario of cosmic evolution, then perhaps we can determine precisely who we are, specifically how life originated on this planet, and, incredibly enough, how living organisms managed to invade the land, generate language, create culture, devise science, explore space, and even study themselves.

As sentient beings, we humans now reflect back on the matter of the Universe that gave us life. And what we find is a natural history, a universal history, a rich and abiding story of our origins that is nothing less than an epic of creation as understood by modern science—a coherent *weltgeschichte* that people of all cultures can adopt as currently true as truth can be.

∞

These writings concern all these things: space and time, matter and life, and the energy exchanges that infuse them. We herein explore our cosmos, our planet, ourselves. We summarize where science stands today regarding answers to some of the time-honored philosophical questions: Who really are we? Where and when indeed did we come from? How did everything around us—the air, the land, the sea, the stars—originate? What is the source of the order, form, and structure characterizing all material things? How do we, as intelligent beings, relate to the rest of the Universe? In short, what are our origin and our destiny? What are the origins and destinies of Earth, the Sun, the Universe?

Written for eclectic individuals having a broad interest in Nature, this book explains valid contemporary science in a mostly nontechnical manner. Accuracy has not been sacrificed, however, and a feeling for the frontiers of science has been included. Even so, readers must recognize that answers to some of the most basic queries are not yet entirely clear. Even among colleagues, scientists are often unable to provide precise and complete solutions for great and profound questions. Only within the past few decades have we gained the technological expertise needed to transfer these issues from the realm of philosophy to that of science.

. . . sentient beings . . . now reflect back on the matter of the Universe that gave us life

Researchers now sense that the cutting edge of knowledge resembles a thinning haze rather than a sharp boundary. The research front resembles the "fog of war," meaning that scientific work is rarely crystal clear in real time, while the work is underway; rather, the intellectual landscape is often revealed only later, after the subjective confusion has abated and a certain objective reality has emerged. That's because the enterprise of science is now advancing rapidly, acquiring new information at a phenomenal rate, and requiring novel interdisciplinary ventures to sort it out. Less than a hundred years ago, we didn't understand how stars shine, how heredity works, that the Universe is filled with galaxies, or even that it had a definite beginning. Furthermore, much of science "as a work in progress" involves the human condition, which ensures many false starts and occasional botched logic among the many valid, proven ideas. As a fair assessment, we might say that a pencil sketch of the answers to some of the most basic questions is now at hand, but that many specifics are yet wanting.

In a descriptive and illustrative way, then, we probe here the essential nature of the cosmos. These pages render the prevailing scientific view that the atoms in our bodies relate to the Universe in general. We elucidate the modern paradigm of cosmic evolution—an astrobiology, a cosmogenesis, a whole new scientific philosophy—whereby changes, both gradual or episodic and generative or developmental, in the composition and structure of matter have given rise to galaxies, stars, planets, and life. We attempt to synthesize the essential ingredients of astrophysics and biochemistry, for these two subjects, more than any others, are greatly affecting our philosophical conceptions of ourselves as human beings and of our place in the Universe.

In short, this book presents the broadest view of the biggest picture. It analyzes, using the best science available, some of the most fundamental questions of all—neither the most relevant nor the most practical questions, perhaps, for twenty-first-century society, but deeply fundamental ones. We develop an appreciation for our rich universal heritage, for an expansive perspective like no other. We seek to know the nature and behavior of radiation, matter, and life on the grandest scale of all. And in deciphering the fabric of Nature, we discover that technological humans now reside at the dawn of a whole new era.

∞

This book is an extensive rewrite of an earlier one, *Cosmic Dawn: The Origins of Matter and Life*, that I authored some twenty-five years ago. That original book, based on an interdisciplinary course that I cocreated at Harvard University in the 1970s (and that I still teach there), was wonderfully received by both students and public alike. Even colleagues uncharacteristically acknowledged it, despite its popularized account, awarding it several literary prizes. Yet much has occurred in the world of science in the intervening decades. Researchers around the globe have acquired vast amounts of new observational data and have gained more theoretical insight into many aspects of cosmic evolution. The intellectual framework has remained much the same, but the details have become richly enhanced.

Astronomers now have intricate models of the early Universe and of the galaxies that formed long ago but have not yet solved some of the most formidable cosmological puzzles. Biologists now better understand the rate and tempo of life's evolution while reaffirming the essence of neo-Darwinism, yet they still debate the mechanisms of change that might supplement the principle of natural selection. Environmentalists have greatly improved their ability to monitor Earth's biosphere yet are unable to predict the adverse long-term trends in climatic change. Chemists now more accurately simulate conditions that likely gave rise to the origin of life, geologists build exquisite maps of Earth's interior to aid comparative planetology, and anthropologists have accumulated a wealth of bones and artifacts from which to unravel our human past—but problems remain everywhere among those devilish details.

Of equal importance to those advances made in the particular disciplines, science during the past decade has also become more interdisciplinary. Highly focused researchers now talk to colleagues across specialized boundaries—astronomers to paleontologists, cosmologists to particle physicists, biologists to mathematicians, neurologists to computer scientists. The breakdown of academic barriers is long overdue, as "thinking out of the box" is increasingly valued today. And with many fields now moving from reductionist to integrationist approaches, multidisciplinarity is in vogue for the twenty-first century. We are entering an age of synthesis, when the drive toward unification is once again at the fore.

That said, my attempted unification concerns what is empirically observed "out the window" in Nature—mainly, detectable things in the world around us, such as atoms, stars, plants, and animals. I see no evidence for cosmic strings, eleven dimensions, or multiple universes. Nor do I feel the need to embrace anthropic reasoning. The weak anthropic principle—that sentient beings eventually emerge in the Universe—is hardly more than cosmic evolution at work, whereas the strong principle—that the Universe is made for us—seems nothing more than teleology at play. Rather than appealing to Providence or "multiverses" to justify the numerical values of some physical constants (such as the speed of light or the charge of an electron), I prefer to reason that when the laws of science become sufficiently robust, we shall naturally understand the apparent "fine-tuning" of Nature. It's much akin to mathematics, when considering the value of π. Who would have thought, a priori, that the ratio of a circle's circumference to its diameter would have the odd value of 3.14159 . . . ? Why isn't it just 3, or 3.1, or some other crisp number, rather than such a peculiar value that runs on ad infinitum? We now understand enough mathematics to realize that this is simply the way geometry scales; there is nothing mystical about a perfect circle—yet it surely is fine-tuned, and if it were not it wouldn't be a circle. Circles exist as gracefully rounded curves closed upon themselves *because* π has the odd value it does. Likewise, ordered systems in Nature, including life, likely exist *because* the physical constants have their decidedly odd values.

Gratifyingly, the concept of pervasive change on all scales remains much as I initially envisioned in *Cosmic Dawn.* If anything, the story of cosmic evolution has been strengthened by advances in nonequilibrium thermodynamics, a frontier subject that models the flow of energy through open, complex structures—whether those structures are galaxies, stars, planets, or life. To be sure, a great deal of new meat has been placed on the bones of the skeletal structure first outlined more than two decades ago.

Much revising, updating, and enlarging has gone into this new book. While still preserving the broad scope, chronological sequence, and literary style that made the original book accessible to a wide audience, I have:

- overhauled completely the science content, bringing everything up to date and thus bolstering the scenario of cosmic evolution with the latest scientific findings;
- supplemented the pencil-sketch drawings of the central ideas with two dozen photographs that provide much observational evidence for those ideas;
- reorganized entirely the chapters on chemical and biological evolution to give each more coverage and to incorporate recent scientific advances;
- provided a glossary of key terms, which are especially helpful for such a wide-ranging, interdisciplinary subject that crosses so many scientific boundaries.

∞

To make the scenario of cosmic evolution readable for a general audience, I have avoided referring in the text to any living authorities. To cite each of the specialist researchers now contributing to the subject would detract from the clarity of the concepts stressed throughout; the apportionment of credit to individuals is less important than the big picture granted by the sweep of the subject writ large. Suffice it to say that the narrative described here is based on countless scientific results advanced by legions of specialists working across the entire spectrum of human knowledge. The bibliography at the end of the book, which may be consulted for further reading, lists a sampling of many fine works that I found useful while synthesizing this survey from big bang to humankind.

Many colleagues have helped mold my views on the grand themes and intricate details of cosmic evolution; some of them have influenced the way I teach, write, and research this highly inclusive subject. I remain especially indebted to George Field and the late Harlow Shapley, both former directors of the Harvard College Observatory—the first for inviting me to join him in exploring this interdisciplinary subject at the start of my professional career a quarter-century ago, and the second for inspirationally paving the way in cross-boundary teaching and research (which he called "cosmography") more than a half-century ago. I am grateful to my wife, Lola, who, in drawing all the freehand illustrations in this work, has beautifully combined the thought-provoking aesthetics of an artist with the technical accuracy of a scientist. Michael Haskell, Robin Smith, Fred Spier, and an anonymous reviewer offered close reading of the manuscript that improved its content and style. Above all, I thank the nearly four thousand students who have taken my course on cosmic evolution during the past generation and who, by embracing its only prerequisite—"persistent curiosity"—have helped crystallize my thoughts and insights on this powerful worldview for the twenty-first century.

EPIC OF EVOLUTION

Prologue

COSMOLOGICAL OVERVIEW

EXPLORING THE ENTIRE UNIVERSE requires big thinking. And there are hardly bigger ideas than cosmological ones. Cosmology is the study of the structure, evolution, and destiny of the Universe—the totality of all known or supposed objects and phenomena, formerly existing, now present, or to come, taken as a whole. Here we strive to gain an appreciation for the properties of the Universe in bulk: its matter and energy, its size and scale, perhaps something about its origin and fate.

Cosmic issues elicit grand perspective, and rightly so. Compared to the whole Universe truly writ large, its smaller contents such as planets and stars—even galaxies, to a certain extent—become nearly inconsequential. To the cosmologist, planets are hardly relevant, stars only point sources of hydrogen consumption, and galaxies mere details in the much broader context of all space.

Time also shrinks in significance when compared to eternity. Reckoning change on human scales pales in comparison to all change on the cosmological stage. Durations of a thousand years seem like nothing, a million years a mere wink of an eye in the cosmic scheme of things. Even a billion years is a rather short interval in the context of all time.

To appreciate cosmology, we must broaden our view and expand our minds to include all of space and all of time. If we have ever wanted to think big, now is the time!

At the outset, take note: Thousands, millions, billions, and even trillions can be used easily in words. Not only are these enormous numbers, but the differences among them are also large. For example, one thousand is familiar enough to understand well; at the rate of one number per second, we could count to a thousand in about fifteen minutes. By contrast, to reach a million surprisingly requires more than two weeks, counting at the rate of one number per second, sixteen hours a day (allowing eight hours a day for rest). And a count from one to a billion, at the same rate of one number per second for much of each day, would take some fifty years. Internalize that fact: nearly an entire human lifetime is needed just to count to a billion!

Here, we shall routinely consider time intervals spanning millions and billions of not merely seconds but rather whole years. And we shall discuss objects housing trillions upon trillions of atoms, even trillions of whole stars. Hence, we must become accustomed to gargantuan numbers of things, enormous domains of space, and extremely long durations of time. Recognize especially that a million is much larger than a thousand, and a billion, much, much larger still.

∞

Viewing the Universe from our vantage point at Earth, we see an abundance and variety of objects and phenomena. Among them are gassy nebulosities glowing with colorful light, explosive stars ejecting matter and energy, and powerful galaxies spinning in the depths of space. Through a telescope on a dark, moonless night, celestial objects present superb examples of astronomical architecture—real jewels of the night. But astronomical bodies are more than works of art, more than objects of elegance. Each is a rich repository of light illuminating a material aspect of our Universe. To the cosmic evolutionist, planets, stars, nebulae, novae, galaxies, and all the rest are of vital significance if we are to realize our human place in the big picture. This intellectual placement of humankind in the wider cosmos will emerge later in this book; for now, we focus on the grand issues addressed by the cosmologist.

Light is only one type of radiation—namely, that type to which our human eyes are sensitive. As light enters our eye, the cornea and lens focus it onto the retina, whereupon small chemical reactions triggered by the incoming light send electrical impulses to the brain, producing the sensation of sight. By contrast, radio, infrared, and ultraviolet waves, as well as X rays and gamma rays, are all invisible radiation, and each goes undetected by human eyes. But regardless of the type, radiation is energy, that physical property best characterizing (and driving) change. Radiation is also information—a primitive form of information that moves from one point to another, such as from a star to our eyes. It is only by means of such one-way information flows that we can hope to fathom the depths of space.

Practitioners of astrophysics acquire information about cosmic objects by interpreting their emitted radiation. We say "astrophysics" because that word best defines the basis on which the interpretations are made. These days, the emphasis is on physics: "astro" is a mere prefix. The space scientist of today who doesn't have a firm grounding in physics is hardly a space scientist at all. Gone are the romantic evenings when individual astronomers made fundamental discoveries by peering through long telescopes and marveling at the sights; gone also are the thick catalog tabulations and stacks of exposed photographic plates. The modern astrophysicist wants to know more than just where objects are or what their brightness and colors may be. Contemporary astronomy has become more of an applied physics than the classical astronomy of old.

Astrophysicists are driven more than most by a need to understand how Nature functions. We not only want to perceive what lurks beyond the range of human eyesight, what the Universe "looks" like in the invisible domain—which is, by the way, where most matter radiates. We also seek to know *how* the myriad celestial objects came to be, *how* they operate in detail, *how* matter and radiation interact, and especially *how* energy guides the ceaseless changes among all known cosmic systems. We are intellectually transitioning from addressing only *what* questions to the more penetrating *how* questions.

In a way, astronomers and astrophysicists have been commissioned by society to keep an eye on the Universe. Our job is to inventory the cosmos, to seek a complete account of the state and nature of all the different types of matter beyond planet Earth. Likewise, the newly emerging field of "astrobiology" seeks to inventory life in the Universe, although thus far life on Earth is our only confirmed example. In contrast to the abundant databases of modern astrophysics, astrobiology is a subject for which there are as yet no data. If and when life is found elsewhere beyond Earth, the interpretive emphasis will be on the biology in a cosmic setting.

Note the essential difference between the majority of scientists, who study terrestrial matter in laboratories on Earth, and astroscientists, who investigate remote, alien matter far from our home planet. On Earth, scientists can control their experiments as an aid to discovering a wealth of properties among terrestrial matter. They can both tangibly manipulate the matter under scrutiny and tinker with the experimental equipment used to inspect it. In the case of a new rocky ore, for example, laboratory scientists could examine its properties by sampling a variety of rocks, each having a different size, shape, or composition. They could probe the ore in many ways—vigorously heating it or cryogenically cooling it, even subjecting it to varying amounts of electricity and magnetism. Or they could just hit it with a hammer, which geologists often do. All the while, researchers would learn a great deal about the rock by testing its responses to many environmental changes. In short, the medium in which a terrestrial experiment operates can be intentionally altered in various ways in order to enhance the study of a piece of local matter.

Distant matter far beyond our planet, however, cannot be so massaged, not even with the very best tools of modern civilization. Remote extraterrestrial environments can be neither controlled nor manipulated. For the most part, astronomers are restricted to working with intangible radiation emitted by cosmic matter—radiation occasionally intercepted by human eyes or detected by earthly instruments, signals momentarily captured while traveling from faraway objects to faded oblivion elsewhere in the dim recesses of the Universe.

Technological advances have recently provided a few exceptions to these statements, enabling space scientists to perform guided experiments on a handful of specimens from nearby extraterrestrial regions: interplanetary meteorites discovered buried in Earth's crust and especially its icy polar caps, lunar rocks retrieved from our dead neighbor via the American and Russian space programs, and Martian soil examined by robot spacecraft now parked on the plains of that alien planet. Yet it's likely to be centuries before our descendants gain the means to conduct hands-on exploration of matter beyond our own Solar System. For now and for a good long time to come, the bulk of cosmic matter must be inventoried and analyzed by extracting information veiled within naturally emitted radiation that just happens to be captured by equipment on or near Earth.

For the time being at least, radiation is the only means whereby we know of the existence of virtually any celestial object.

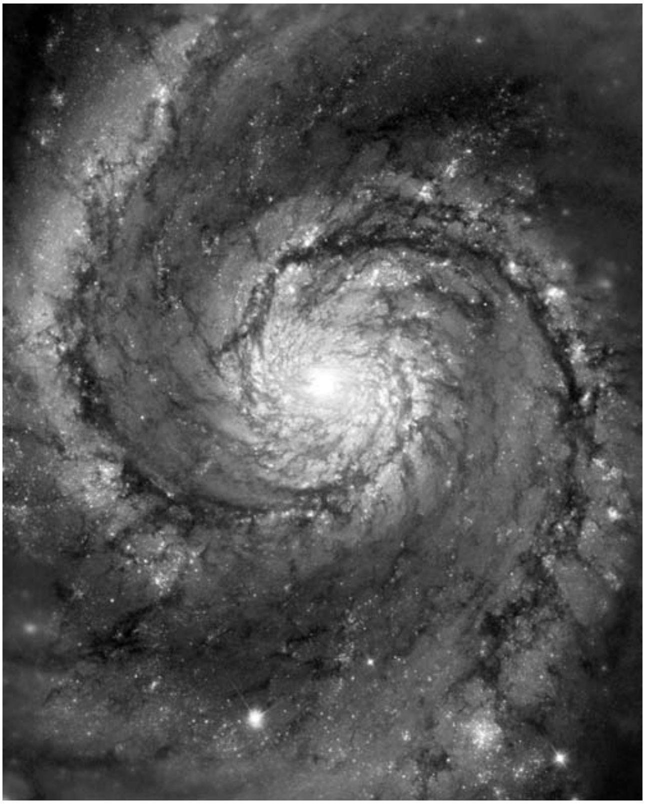
A further restriction comes to mind when contemplating distant extraterrestrial matter. Not only are we prohibited from studying celestial objects at their present locations in space, but we are also denied the chance to examine them now in time. The reason is that radiation does not travel infinitely fast; it moves at a finite speed—the velocity of light. Consequently, it takes time—often lots of time—for light or any type of radiation to travel through the vast expanses of space separating objects in the Universe. Yet most people don't realize the long time intervals needed even for light to traverse the great realms beyond our home in space.

The bright red star in the northern winter constellation Orion provides a classic example. Betelgeuse is known to be a bit more than four hundred light-years away—a terribly long range given that a light-year is the *distance* traveled by light in a full year at the fastest velocity known. One light-year equals about ten trillion kilometers, or six trillion miles; even a light-day measures some thirty billion kilometers. So radiation is fast, there is no doubt—which makes the distance to this relatively nearby star all the more impressive. To be sure, Betelgeuse's radiation takes more than four centuries to travel to Earth. Since nothing known surpasses the velocity of light, its radiation simply could not get here any quicker. Expressed another way, the light we see while looking at Betelgeuse tonight left that star before the invention of the telescope. It has been cruising through the near void of outer space ever since.

The nearest spiral galaxy, called Andromeda for short, provides an even more dramatic example of light's finite speed. It, too, can be seen with the naked eye as a fuzzy "cotton ball" just south of the bright, sharp stars of the big-W constellation Cassiopeia in the northern summer sky. Roughly two-and-a-half million light-years distant, this galaxy's radiation takes some twenty-five thousand centuries to reach us—meaning that Andromeda's light left that galaxy well before *Homo sapiens* emerged on planet Earth. And yet it's the *nearest* major galaxy to us!

Radiation from distant objects, therefore, harbors clues to the past—but not to the present. The farther an object is from Earth, the longer its light takes to reach us. In the case of the truly remote galaxies, some of which are billions of light-years away, radiation left those objects well before Earth or the Sun even formed. In fact, radiation now reaching us from the most distant cosmic objects was launched in earlier epochs of the Universe, when none of the familiar stars and planets yet existed.

By collecting radiation, astronomers can learn what the conditions were like long ago when distant objects emitted their light. The light it-



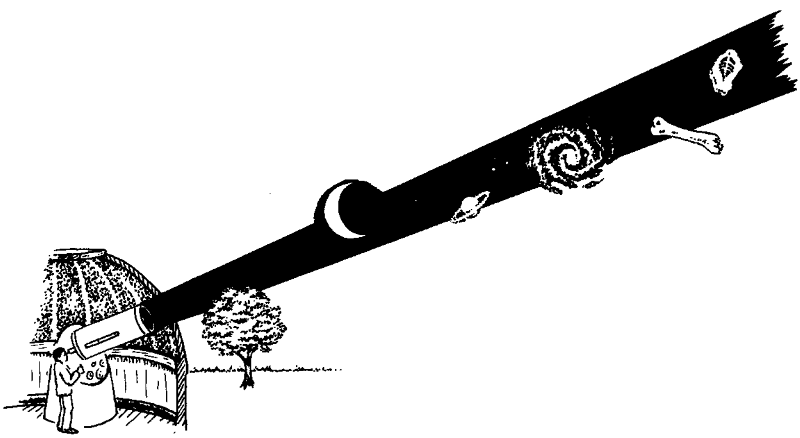
Evidence of extraterrestrial objects.

A typical galaxy is a collection of a couple hundred billion stars, each separated by vast regions of nearly empty space. Shown here face-on is the Whirlpool Galaxy, a lens-shaped colossal spiral roughly thirty million light-years away. It measures about a hundred thousand light-years across, or a thousand quadrillion kilometers. Our Sun is a rather undistinguished star near the edge of another such galaxy, called the Milky Way. *Source: Space Telescope Science Institute.*

self resembles a letter mailed some time earlier; the letter's contents grow no older while being delivered, thus bringing to the recipient information about the time when the letter was written. Likewise, light embodies data about earlier times when the light was launched; the light itself does not age. Deciphering the information within that radiation, we can not only determine the general conditions in the Universe before the dawn of the Sun and Earth, but we can also specify values for the two most important factors—temperature and density—characterizing the Universe in some of those ancient times.

Our perspective of the Universe is delayed. We see the Universe as it was, not as it is. Even more useful than the wish of many philosophers that light speed be infinite so as to reveal the whole Universe presently, is the fact that precisely because light speed is finite we can discover a fascinating record of many past events, including perhaps knowledge of our own cosmic origins.

Astronomers, then, are the ultimate historians; our telescopes, effectively time machines. We go all the way back (or nearly so) into "deep time," indeed times much, much earlier than those studied by scholars traditionally called historians—before Rome, before Egypt, to be sure well before any recorded history. Looking out from Earth, we see a "big history" of the Universe arrayed before us, including epochs early enough to reveal ways and means that may have led to our being. Much like anthropologists who sift through ancient rubble for bones and artifacts containing hints and clues about the origin and evolution of human culture, astrophysicists dissect radiation only now arriving at Earth, seeking to interpret its embedded information about the origin and evolution of matter itself.



Looking out in space is equivalent to probing back in time.

So never forget the cosmologists' dictum: *Looking out in space is equivalent to probing back in time.* We do not perceive the Universe as it is now; rather, we see it progressively younger the farther out we probe. Since our field of view extends for billions of light-years into space, we necessarily explore billions of years earlier in time. By examining deep space and capturing radiation from the most distant objects, researchers gain an increasingly better picture of what the Universe was like long ago, including near the time when time itself began. This is the task before us—to construct a chronological narrative that relates, using the best science available, how all things came to be.

∞

Cosmic activity permeates the Universe, yet so does quiescence. Perspective often determines which dominates. Surveyed casually, celestial objects usually display stability. Yet higher resolution often reveals some violence. Generally, the larger the perspective, the more stable things seem. For example, that our Earth is ruptured by quakes and volcanoes is obvious to those of us who live on it and witness its daily activity up close, but our planet appears tranquil when viewed from afar in those striking lunar earthrise photos taken by the *Apollo* astronauts. Likewise, telescopic studies of our Sun show it to be peppered with bright flares, dark spots, and surface explosions, as are presumably all stars; yet to the naked eye, the Sun and most stars assume a rather peaceful, steady pose.

We might then expect that while pockets of violence will be surely tucked here and there throughout the fabric of the Universe, the largest possible, cosmic perspective would display perfect quiescence. Not so, however. In bulk, the Universe is not calm and stable. Surprisingly, the whole Universe in toto displays much dynamism.

Knowing, then, that the Universe harbors a certain verve, we might further expect the largest material structures—among them the galaxies—to have random, disordered motions, some hurtling one way and some others another. Chaotic motions of fireflies trapped in a jar come to mind, or the nearly scattered motions of skaters in a hockey rink. For the Universe, however, these are not good analogies. Our expectations are wrong again, for the galaxies are not moving chaotically. The Universe is indeed active, but in an awesomely ordered fashion.

For well more than half a century now, scientists have realized that galaxies have some definite organized movement in space—a universal traffic pattern of sorts. Surprisingly, virtually all the galaxies are steadily receding, propelled away from us as though we had a kind of cosmic plague. (Only a few nearby galaxies, including neighboring Andromeda, are known to have a component of their velocity toward us, but that's due to the random, small-scale motions that all galaxies display in addition to their more directed, large-scale recession—like confused fireflies in a jar that has been heaved away, which *is* a good analogy.) What's more, the galaxies are also receding in a grand overall manner. Each one drifts away at a velocity proportional to its distance from Earth. This is a fact of great significance: the greater the distance of an object from us, the faster that object recedes. These two quantities—velocity and distance—are highly correlated.

Astronomers know this because the galaxies' light is red shifted—that is, stretched to longer wavelengths because of the Doppler effect. Just as sound waves from a police car's siren seem to produce a higher pitch when the vehicle approaches and a lower pitch while moving away, so light waves from an approaching object are squeezed to shorter wavelengths—toward blue—and stretched to longer wavelengths—toward red—as it recedes. The extent of the shift, which occurs in light much as it does in sound, reveals how fast the object is traveling. To be sure, the Doppler effect is also used to spot speeders on the highway and to measure the speed of a fastball at the park.

Now, if we think about it for a moment, the entire pattern of distant objects receding more rapidly than nearby ones implies that an "explosion" must have occurred at some time in the past. Visualizing the past by mentally reversing the outward flow of galaxies, we reason that all such galaxies were once members of a smaller, more compact, and hotter Universe. The more distant an object is from us, the more forcefully it—or whatever preceded it—must have been initially expelled; their greater distances result directly from their greater velocities. In other words, the faster-moving galaxies are by now farther away *because* of their high velocities. This is precisely the flight pattern of shrapnel fragments when a conventional bomb explodes. The galaxies are simply the debris of a primeval "explosion," a cosmic bomb whose die was cast long ago.

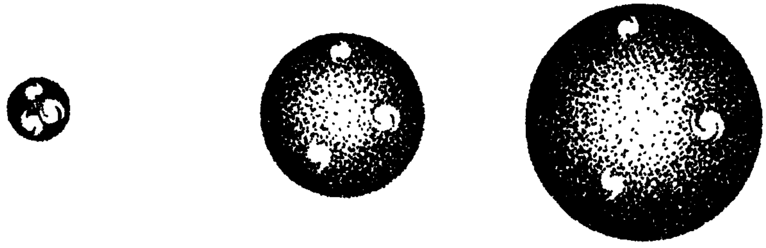
The word *explosion* is in quotes above because, technically, most astronomers don't like that description. Since there was no preexisting space, nor any matter per se at the start, that word can be misleading. Yet if we keep this bomblike interpretation in mind as merely artistic license—here with energy initially expelled into time, rather than matter into space—then the analogy serves a useful purpose.

This implied, titanic event is commonly known as the "big bang," a derisive term introduced by skeptics who decades ago preferred a more steady, less violent Universe. But the term has stuck and is now synonymous with the standard model of cosmology—a widely accepted description of macroscopic phenomena on the largest scales. Note again and despite the word "bang" that the primordial matter did not actually explode into any already existing space, nor are the galaxies now moving through space or rushing into "empty space" beyond. Rather, owing to the initial conditions at the moment of the big bang, space itself began expanding at high speed, much like a crumpled fabric rapidly unraveling. The galaxies now seen are part of that expanding fabric of space, or perhaps more like raisins in a baking bread, and are basically "along for the ride."

Recessional motions of the galaxies virtually prove that the whole Universe itself is in motion. On the largest scale of all, the Universe is active and by no means a pillar of stability. Instead, much like everything within it, the Universe changes with time—in short, evolves.

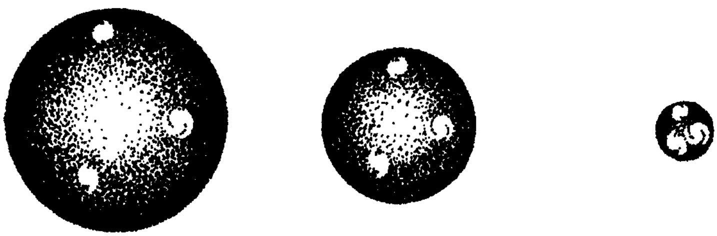
Be assured that neither Earth nor the Solar System nor individual galaxies are physically ballooning in size. Planets, stars, and galaxies are all gravitationally bound, intact systems. Only the largest framework of the Universe—the ever-increasing distances separating galaxies and especially clusters of galaxies—manifests cosmic expansion.

Astronomers, philosophers, theologians, as well as people from all segments of society would like to know if the Universe will continue to expand forever or whether its expansion will someday stop. It's the destiny issue, hereby scientifically stated: If the Universe eternally expands, unimaginable amounts of time would be available for the continued evolution of matter and life. By contrast, if the Universe embodies enough matter, the combined pull of gravity could conceivably bring the expansion to a halt and even reverse it into contraction.



. . . will the Universe continue to expand in this way forever?

Several questions come to mind: How long has the Universe been expanding? How much more time will elapse before it ceases expanding? If the Universe does start to contract, what will happen upon its eventual collapse? Will the Universe simply end as a small, dense point much like that from which it began? Or will it perhaps bounce and begin expanding anew? If the Universe has rebounded in this way before, we may well inhabit a cyclically expanding and contracting Universe—one having a continuous cycle of birth, death, and rebirth, though neither a true beginning nor an ultimate end.



. . . or will it contract to a virtual point and end?

These are the basic large-scale fates of the Universe in bulk: It can expand forever. It can expand and then contract to a virtual point and end. Or it can cyclically expand and contract indefinitely. Each model represents a hypothesis—a theory based on available data and awaiting further tests. But unless we take that final step in the scientific method and put the models to the experimental test, we cannot know which one, if any, is correct.

We also welcome more information about the nature of the primeval event that triggered the expanding pattern in the first place. What was the original, primordial state that gave rise to the energy that would later help form galaxies, stars, planets, and life? Can we really expect to probe all the way back in time? After more than ten thousand years of civilization, indeed after many cultures had earlier invented their own worldviews based on beliefs and thoughts, modern science now seems ready to provide some data-driven insight into the origin of all things.

As tricky a task as this may seem, several cosmological models are now being subjected to observational tests by today's astrophysicists. We live at a remarkable time when truly fundamental issues can be addressed, if not yet solved, by observational means. Our experiments, together with the theories underlying them, seek direct answers to many of the above questions. Even a superficial understanding of the current status of the solutions, though, requires a deep appreciation for the nature of space and time on the grandest scale. And to gain this appreciation, we need a tool of deep and powerful insight—Einstein's theory of relativity.

∞

Some people become hot, bothered, and tense upon hearing the word "relativity." This subject is surrounded by a mystique implying that only geniuses can understand it—and that might well be true at the mathematical level. But, conceptually, relativity theory is relatively simple. Its foundations are clear and explicit, provided we are willing to forgo common sense and human intuition. Indeed, that's the key: to put aside our everyday, Newtonian (even Aristotelian) ways of reasoning and adopt a broader, innovative stance that allows for unorthodox thinking.

Relativity is simple in its symmetry, its beauty, its elegant ways of describing grandiose aspects of the Universe. Sure, it employs higher mathematics—advanced calculus and beyond—to quantify its application to the real Universe, yet everyone should strive to gain at least a nonmathematical feeling for some of the underlying concepts of relativity theory. In this way, we shall be better positioned to appreciate, albeit only qualitatively, some of the weird physical effects encountered while modeling the Universe, exploring the bizarre black holes, and even contemplating the origin of all things.

Relativity theory has two principal tenets, both enunciated in 1905 by the German-Swiss-American physicist Albert Einstein. Together they lead to the famous $E = mc^2$ equation, where E, m, and c are symbols representing energy, mass, and the speed of light, respectively. The first tenet is straightforward: Nature's laws are the same everywhere and for all observers. Regardless of where a person is, or how fast a person may be moving, the basic physical laws are invariant.

The second tenet of relativity is a little more subtle: there is a fourth dimension—time—which in every way is equivalent to the usual three spatial dimensions. In other words, by using the three well-known dimensions of space, an object's position can be generally described as either right or left, either up or down, and either in or out. Three dimensions are sufficient to describe *where* any object is in *space.* A fourth dimension of time is necessary to describe *when*—either past or future—an object exists in that space. By coupling time together with the three dimensions of space, Einstein was able to reconcile previous inconsistencies in Isaac Newton's post-Renaissance view of our world by arguing that the velocity of light is an absolute constant number at all times and to all observers, regardless of when, where, or how radiation is measured. Space and time are in fact so thoroughly intertwined within Einstein's view of the Universe that he urged us to regard these two quantities not as space *and* time but as one—*spacetime.*

Many important consequences of relativity theory can be qualitatively explained only by analogy. Here is one of them: Suppose we are in an elevator that has no windows. As it rises, we feel the floor pushing, especially on our feet. It's easy to attribute this pushing sensation to the upward acceleration of the elevator. Now, imagine such a windowless elevator in outer space far from Earth. Normally, we would experience the weightlessness made familiar by watching astronauts floating around where there are no net forces. But if we *did* experience a sensation of pushing on our bodies, we could draw one of two conclusions: Perhaps the elevator is accelerating upward in the absence of gravity, thus pinning us to the floor. Or maybe the elevator is at rest in the presence of gravity, which is pulling us from below. There is no way to tell which of these explanations is correct without performing an experiment—that is, without observing objects outside the hypothetical elevator. In either case, pendulum clocks swing normally, released stones fall just as Galileo taught us, water pours from a glass in customary fashion, and so on. If we did build a window to look out, we would have no trouble establishing whether the elevator is really at rest or really accelerating. Relative to the Universe outside the elevator, it's easy to assess the real status of that elevator.

The important point is that the effect of gravity on an object and the effect of acceleration on that object are indistinguishable. Physicists call this keystone of relativity theory the "Principle of Equivalence": The pull of gravity and the acceleration of objects through spacetime can be viewed as conceptually and (almost) mathematically equivalent. Consequently, Einstein postulated as unnecessary the Newtonian view of gravity as a force that pulls. Not only is that view obsolete, but Newton's theory is today known to be less accurate than Einstein's.

Let's briefly examine how the notion of an accelerated object can replace the commonsense idea of gravity. The upshot is this: Einstein's theory of relativity allows us to inquire how it is that matter, which conventionally gives rise to Newton's theory of gravity, alters the nature of spacetime. Bypassing the details, matter effectively shapes the geometry of spacetime. Put another way, mass is said to "curve" or "warp" spacetime.

Ordinary Euclidean geometry—the type learned in high school—holds valid when the extent of curvature is zero, that is, when spacetime is flat. Even when that curvature is slight, Euclidean geometry of flat space is approximately correct. At any one location on Earth's surface, for instance, an architect can design a building, or a contractor build one, using the procedures laid down twenty-five centuries ago by the Greek mathematician Euclid. However, although terrestrially familiar flat-space geometry is used regularly in our daily tasks, it's not absolutely correct. Earth, after all, is not flat; it's curved. On the surface of a sphere, flat Euclidean geometry works satisfactorily at any small locality, but that's because it's nearly impossible to perceive our planet's curvature from any single place on its surface. Once the curvature of Earth becomes discernable, as in the case of intercontinental aircraft or shipboard navigation, for example, a more sophisticated geometry must be used—a curved-space geometry.

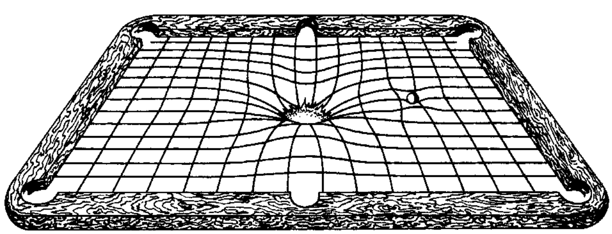
Thus, in the absence of matter, the curvature of spacetime is zero, the appropriate flexure is flat, and objects move undeflected in straight lines. Newtonian dynamics and Euclidean geometry are fine, for all practical purposes, wherever spacetime is unappreciably curved. To be sure, flat space isn't entirely hypothetical, since beyond the reaches of galaxies very little matter presumably exists. As noted later in this prologue, the Universe itself, on average and in sum, may well be flat.

On the other hand, the geometry of spacetime is strongly warped near massive objects. It's not the object or the surface of the object that is warped, just the near-void of spacetime in which the object exists. The larger the amount of matter at any given location, the larger the extent of curvature or the warp of spacetime there. Furthermore, far from a massive object, the warp lessens. As with gravity, the extent of curvature depends upon both the amount of matter and the distance from that matter. But, since this newer notion of warped spacetime is more accurate than the older, traditional idea of gravity, the universal worldview of Newton must be replaced by that of Einstein.

No one ever said that relativity wasn't strange. How can a curve replace a force? The answer is that the topography of spacetime influences celestial travelers in their choice of routes, much as Newton imagined gravity to hold an object in its path. Just as a pinball cannot traverse a straight path once shot along the inside of a bowl, so the shape of space causes objects to follow curved paths (called geodesics). Any object whose motion changes direction, even though its speed remains steady, is said to be accelerated. Earth, for example, accelerates while orbiting the Sun—not because of gravity, as Newton maintained, but because of the curvature of spacetime, as Einstein preferred.

To see this, consider another analogy—not an example, an analogy. Imagine a pool table with a playing surface made of a thin rubber sheet, rather than the usual felt-covered slate. Such a rubber sheet would become distorted if a large weight were placed on it. A heavy rock, for instance, would cause the sheet to sag or warp. The otherwise flat rubber sheet would become curved, especially near the rock. The heavier the rock, the greater the curvature. Trying to play billiards, we would quickly find that balls passing near the rock are deflected by the curvature of the tabletop.

In much the same way, both matter and radiation are deflected by the curvature of spacetime near massive objects. For example, Earth is deflected from a straight-line path by the slight spacetime curvature created by our Sun. The extent of the deflection is large enough to cause our planet to circle the Sun repeatedly. Likewise, the Moon or a baseball responds to the spacetime curvature created by Earth and they, too, move along a curved path. The deflection of the distant Moon is slight, causing it to orbit Earth endlessly. The deflection of a small baseball is much larger, causing it to return to Earth's surface.



. . . the geometry of spacetime is strongly warped near massive objects.

The commonsense notion of gravity, then, is just a convenient word for the natural behavior of objects responding to the curvature of spacetime. Accordingly, we can use a knowledge of spacetime to predict the motions of objects traveling through space and time. More appropriately, we can turn the problem around: by studying the accelerated motions of objects, we can learn something about the geometry of spacetime near those objects.

And so it is with the whole Universe. When seeking the size, shape, and structure of the entire Universe—the biggest picture of all—we need to consider, in principle, the net effect of spacetime curvature caused by each and every massive object in the cosmos. By studying the motions of representative pieces of matter within the Universe, we can discover much about the curvature of the whole Universe. In practice, it's a lot more difficult.

By infusing relativity's basic tenets into a full-blown, mathematical treatment of Einstein's theory, researchers have learned to map various ways that matter warps spacetime. This is the area where relativity theory becomes notoriously complex; here, theorists scamper away, leaving us in an imponderable dust. Our gleanings from their labored calculations can only be appreciative. The results, in a nutshell, are the so-called Einstein field equations—a dozen or so equations that must be solved simultaneously to determine how the Universe is grandly structured, namely, how spacetime is curved by all the matter present. On the one hand, these equations are nearly intractable to solve quantitatively, yet on the other hand, they contain remarkable symmetry qualitatively. Much like works of art, they often inspire a sense of wonder, a certain awe. Their complexity arises largely because, in addition to the field equations specifying the shape of the Universe, astrophysicists using relativity must also solve several geodesic (geometrical) equations to determine how it is that any individual object behaves dynamically at any given place among all the other matter in the Universe. The bottom line of much technicality is this: matter determines how space is curved, and space determines how matter moves.

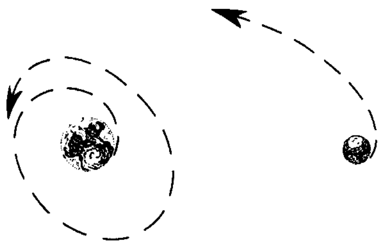
To illustrate further the curvature of spacetime, ponder the following hypothetical example. Imagine two planets, each inhabited by equally advanced technological civilizations capable of launching identical rockets. Earth can be one and the less massive planet Mars the other. For the sake of discussion, let's assume that these rockets can achieve only a fixed amount of thrust at launch, after which they glide freely through space. When the rockets are launched from both planets, the shapes of their paths differ. In the Newtonian view of space, the rocket paths are determined by the gravitational interaction between the rocket and each planet. In the Einsteinian view of spacetime, these trajectories are determined by the response of the rocket to the spacetime warp produced by each planet.

Consider first a typical path of the rocket launched from the more massive Earth. The initial kick is chosen in this case to be large enough to place the rocket into an elliptical orbit. Like gravity, whose strength decreases with increasing distance from a massive object, the curvature of spacetime is also greater close to the massive planet. The rocket accordingly speeds up (or accelerates) when close by and slows down (or decelerates) when far away. General relativity thus agrees with the laws of planetary motion empirically discovered a few centuries ago by the German astronomer Johannes Kepler. Relativity maintains that the rocket accelerates near massive objects, owing to the greater degree of spacetime curvature there.

The ellipse, a "closed" geometric path, is only one possible type of motion. It is a trajectory of minimum energy, so labeled because a rocket in such an orbit doesn't have enough energy to escape the planet's influence. It keeps orbiting endlessly like an artificial satellite.

Rockets can have other paths as well. Consider the trajectory taken by an identical rocket launched from the less-massive-planet Mars. The same thrust used to launch the Earth rocket into elliptical orbit is now great enough to propel the rocket entirely away from Mars. Less energy is used in the launch from Mars than in the one from Earth, and thus more energy can be imparted to the motion of the rocket. The rocket escapes the influence of Mars because, as a Newtonian classicist would say, Mars has less gravitational pull than Earth. By contrast, Einsteinian relativists claim that such a rocket escapes Mars because the less-massive Mars warps spacetime less than does Earth. The two views—Newtonian and Einsteinian—predict virtually identical paths for the rocket as it recedes toward regions of spacetime progressively less curved by Mars.

The resultant path away from Mars is called a hyperbolic trajectory. This is the type of flight path taken by robot spacecraft that have been exploring other planets of our Solar System in recent years. Its geometry is said to be "open," in contrast to the closed, elliptical path around Earth. Any object traveling along such a hyperbolic path has more energy than one on an elliptical trek, either because the initial kick needed to achieve a hyperbolic trajectory was large or because the mass of the parent object from which the launch was made is small. In this particular example, the rockets are identical, so the increased energy of the hyperbolic case results from the relatively small mass of Mars.



. . . trajectories are determined by the response of the rocket to the spacetime warp of each planet.

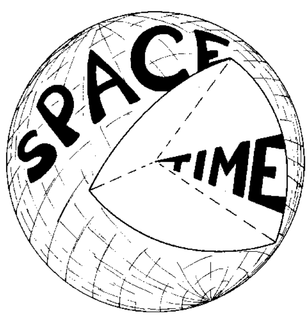
Even while receding far from its parent planet, a rocket is still affected by the pull of gravity or the warp of spacetime created by the mass of that planet. Although large only in the immediate vicinity of the planet itself, Mars' influence over the rocket never diminishes to zero. Mathematical analyses predict that, in the idealized absence of all other astronomical objects, such a hyperbolically launched rocket should approach infinity—that is, withdraw from Mars indefinitely.

The hyperbolic path contrasts slightly with another type of trajectory conceivably taken by an escaping rocket. A third geometrical path, also open in form and called a parabola, is one taken by a rocket from some hypothetical planet having a mass between that of Earth and Mars. The parabolic path closely mimics the hyperbolic one in that they both approach infinity, though they differ a little in energy content. Mathematicians distinguish between the two open paths by saying that a parabolically moving rocket will have a velocity of zero when it gets to infinity—and will then stop!—whereas its hyperbolic counterpart will theoretically reach infinity with some finite velocity—and move beyond! The academic language of mathematics notwithstanding, we realize that in actuality no object can ever really reach infinity, thus this is tantamount to saying that the rocket will continue to recede forevermore.

The above cases conveniently describe the motion of any object in terms of its energy content and its response to the curvature of spacetime. Actually, the intermediate case of the parabolic path is a very special, precisely balanced one for which the net energy is zero and the overall geometry of space is flat, and Euclid would have loved it. These cases will be useful analogies when later considering the essentials of cosmology, for then the "object" will be the entire Universe itself.

∞

Einstein, as the originator of relativity, clearly had an advantage in initially using his equations to deduce the nature and structure of the Universe; he knew them better than anyone else. His equations predicted in 1917 that the curvature of the entire Universe must indeed be large owing to all the matter contained within it. The flat geometry of Euclid just didn't seem to work when examining the bulk properties of the whole Universe. Unfortunately, Einstein's most popular solution—one of many possible at the time—can be cast only in terms of nearly unimaginable four-dimensional spacetime. It's quite imaginable mathematically, but it's tricky verbally. Even if we suspect now, nearly a century later, that the Universe in toto is not very curved (and in fact may be flat, on average), what follows is useful conceptually.



. . . a three-dimensional analogue of Einstein's four-dimensional Universe.

To visualize the essence of this solution, we employ another analogy. Since no one has ever built a viewable model of anything in four dimensions, in this analogy we suppress one of those four dimensions. For sake of argument, imagine consolidating the three dimensions of space into only two dimensions. Then, with time as the remaining dimension, we can construct a three-dimensional analogue of Einstein's four-dimensional Universe. That analogue is a sphere, sometimes colloquially termed "Einstein's curveball." Here, all of space is taken to be spread *on the surface* of this sphere. The other dimension—time—is represented by the radius, or depth, of the sphere.

To counter an oft-misunderstood aspect of this analogy, note that the Universe and all its contents are *not* envisioned to be scattered inside the sphere. Rather, they are distributed *just on its surface.* All three dimensions of space are warped—in this special case, into a perfect sphere. Thus, all the galaxies, stars, planets, and people, and even all the radiation reside only on the surface of the sphere of this model Universe.

Note also that since the radius of this model sphere represents time, this spherical analogue grows with time. After all, the galaxies are observed to be receding; the Universe is expanding. As time marches on, the radius of the sphere increases and so does its surface area. In this way, our three-dimensional analogue mimics cosmic expansion.

Actually, Einstein didn't know in 1917 that the Universe is expanding. Astronomers, notably Edwin Hubble and his American colleagues, didn't establish that until the 1930s. Einstein's own equations had allowed cosmic expansion (or contraction), but he didn't believe it. He was probably fooled by the then still-popular Aristotelian philosophy that few things change (and nothing at all beyond the Moon). So he tinkered with his equations, introduced an additional factor that offset the predicted expansion, and thereby forced his Universe models to remain static. Einstein later came to think that he was wrong in doing this, calling this "cosmological constant" the biggest mistake of his career. But now, in the early twenty-first century, this poorly understood factor has again become fashionable, suggesting that Einstein may well have been onto something truly fundamental, yet truly odd, that no one has yet deciphered. We shall return to discuss the implications of the cosmological constant later in this prologue.

Even if not a good model for the Universe, this spherical analogue enabled Einstein and colleague relativists to uncover many telling features of curved spacetime. One of their most important findings is known as the cosmological principle—the notion that all observers perceive the Universe in roughly the same way regardless of their actual locations. To be sure, all our large-scale studies to date strongly imply that the Universe is homogeneous (the same everywhere) and isotropic (the same in all directions). Excluding directions obscured by our Milky Way and considering realms beyond a billion light-years away (closer than which cosmic structure *is* seen), the contents of the Universe look virtually identical. On the grandest scales of all, then, the Universe seems smooth, and even a bit boring.

To grasp the essence of the cosmological principle, consider a sphere again. It can be any sphere, so let it be Earth. Imagine ourselves at some desolate location on Earth's surface, perhaps in the midst of the Pacific Ocean. To validate this analogy, we must confine ourselves to two dimensions of space; we can look east or west and north or south but not up or down—the life of a fictional "flatlander." Perceiving our surroundings, we note a very definite horizon everywhere. The surface *appears* flat and pretty much identical in all directions. Accordingly, we might get the impression of being at the center of something. But we're not really at the center of Earth's surface at all. The surface of a sphere has no center. Such is the cosmological principle: there is no preferred, special, or central location on the surface of any sphere.

Likewise, regardless of our position in the real, four-dimensional Universe, we observe roughly the same spread of galaxies as would be noted by any other observer from any other vantage point in the Universe. Despite our observation that galaxies literally surround us in the sky, this need not mean that we reside at the center of the Universe. In fact, if our spherical analogy is valid, then the Universe has no center. Nor does it have any edge or boundary. The case of a flatlander roaming on the surface of a three-dimensional sphere is completely analogous to a space traveler voyaging through the real four-dimensional Universe. Neither ever reaches a boundary or an edge. Proceeding far enough in a single direction on the surface of the sphere, the traveler (or any radiation) would eventually return to the starting point, just as Magellan's crew proved long ago by circumnavigating planet Earth. In much the same way, if four-dimensional spacetime is structured according to this spherical analogue, an astronaut could be launched in one direction, only to return at some future date from the opposite direction. Einstein's curveball, indeed.

Today, we realize that the Universe is not static. The recessional motions of the galaxies make its expansion indisputable. Following the lead of the Russian meteorologist Alexander Friedmann and the Belgian priest Georges Lemaitre from the 1920s, modern relativists seek more realistic models of the Universe, especially ones that take account of the measured rate of cosmic expansion. In this way, observations of galaxy recession become a boundary condition, or demanding constraint, on any plausible model of the Universe, helping refine our twenty-first-century view of the big picture.

The cosmological principle is valid even though the Universe is expanding. Like that of any static sphere, the surface of an expanding sphere has no center, edge, or boundary. To see this, imagine a sphere again, though now one that can swell like a balloon. For example, visualize the entire Earth to be expanding, causing the surface area of our planet to increase as time advances. Standing on such a hypothetically expanding "Earth," we would see familiar objects moving away. Surface objects all around—whether trees, homes, or mountains—would appear to recede. Now, more than ever, we may want to conclude that our position is special—that we exist at the center of some explosion. But we do not. Our position is no more special than anyone else's on the sphere's surface. In fact, everyone everywhere on an expanding surface would observe their surroundings to be receding. Who is correct, then? Everyone is correct. Recessional motions are observed from *any and all* positions on the surface of an expanding sphere.

Another popular way of visualizing the same concept is to tape small coins onto the surface of a balloon. The coins are meant to represent the galaxies, and the balloon the "fabric" of space itself. As the balloon inflates, space expands and all the coins recede from one another (though the coins themselves do not expand). Regardless of which galaxy we inhabit, we would see all the other galaxies receding (though the galaxies themselves, held together by gravity, are also not expanding). The galaxies would appear to recede for any and all observers in the Universe. Nothing is special or peculiar about the fact that all the galaxies are receding from us. Such, again, is the cosmological principle: no observer anywhere in the Universe has a privileged position.

And so it is in the real, four-dimensional Universe. Although the galaxies recede from us, this is not a peculiarity of our vantage point. All observers everywhere in the Universe witness essentially the same sort of galaxy recession. Neither we nor anyone else reside at the center of the expanding Universe. There is no center in *space*—no position on the sky that we can ever hope to identify as the location from which the cosmic expansion began.

Do note that all these analogies have their shortcomings and this one is no different. The issue here is that we must imagine the balloon, whose surface is a two-dimensional analogue of space, expanding into a third dimension. That might suggest that, in the real world of three spatial dimensions, the Universe is expanding into some additional spatial realm—which, as noted earlier, is wrong. In our analogy, that balloon is properly visualized as expanding into time—namely, into the future. As best we can tell, even if higher dimensions of space do exist, they are irrelevant to the macroscopic models of the Universe discussed here.

Surprisingly, there *is* a center in *time*—at least in our analogy. This is the origin of time, and it corresponds in our three-dimensional spherical analogue to a sphere having zero radius. In other words, at the beginning of the Universe, the three-dimensional sphere was a point. This marked the beginning of time, the moment of the big bang. It's proper to think of it as the edge of time. But there's no edge in space.

Basic and profound queries come fluxing forward: When did the sphere have zero radius—a mere point? That is, how long ago were all the contents of the Universe squashed into a single speck? Fundamentally put, when did time begin?

To appreciate answers to these questions, imagine that time can be reversed. Not that we have any evidence that time actually does reverse, or flow backward; rather, this is another mental exercise to visualize when all the galaxies in space (or all the coins on our analogous balloon) were effectively piled one upon another. To do this, we imagine reversing the expansion of the Universe by contracting it backward at the same rate as we currently observe it expanding forward. The galaxies would come together, eventually touch, and finally mix. If we can estimate how long it would take for the whole Universe to shrink back to its starting point, we shall then have a measure of the time it did take to reach its present state—the age of the Universe.

The answer, as best we can determine, is about fourteen billion years. Thus, the singular, compact region of space often associated with the origin of the Universe must have existed about fourteen billion years ago. Alternatively stated, fourteen billion years have passed since the expanding debris of universal matter raced out to the places where they are now observed.

∞

The issue of absolute ages of cosmic systems as well as their relative timescales is an important one in cosmic evolution. For the cosmic-evolutionary narrative to hold, all those ages must be consistent, each arranged sequentially along the arrow of time. Here, we intentionally become a bit more technical, mainly to give those readers who wish it a slightly deeper treatment of the central topic of time.

The age of the Universe has been a particularly vexing quandary for decades. Teams of researchers joust in heated argument, and not with just a little acrimony. Clear-cut biases are evident and reputations of some astronomers are on the line. The media, too, has caught on, viewing this issue as another kind of "Hubble wars" while claiming, often wrongly, all sorts of dire consequences for big bang cosmology. Yet longstanding problems of age among principal systems—cosmos, stars, life—have plagued science off and on for well more than a century.

In two paragraphs, here is a statement of today's concern: The simplest analysis of a uniformly expanding Universe implies an age of some fifteen billion years. This is based on a so-called Hubble constant of twenty kilometers per second per million light-years, our best current value specifying the rate at which the Universe expands. (In units more commonly used by astronomers, this number is nearly seventy kilometers per second per million parsecs; there are 3.26 light-years in a parsec, a dreadful unit that does nothing but help keep the beginners out.) That is, for each additional million light-years of distance, the galaxies seem to recede with an added twenty kilometers per second, in accord with the established finding that distances and velocities of galaxies are well correlated. However, this age is correct only if the cosmic density is much lower than the "critical density" of a marginally bound Universe whose space balloons to infinity—namely, a Universe containing little or no matter. (Mathematicians say that such a Universe has a "trajectory" that will then stop at infinity, but since nothing can actually reach infinity, this is tantamount to the Universe expanding forever, much like the analogy of a rocket escaping its parent planet.) If the Universe does have matter (and of course it does) and if its density does equal the critical value (which many astronomers favor), then the Universe might ordinarily be expected to decelerate with time, owing to the mutual gravitational attraction of matter everywhere, making the true age less than fifteen billion years. This is one of the more famous Einstein solutions to his field equations, for which the Universe's age, using today's value of Hubble's constant, would be more like ten billion years.

By contrast, key parts of the Universe—namely, some of its stars—seem older than ten billion years. These are the ancient stars of the globular clusters (which we shall meet later in the Stellar Epoch), tight-knit groups of typically hundreds of thousands of stars that are strewn throughout the halos of galaxies and are probably as old as the galaxies themselves (a topic soon discussed in the Galactic Epoch). Astronomers estimate such stellar ages based on the rates at which stars undergo nuclear fusion—in particular, according to the theory of stellar evolution that specifies when mature, normal stars begin changing into swollen, red-giant stars. Many globular clusters were examined for this color change during the past few decades and most of them imply ages in the range of twelve to sixteen billion years. Hence, the paradox at hand: at face value, some stars seem older than the Universe itself—a possible inconsistency of timescales and a clear embarrassment to astronomy if not resolved.

Actually, this problem is not really a new one. Debate has swirled around it in one form or another for well more than a century. For example, in the mid-nineteenth century, when the pioneers of geochronology sought to assess the age of Earth on grounds other than religion or philosophy, they essentially made two assumptions: Earth probably formed at the same time as the Sun, and the Sun shone by the burning of some known chemical, like the wood or coal commonly used during the Industrial Revolution. The answer they got for the age of the Sun, and hence the age of Earth, was a few thousand years, a value less than that of recorded history. So an age controversy developed, not so much heated as merely amusing to most theologians of the time who thought poorly of science: How could Earth be younger than the duration of human existence?

The first assumption of early Victorian science was a good one—we do now judge the births of Earth and the Sun to have been contemporaneous, as noted later in the Planetary Epoch. But the second one was most definitely not—the Sun is assuredly not made of wood or coal! Physicists such as Lord Kelvin of Britain and Hermann von Helmholtz of Germany later revised these calculations in the late nineteenth century, taking the Sun to be made of an incandescent liquid mass (such as gasoline or kerosene) and allowing for some energy generation via gravitational infall (including meteors crashing into the Sun). Yet they were unable to increase the age estimate for the Sun to much more than a hundred million years—a value surely older than recorded history but still much less than that then needed by the British naturalist Charles Darwin to explain the fossil record in terms of biological evolution by natural selection. Long-dead life-forms seemed at the time to be at least several hundred million years old, and we now realize they are even older, as noted in the Biological Epoch. Kelvin got similarly low values for Earth's age when trying to estimate the rate at which our planet cooled, largely because he overlooked the poor thermal conductivity of the rocky interior—all of which put geological evolution into conflict with biological evolution. Thus, the age controversy continued, dominating scientific circles about a hundred years ago, some of the debate (then as now) being quite vehement: How could life on Earth be older than the planet itself?

These early age discrepancies eventually went away. As radioactivity became better understood, mainly by the French scientists Henri Becquerel and Pierre and Marie Curie around the turn of the twentieth century, geologists could then measure the age of Earth directly. And what they found was a planet of a few billion years, fully enough to provide the long timescales needed to explain Darwin's fossils. Scientists now know that biological evolution has occurred for more than three billion years, yet there is no problem here since modern radioactive methods currently date Earth at nearly five billion years.

Alas, in the 1930s, a version of this "age-old" problem resurfaced. At issue were some of the first measurements of the Hubble constant by Edwin Hubble himself and some of his colleagues. Owing to observational uncertainties in the brightness of the galaxies and especially to calibration errors in the analysis of the acquired data, they found a Hubble constant of more than a hundred kilometers per second per million light-years. This then implied that the Universe expands much faster than we now know it does, meaning that the galaxies would have gotten out to where they are now observed much quicker. In fact, Hubble's original analysis implied a Universe age of less than a few billion years, and suddenly the general problem was back: How could Earth be older than the Universe?

In turn, this problem gradually faded away as many astronomers undertook, over the course of several decades in the mid-twentieth century, better observations and data analyses of the brightness and distances of the galaxies. By the 1950s, the value of the Hubble constant had decreased five-fold, and the Universe age consequently lengthened to at least ten billion years. Hence, the Universe was then safely older than Earth and the age problem went away again . . . for a while. To be sure, it has returned in more recent years given the claimed ages of some globular clusters. By the 1980s and into the 1990s, we had the modern version of a recurring age discrepancy, to wit: How could some stars within the Milky Way be older than the Universe itself? Well, they can't be. It's as simple as that. Something is awry—again.

Fortunately, this lingering age controversy seems to be fading away again, just as better observations and improved models caused similar glaring contradictions to evaporate throughout the past century. Indeed, several recent developments favor the dissolution of this problem altogether. For example, today's astronomers are converging on a model Universe that is decidedly "open"—again, like the escaping rocket whose open-ended geometric path extends forever. In particular, our best value of the Hubble constant currently seems to be pointing at a somewhat older cosmic age of about fourteen billion years. What's more, recent reanalyses of the globular star clusters, especially based on data acquired by Europe's *Hipparcos* satellite, imply that the globulars have had their ages previously overestimated by nearly twenty percent. If confirmed, then the average age of the oldest stars needs to be readjusted downward to ten to twelve billion years, making them safely younger than the Universe.

We need not be overly concerned about this periodic age controversy, other than to note it as an active area of research that seeks to specify a number (the value of Hubble's constant) to an accuracy of ten percent, when many other cosmologically significant numbers (such as the cosmic density) are known only to within a factor of about ten. Maybe the universal age will eventually turn out to be twelve, or thirteen, or fifteen billion years; all arguments presented in this book aim toward a best current estimate of fourteen billion years. The specific number is not that important and will not likely be pinned down for many years, if ever. What is most remarkable is that the ages of the cosmos, stars, and life now stack up so well chronologically along the arrow of time—and are indeed consistent with increasing order and rising complexity over the course of all natural history.

As for the cosmic-evolutionary scenario presented here, our narrative is hardly affected by this on-again, off-again age controversy—even if it reemerges. The arrow of time itself can be contracted or expanded, a little like an accordion, in order to match whatever is the true age of the Universe. The historical sequence of events along the arrow of time is more important than the magnitude of the arrow itself.

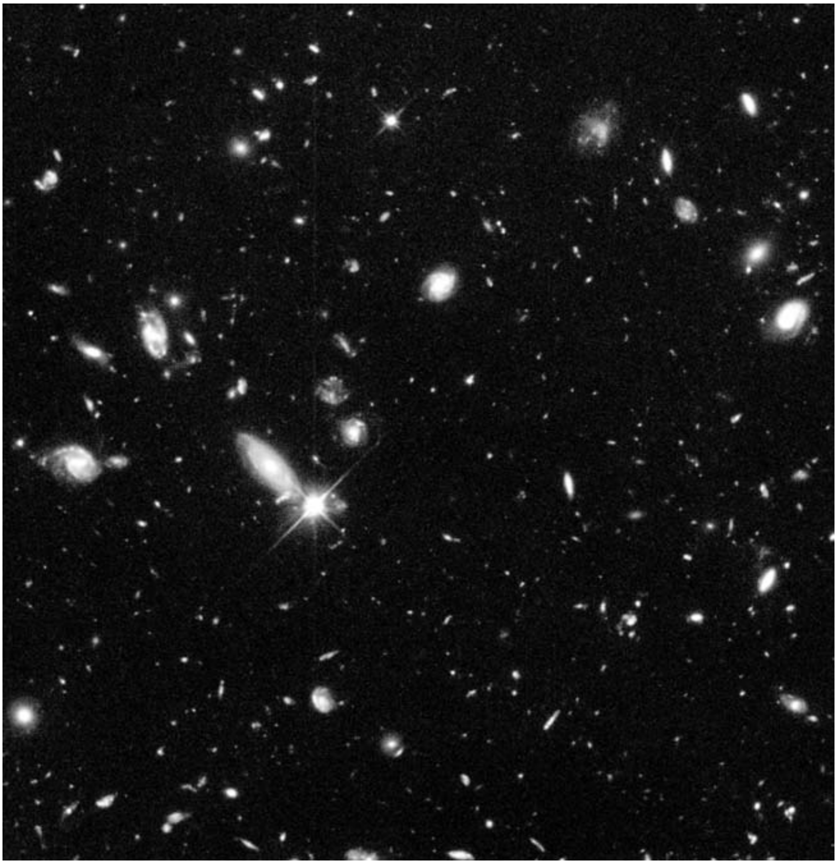
∞

At the origin of time, the Universe burst forth. Like an inflated balloon, it flowed out into the future—the Universe expands and the galaxies recede. Initially, it all changed at a rate dependent on the density of matter contained within it. After all, each clump of matter in the Universe gravitationally pulls on all the other clumps. Since the gravitational force is always attractive, it tends to counteract the expansion. So a tightly packed Universe is expected to cause a strong gravitational pull and eventually a slowing of the universal expansion. (Notice that we've returned to the notion of gravity; though warped space is more correct, the familiar concept of gravity often makes the argument easier to comprehend.)

At face value, universal expansion is not unlike what happened with the rockets noted earlier. Each rocket departed from its parent planet at a rate dependent on that planet's mass. Mars, for example, pulled on the launched rocket, but was unable to slow the rocket's escape; the more massive Earth exerted an even stronger pull on the rocket and was able to halt its escape. The parallel between the orbital dynamics of a rocket and the cosmic dynamics of the Universe is quite a good one. As for rockets, there are two diametrically opposed models of a dynamic, changing Universe—and one perfectly balanced between the two extremes.

The first model Universe is one that evolves from a powerful initial "explosion"—again, a bang of some sort at the origin of time. The Universe then expanded from what must have been an exceedingly dense primeval clump. As time progressed, space diluted the matter throughout the Universe, causing its average density to decline. In this first model, insufficient matter exists to counteract the expansion. Accordingly, the Universe simply expands forever, with the density of matter thinning eventually to nearly zero. It's specifically analogous to the rocket moving away from Mars; this type of Universe has too little mass ever to halt the matter's outward motion. Since this model Universe will theoretically arrive at infinity with some finite (nonzero) velocity, some astronomers term this case the hyperbolic model of the Universe, for that is the trajectory such a Universe takes while racing toward infinity.

A hyperbolic model is said to imply an "open Universe." It's open in the sense that the initial bang was large enough and the contained matter spread thinly enough to ensure that this type of Universe will never stop expanding. Although matter everywhere mutually pulls on all other parts of the Universe, such a Universe will never collapse back on itself. There's simply not enough matter.



Evidence of universal expansion.

Thousands of galaxies, each housing hundreds of billions of stars, can be seen in this optical image. Covering only one percent of the area subtended by the full Moon, this image resembles a "core sample" of celestial objects—extending from relatively near Earth to the distant Universe billions of light-years beyond. Doppler measurements indicate that all these galaxies are receding, all of them partaking of a grand expansion that began more than ten billion years ago. *Source: Space Telescope Science Institute.*

Of course, the Universe can never really become infinitely large. An infinite amount of time would be needed to reach infinity. This is just the mathematician's way of saying that a hyperbolic, or open, Universe will expand endlessly. Properly stated, an open Universe *approaches* infinity.

Should this model be correct, the galaxies will recede forevermore. With time, for an observer on Earth, they will fade away toward invisibility, their radiation weakening with increasing distance. Eventually, even some of the closest galaxies will be so remote as to be hardly visible. Someday, all the galaxies might become unobservable; they will be too distant, their radiation too faint. Our home Milky Way Galaxy will then be the only object within the observable Universe. All else, even through the most powerful telescopes, will be dark and quiet. And even beyond that in time, the Milky Way too will someday peter out as its fuel supply is consumed, the hydrogen in all its stars totally spent. This type of Universe and all its contents eventually experience a "cold death." The radiation, matter, and life in such a Universe are destined to freeze.

Quite a different fate awaits the Universe if it has a larger matter density. As for the open Universe, this model also expands with time from a superdense original point. But unlike the open Universe, this model contains enough matter to halt the cosmic expansion before reaching infinity. That is, after the bang initially pushed out the Universe, the galaxies lost so much momentum that they will eventually skid to a stop sometime in the future. Astronomers everywhere—on any planet within any galaxy—would then announce that the galaxy recession has ended as their radiation is no longer red shifted. The cosmological principle guarantees that this new view will prevail everywhere. The bulk motion of the Universe, and of all the galaxies within, will be stilled—at least momentarily.

Cosmic expansion may well stop, but gravitational pull does not. Gravity is relentless. Accordingly, this type of Universe will necessarily contract. It cannot stay motionless; nothing fails to change. Astronomers will witness the galaxies' red shift gradually change to a blue shift. The contraction of this model Universe is a mirror image of its expansion. Not an instantaneous collapse, it's rather a steady movement toward an ultimate end, requiring just as much time to fall back as it took to rise up.

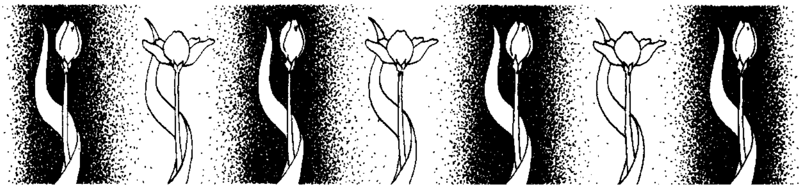
This model in many ways resembles the rocket trajectory for which, in our earlier example, the gravitational pull was great enough to cause the rocket's path to become elliptical. Since it has a similar geometrical pattern, a cosmic model containing enough matter to reverse the expansion is often called an elliptical Universe. It's also sometimes termed a "closed Universe"—closed because it represents a Universe finite in size and in time. It has a beginning and it has an end.

The change of density in a closed Universe is interesting—and ominous. From what must have been an enormously high initial value, the density thins dramatically by the time the Universe stops expanding, then returns again to a huge value when, at some future epoch, all matter collapses onto itself. Some astronomers call it the "big crunch."

The expansion-contraction scenario of a closed Universe has many fascinating (and dire) implications. Life, in particular, which has evolved from simplicity to complexity during the expansion, will begin breaking down into simplicity again while inevitably heading toward its demise during collapse. Toward the end of the contraction phase, the galaxies will collide frequently as the total amount of space in which they exist diminishes—and that means trouble for any life-forms. For just as compressing air in a bicycle pump or rubbing our hands causes heating via friction, collisions among galaxies will generate heat as well. The entire Universe will grow progressively denser and hotter as the contraction approaches the end. Near total collapse, the temperature of the entire Universe will have become greater than that of a typical star. Everything everywhere will have become bright—so bright that stars themselves will cease to shine for want of contrasting darkness. This type of Universe will then shrink toward the superdense, superhot state of matter similar, if not identical, to the one from which it originated. In contrast to the open Universe that terminates as a frozen cinder, this closed Universe will experience a "hot death." Its contents are destined to fry.

Cosmologists are uncertain of the fate of a closed Universe upon reaching this (perhaps infinitely) hot, dense, and small state, known among scientists as a "singularity." The Universe might just end. Or it might bounce—into another cycle of expansion and contraction. Frankly, the mathematics of singularities have not yet been fathomed; physical laws there are suspect. This ultimate state of matter poses one of the hardest problems in all of science. Though they don't like to hear it said out loud, astrophysicists are experimentally and theoretically ignorant of the physics of singularities.

Frontier research seeks to understand better the nature of such a singular state of matter, a topic to which we shall return when examining the universal origin more closely in the Particle Epoch, and also when exploring black holes in the Galactic Epoch. For now, suffice it to note that with both density and temperature increasing as the contraction nears completion, pressure—the product of density and temperature, at least in some classical sense—must increase phenomenally. The question as yet unanswered is, Will the Universe just end as a final miniscule speck, or will this pressure be sufficient to overwhelm the relentless pull of gravity, thereby pushing the Universe back out into another cycle of expansion and contraction? In other words, will a closed Universe bounce?



. . . a cyclic Universe oscillates forever, each expansion a "day," each contraction a "night."

A certain aesthetic beauty pervades such a model of a "cyclic Universe." Subjectively, in our guts, many researchers prefer it. For starters, there's no need for a unique, once-and-for-all-time initial event—no need for a *big* bang. Nor does this model embody a definite beginning or a definite end. The cyclic model merely goes through phases—perhaps an infinite number of them—each initiated by a separate "bang," each ending in another "bang," ad infinitum. Indeed, such a cyclic Universe would presumably oscillate forever, each expansion a "day," each contraction a "night." But none of these bangs is unique, none of the origins any more significant than any other. Oscillation avoids the potential philosophical problem of what preceded a unique big bang of either a one-cycle closed Universe that has a real beginning and final end or of an open Universe that expands indefinitely from a single event without any prospect of ever having an end.

Should the oscillating model be valid, we need not trouble ourselves with the concept of "existence" before the beginning of time. In this model, there is no beginning of time, no genuine start to the cosmos; its contents are endlessly recycled. Such a Universe always was and always will be.

The above models of the Universe stipulate evolutionary change as their guiding principle. Each is derivable from Einstein's general theory of relativity, and together they are favored, in one form or another and with modifications, by the majority of today's cosmologists. However, several other Universe models have been proposed over the years. Most of them do not follow directly from relativity; some don't even call for change with time or embrace evolution as their central theme. It's worth considering one of the more prominent ones, for until a few decades ago it was favored by leading members of the scientific community.

The "steady-state" model stipulates not only that the Universe appears roughly the same to all observers, but also that such a Universe appears unchanging to all observers *for all time*. Its fundamental tenet is embodied within what is sometimes called the *perfect* cosmological principle: To any observer at any time, the physical state of the Universe is much the same. In other words, the average density of the Universe remains eternally constant. It holds steady.

Initial motivation for a steady-state model was based as much on philosophy as on science. The cyclic Universe aside, many scientists and philosophers were (and still are) unwilling to concede that nothing could have existed prior to a unique big bang. Admittedly, it's challenging indeed to inquire about time and events preceding the origin of the Universe. What existed before the big bang? Why was there a big bang? What or who caused it? These are queries unaddressable within the realm of modern science. When there are no data or ways to experimentally test ideas, the scientific method is useless. Philosophies, religions, and cults of all sorts can offer hypotheses to the nth degree, but science remains mute. The steady-state model avoids these thorny questions, as does the oscillating model. For them, neither beginning nor end pertains. The Universe just *is* for all time.

Steady-state cosmologists concede that the Universe is expanding, for the recession of the galaxies is irrefutable. They nonetheless demand that the bulk view of the Universe—the average density of matter—remains constant forever. Accordingly, since the recession of the galaxies demonstrates the distances among galaxies to be increasing, the steady-state model requires the emergence of additional matter. Otherwise, with the galaxies separating, the average density would inevitably dilute. Odd as it may seem, the steady-statists proposed that this new matter is created from nothing. Despite the observed recession of the galaxies, the creation of additional galaxies in just the right amount can keep constant the number of galaxies per unit volume, thus preserving the same universal density forever.

The most vexing problem with the steady-state model is its failure to specify how the additional matter is created. Nor does it specify where.

Some researchers theorize its injection in the voids well beyond the galaxies in intergalactic space, whereas others prefer infusion within the bright centers of galaxies. Not much new matter is needed to offset the natural thinning as galaxies speed apart. Creation of a single hydrogen atom every few years in a volume the size of the New Orleans Superdome would suffice. Unfortunately, the sudden appearance of such a minute quantity of matter, either inside or outside galaxies, is currently quite impossible to detect and therefore to test.

Regardless of where matter is created, the real quandary is about how it's created. The sudden appearance of new matter from absolutely nothing violates one of the most cherished concepts of modern science—the conservation of mass and energy. This widely embraced principle of physics maintains that the sum of all matter and all energy is constant in any closed system. Matter can in fact be created from energy (and energy from matter), but it's tricky to understand how that matter can be spontaneously fashioned from nothing at all.

The grand puzzle of the steady-state model, then, is the process of material creation. Nonetheless, the lure of a Universe that always has existed and always will exist is strong, for it provides a way to skirt the need for a unique big bang and all the other awkward questions about the very start of an evolving Universe. All things considered, the big bang model is as troubling for a steady-state cosmologist to swallow as is this continual-creation idea for an evolutionary cosmologist. At any rate, and mental hang-ups aside, current observations have virtually destroyed any notion of a steady state, while fully embracing dynamism as key to the most feasible model of the Universe.

∞

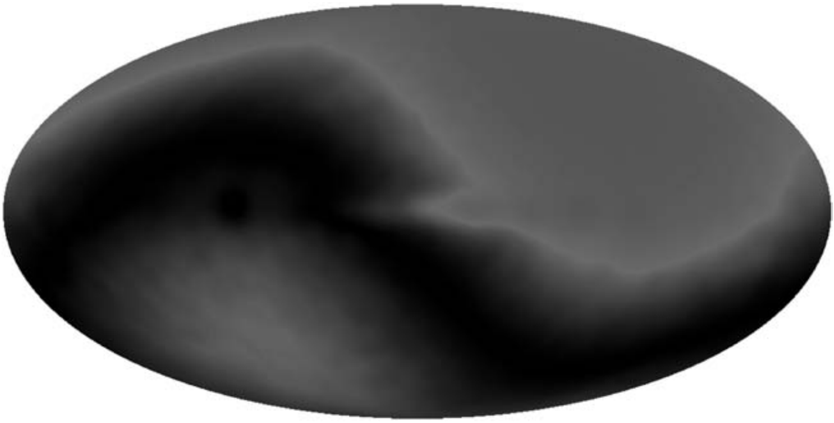
How can we distinguish among the various possible models of the Universe? Are there ways to rule out some of them and thereby converge on the best model by a process of elimination? Observational tests designed to answer questions like these have driven us further into the embrace of evolution as a guiding principle in cosmology—to be sure, evolution writ large as a unifying theme in all of science.

The steady-state model is widely judged untenable for at least two reasons. First, the spread of galaxies is not uniform throughout space. As noted in the Galactic Epoch, active (spasmodic) galaxies at great distances from Earth far outnumber those nearby; most neighboring, normal galaxies (including our Milky Way) are calmer, less active. Had we lived some ten billion years ago, when active galaxies were presumably the dominant astronomical objects, our view would have been filled with other active galaxies—many more than now surround our vantage point on Earth. The perfect cosmological principle is clearly violated: the large-scale view of the Universe was not the same eons ago as it is now; it's changed.

Second, a serendipitous discovery has turned out to be rather fatal for the steady-state model. Observations made with radio telescopes always yield a signal, regardless of the time of day or night. Unlike optical observations that often show a complete void of light toward dark and obscured regions of space, radio receivers never fail to detect some radiation. Sometimes the radio signal is strong, especially when the telescope is aimed toward an obvious source of radio emission. At other times it's weaker, particularly in regions devoid of all known radio sources. Yet, whenever the accumulated emissions from all known celestial objects and from all atmospheric and instrumental noise are accounted for, a minute radio signal always remains—a sort of weak hiss like static on a home AM radio or the "snow" on an inactive (noncable) television channel. Never diminishing or intensifying, this weak signal is detectable at any time of the day, any day of the year, year after year—it's omnipresent, apparently inundating all of space. What's more, it's equally intense in any direction of the sky—that is, it's isotropic. The whole Universe is apparently awash in this feeble but persistent radiation.

This ubiquitous radio signal was accidentally detected several decades ago, in the early years of the Space Age, while technicians struggled to improve America's telephone system. In their data, they unexpectedly noticed the bothersome radio hiss that just wouldn't go away. Unaware that they had detected a signal of cosmological significance, the researchers sought many different sources for the excess emission, including atmospheric storms, ground interference, equipment short circuits, even pigeon droppings left inside their radio antenna! Later, conversations with theorists enlightened the experimentalists about the static's most probable source: the fiery origin of the Universe itself.

This weak, isotropic, radio radiation is widely interpreted as a veritable "fossil" of the primeval event that began the universal expansion long ago. The leftover hiss, often termed the cosmic background radiation, floods every nook and cranny of space, including that surrounding us presently. Its existence is fully consistent with any of the evolu-

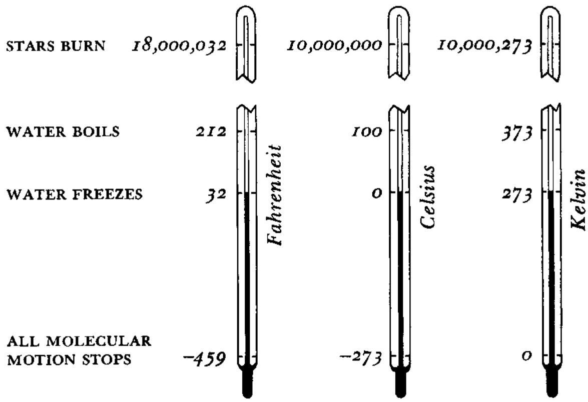


Evidence of the big bang.

This map of the entire sky was made by capturing weak, omnipresent radio waves launched from deep space and displayed here as a flat oval, much as maps of Earth's surface are often projected as oval shaped. It reveals that the cosmic background radiation is a little hotter in one direction (upper right) and a little cooler in the opposite direction. The difference is only a few thousandths of a degree Celsius and is caused by Earth's motion through space. This radiation is the Doppler-shifted remnant of a hot and dense Universe shortly after the big bang, but which is now much cooler and thinner some fourteen billion years later. *Source: Cosmic Background Explorer Satellite.*

tionary models of the Universe, but there's no role for it in the now defunct steady-state model.

The cosmic background radiation is presumed to be an ancient remnant of the extremely hot early Universe—a Universe that has greatly cooled during the past fourteen billion years or so. Regardless of whether the initial event was a unique big bang producing an open and infinite Universe or a closed and finite one, or even one of several repeated bangs of a cyclic Universe, the primeval, seething, dense matter must have emitted thermal radiation (as elementary particles naturally released energy while interacting with one another). All objects having any heat emit such radiation; a very hot piece of metal (a branding iron, for instance) glows with red- or white-hot brilliance, whereas less-hot metal (such as a home radiator) feels merely warm to the touch while emitting less-energetic infrared or radio radiation. In its fiery beginnings, the Universe almost certainly launched highly energetic radiation, but with time it expanded, thinned, and cooled, causing its emitted radiation to shift steadily from the lethal, high-energy gamma-ray and X-ray varieties normally associated with intensely hot matter, down through the less-energetic ultraviolet, visible, and infrared types, eventually becoming the harmless, lowest-energy radio waves usually released by relatively cool matter.



. . . some benchmarks on three temperature scales . . .

Evolutionary models predict that some fourteen billion years after the start of all things, the average temperature of the Universe—the relic of the big bang—should now be quite cold, in fact no more than about −270 degrees Celsius. That's far below the zero-degree-Celsius temperature at which water freezes and only a few degrees above the absolutely coldest value at which all atomic and molecular motions virtually cease. On the scientific scale, −270 degrees Celsius equals a mere 3 kelvins.

To confirm the theory, astronomers have carefully measured the intensity of this weak isotropic signal at a variety of frequencies up and down the radio band. All the data collected during the past few decades, especially those acquired by the *Cosmic Background Explorer* satellite in the early 1990s, are indeed consistent with a universal temperature of approximately 3 kelvins. Furthermore, this oldest fossil really does seem to pervade the whole Universe, including Earth, your house, or wherever you are now reading this. The amount of cosmic radiation present at any one time, however, is miniscule, totaling about a billionth of the power shone by a hundred-watt light bulb.

Existence of the cosmic background radiation, together with the spread of galaxies in space, discredits the steady-state idea as a viable model of the Universe. Clearly, the Universe has changed with time; it has not been steady at all. The choice of correct Universe type must then be made from among the evolutionary models. Other data must be obtained to sift through each of them.

The most straightforward way to distinguish between the open and closed Universe models requires an estimate of the average density of matter in the cosmos. More than anything else, density is what differentiates the closed model, which has enough matter to halt the expansion before it reaches infinity, from the open model, wherein there simply isn't enough to bring it all back.

We would be foolish to try to inventory all the matter in the Universe. Authors don't try to count by hand all the words in a written manuscript; rather, they make an estimate by counting the words on a single page and then multiplying by the number of pages (or nowadays let those incredibly fast morons known as computers count all the words for us). Likewise, astronomers try to measure the amount of matter within a certain volume of space and then extrapolate that amount to include the whole Universe. This is tantamount to estimating the mass density, for density is nothing more than mass per unit volume.

The precise density of matter—known as the "critical density"—needed to halt the expansion just as the outer limits of the Universe reach infinity can be computed theoretically. For today's thinned-out Universe, the answer is some million million million million million times less than one gram per cubic centimeter. (A cubic centimeter is just about the volume contained within a small sewing thimble, and a gram is about one five-hundredths of a pound. A thimbleful of water would have a mass of about one gram, the density of that liquid being one gram per cubic centimeter.) This extraordinarily small density amounts to a few hydrogen atoms within a volume the size of a typical household closet. That's terribly tenuous; in fact, many orders of magnitude thinner than the best vacuum attainable in laboratories on Earth. But remember, this is an average density of the entire Universe—calculated by lumping groups of galaxies, where the matter is most concentrated, together with intergalactic space, where little if any of it exists.

If the actual density of the Universe is less than this theoretically computed critical value, then the Universe is destined to expand forever, the hallmark of an open, infinite, hyperbolic model. If, on the other hand, the actual density exceeds this value, the Universe will someday stop expanding and start contracting, the fate of a closed, finite, elliptical model.

Theory aside, how can we determine the actual density of matter in the Universe? At first, it would seem simple. Just measure the total mass of all the visible galaxies residing within some large parcel of space, estimate the volume of that space, and compute the average density. Having done this many times for many pieces of cosmic real estate, astronomers usually find about ten times less density than the amount needed to halt the expansion of the Universe. As best we can tell, this calculation is independent of whether the chosen region contains only a few galaxies or a rich cluster of them; the resulting density is roughly the same, within a factor of two or three. Galaxy-counting exercises of this sort therefore imply that the Universe is open, meaning that it originated from a unique big bang and will expand forever. Such a Universe has no end, though it definitely had a beginning.

But—and this is a crucial but—an important caveat deserves mention. All the matter in the Universe is not likely housed exclusively within the brightly visible galaxies. Observations imply that invisible matter exists beyond each of them—"dark matter" sensed only indirectly by means of its gravitational effects mostly outside galaxies. The extent and amount of this dark matter is presently unclear, but if much additional matter resides outside the galaxies as within them, then the universal density would correspondingly increase. Reservoirs of as-yet-unseen matter skirting the galaxies could reverse the solution to this first cosmological test, forecasting a closed Universe possibly having an end as well as a beginning. Whether such a Universe originated from a unique big bang prior to which nothing at all existed or whether such a Universe ends for all time without bouncing, cannot be addressed by this test.

Frankly, that astronomers are deeply puzzled about the nature of this dark, or hidden, matter is an understatement. We don't know what it is, only that it almost surely exists. Nor do we know much about how it's distributed in space, but there are some clues. We can only indirectly infer its effects by two methods, each of which measures the dynamical behavior of individual galaxies: First, the outer parts of galaxies rotate faster than expected for the visible matter seen, implying that invisible (or dark) matter must be present to gravitationally prevent those outer parts from dispersing. Second, within much larger groups of galaxies, some galaxies have motions so large that they should have escaped from their group long ago—unless, again, some sort of dark matter were gravitationally binding the groups together.

What is this dark stuff and where is it hiding? For the past few decades, astronomers have sought unconventional forms of normal, or "baryonic," matter, suspecting that they may have overlooked an important part of their cosmic inventory. (Baryons include the atoms of which all stars, planets, and life-forms are made—mostly the protons and neutrons that constitute our tangible world, namely, the basic ingredients in the chemist's periodic table of the elements.) For example, cold, tenuous matter might be lurking in and amongst the galaxies, but radio astronomers, whose equipment is most sensitive to this kind of low-energy gas, have found little of it. Hot, tenuous matter is another possibility, but X-ray astronomers who are best equipped to detect such intensely glowing, high-energy gas have also found hardly enough of it to account for the hidden matter. Dwarf stars that are not only small but very dim, especially among the rich globular star clusters in the large, spherical halos of galaxies, are yet another candidate for locales where matter might have gone unseen, but recent, direct observations of such clusters have found surprisingly few dwarf stars. Wandering blobs of compressed matter, either clumps of gas that never achieved official stardom or burned-out cores of erstwhile stars—collectively called massive compact halo objects, or MACHOs for short—were once a leading possibility, but few of them have been spotted in the halo of our Milky Way and none at all in distant galaxies. Even black holes, as noted later in the Stellar Epoch, are not found in great abundance, making them unlikely places to trap lots of matter that cannot be seen. And so on, down the list of many candidates for normal matter, none of which has panned out in recent years, despite direct, exhaustive observational searches for them.

In contrast, most astronomers are now agreed that the bulk of the suspected dark matter is probably made of material that is abnormal. Dark matter is more likely "nonbaryonic," that is, composed of matter completely different from atoms as we know them. This type of dark matter probably exists as exotic subatomic particles, formed in the early Universe and now moving around sluggishly and elusively. Known as "cold dark matter," it also collectively goes by the acronym WIMPs, which stands for weakly interacting massive particles —weakly interacting since they ostensibly remain aloof from normal matter, yet massive since they still exert gravity even if they are "dark" and emit no light. To solve the dark-matter problem, these putative particles would have had to survive to the present day in gargantuan quantities and to pervade virtually every part of the cosmos. Alas, no one has ever seen such particles directly or even evidenced them indirectly, and no telescope has been built to detect such peculiar stuff. Our best bet to do so is in the high-energy accelerators where elementary particles can be created from packets of energy, as noted in the Particle Epoch, but thus far no WIMPs have emerged. Whatever the dark matter is, its presence seems unambiguous. Until its nature and composition are resolved, the issue of dark matter will remain one of the thorniest challenges in the world of astronomy.

How much dark matter are we talking about? Some observations imply that each galaxy could conceivably contain as much as ten times more dark matter than its luminous material, and the figure for groups of galaxies might even be higher; astonishingly, perhaps as much as ninety-five percent of the total mass in the huge galaxy clusters is invisible. Even so, based on the best data available today, most astronomers now reason that dark matter probably raises the overall cosmic density to no more than about a third of that critical value needed to collapse the Universe in some far-future time.

The observed universal density determined by this galaxy-counting method is thus quite uncertain at present. This test cannot clearly distinguish between the open and closed models, though at face value it favors an open Universe destined to expand forever.

Another observational test seeks to determine the ultimate fate of the Universe, and here a new and unexpected result has been recently reported. Apparently, the real Universe might be a great deal stranger—and more complicated—than the simple models outlined earlier. The newest data suggest that the Universe may be not merely changing nor merely expanding but actually receding at ever-faster rates—a shocking development that has profound implications for cosmology.

Like the first destiny test, this second test also seeks to estimate the average mass density of the Universe. And it again relies on the fact that each and every piece of matter gravitationally pulls on all other pieces of matter. This second test addresses the question, How fast is gravity applying the brakes, which would ordinarily cause cosmic expansion to decelerate? Put another way, what is the rate of change of evolutionary change?

Given that the Universe began in a violent bang, it must have expanded rapidly at first, thereafter gradually growing more sluggish. The expansion of anything—the debris of a bomb, the sound of a thunderclap, whatever—is always greater at the moment of explosion than at some later time. Hence, since looking out into space is equivalent to probing back into time, the recessional motions of the galaxies should be larger for the distant galaxies and somewhat smaller for those nearby.

Observationally, cosmologists try to detect any change in the Doppler-shifted velocities of our neighboring galaxies compared to those far away. This change is presumed greater for the finite, closed model of the Universe, since the large amounts of matter needed to stop and then contract the Universe would have well slowed its expansion over the course of fourteen billion years. The infinite, open model is predicted to show smaller changes in the galactic recessional velocities; in this case the deceleration of the Universe would be less.

Surprisingly, data acquired in the late 1990s show none of this expected deceleration. Instead, observations of supernovae (exploded stars) in distant galaxies imply that the Universe is speeding up—in short, accelerating! Basically, the brightnesses of the supernovae are fainter than expected, meaning that they are probably farther away and somehow they had to get out there. If it were not for the fact that two independent groups of astronomers found the same startling result, no one would believe it. But science is not a matter of belief, and the data do clearly imply that the galaxies at large distances (hence seen far back in time) are receding less rapidly than expected. These data are not yet foolproof and their interpretation is still subject to debate, but, if they hold up, the new results will force a major revision in our Universe models.

Not that astronomers need to return to the drawing boards; that would be too drastic. Contrary to many hyped news reports, this surprising finding does not mean that big bang cosmology has been overthrown. The new results mandate a *revision* but not a revolution in our previous thinking. Some "wiggle room" still exists in the analysis, and some astronomers prefer to discount the new data as still inaccurate or poorly acquired or observationally biased. Others argue that faraway supernovae might be dimmed because of tainted radiation from ancient stars or attenuating dust along our line of sight to them. Nonetheless, most researchers have reluctantly accepted these new results, and, for now at least, the speeding Universe seems to be real.

What could be the cause of such cosmic acceleration whose effects (as for any accelerated object) are likely to have been minimal in the past yet will be more dramatic in the future? Frankly, it's unknown, but one ironic possibility is that the culprit is the same "cosmological constant" invented (largely out of thin air) by Einstein decades ago to act as a repulsive force to counter gravity and thus keep his Universe models from collapsing. This factor acts only on the largest scales, thus potentially explaining its dormancy for the first many billions of years of universal history; only today would it be emerging as a major factor in cosmic expansion. What's more, it's thought to arise from "vacuum energy" associated with empty space itself, thus potentially accounting for a negative, or outward pushing and repulsive, pressure that might increasingly challenge gravity on the largest scales. In other words, according to quantum theory, any region of "empty" space—traditionally called a vacuum—actually seethes with energy as subatomic particles burst in and out of existence for extraordinarily short periods of time. Not to do so would be a violation of physical law, specifically of Heisenberg's uncertainty principle, which cannot bear the certainty of true emptiness. Vacuum energy thereby gives even to empty space a push at every point.

That said, astronomers have no clear understanding of the cosmological constant, and physicists can't even define it. At best, we know that it must be related to a new kind of force whose strength, quite unlike gravity, must *increase* with distance. It would therefore grow stronger over the course of time, thereby escalating to runaway expansion on large scales yet remaining negligible on small scales so as to avoid interfering seriously with Einstein's gravity, which has been so well tested locally in the Solar System. This wholly new force, whose physical significance is mostly a mystery and whose numerical value is largely unknown, is neither required nor explained by any currently known law of physics.

Quintessence is the fanciful name of another candidate phenomenon that might force the Universe outward, ever faster with time. Beyond the Aristotelian notion that all terrestrial things are made of four interchangeable elements—air, earth, fire, and water—the "fifth essence" of ancient Greek philosophy was responsible for celestial phenomena. Hence, an archaic term returns (in name only) to describe an omnipresent property of spacetime that has the dual, and most peculiar, effect of both positive mass to gravitationally clump matter locally and negative pressure to accelerate the expansion of the Universe globally.

But if quintessence does exist, where did it come from, and is it any less ad hoc than the cosmological constant?

Whatever it is and however it works, the mysterious force that might cause the Universe to accelerate apparently derives from neither conventional matter nor ordinary radiation. For now, and partly as a pun—both on the nature of the substance and the extent of our ignorance—astronomers have given it the name "dark energy."

Dark matter and dark energy have become embarrassing for astronomers struggling to inventory the Universe. Dark matter itself—whatever it is—seems to outnumber by a factor of ten the normal (baryonic) matter of which galaxies, stars, planets, and life are made. And now, dark energy—whatever that is—dominates them all. Numerically, normal matter probably makes up only a few percent of the Universe, dark matter about thirty percent, and dark energy the rest—implying that more than ninety-five percent of the Universe is unaccounted for! Having so much of the Universe on "the dark side" is highly disconcerting, and most scientists are more than a little uneasy about it.

Do note that an outside chance remains for dark matter and dark energy to be hardly more than theoretical artifacts devoid of reality. Both quantities are merely inferred to keep the Universe "balanced"—that is, to theoretically grant it that precise critical density demanded by the nearly equally peculiar concept of "cosmic inflation," thereby making the Universe globally flat and of zero net energy (all of which is to be described shortly in the Particle Epoch). This troubling state of affairs sits like a bone in the collective throat of many astrophysicists, and someday it could conceivably be shown to be incorrect, especially the need for dark energy. All this uncertainty makes the astronomical community feel insecure, as though the mounting complications may well bring down our intricately constructed house of cards—the standard model of modern cosmology—as the next generation of scientists seeks to reinfuse simplicity into one of humankind's greatest intellectual adventures.

Luckily for us in this book, the recent findings of possible cosmic acceleration do not strongly affect our cosmic-evolutionary narrative. Dark energy (if it really exists) was likely a negligible factor for the first many billions of years, only recently becoming more relevant as the cosmos geared up. Nor does our present ignorance of dark matter much affect this story of natural history, for it too obeys gravity, which dominates on large scales and doesn't care much about particulars. Just as our scenario's validity holds whether the Universe is as young as ten or as old as twenty billion years—provided the relative ages are consistently and chronologically sequenced along the arrow of time—cosmic evolutionists are prepared to revise the narrative to incorporate the latest data. As for the future, an accelerating Universe portends an even more dramatic rise in energy flows, novel environments, and ordered structures—the likely result being ever-greater diversity and richness among all types of complex systems, including life.

∞

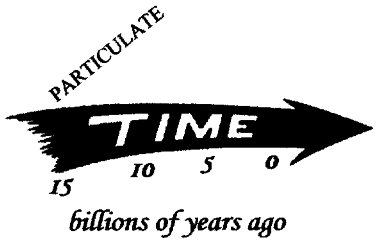
The big picture of the Universe seems both gratifyingly well understood in gross fashion yet puzzlingly unresolved in its devilish details. Present consensus suggests an evolutionary Universe that expands forevermore, but its origin, destiny, and basic composition remain concealed in lingering uncertainties. As we now enter a golden age of cosmology, many astrophysicists are inclined to say that we should expect definite answers within a few years. This is perhaps overly optimistic, for the final solution requires the agreement of three often disparate groups of human beings:

First, there are the theoreticians, whose imaginative minds invent the model Universes, in the process striving to stay within the bounds of good, solid, accurate science. They try to determine what the Universe is supposed to be like. Second, there are the experimentalists, constantly testing the theories, all the while extending their observations to more distant realms within the real Universe. They try to determine what the Universe actually is like. And third, there are the skeptics, who regard the models of the first group as mere speculation and the results of the second group as overinterpretation of the data without due regard for observational error.

In the end, all three attitudes are helpful and necessary, for only by their cooperation and counteraction can we ever hope to approach the truth. Fortunately, the cosmic-evolutionary story—a telling of natural history from shortly after the big bang until now—is largely independent of which specific cosmological model is correct or what may be the ultimate fate (of the models and of the Universe!). All models include universal expansion, as well they must given the indisputable fact of galaxy recession—and it is expansion, more than anything else, that drives the potential for the rise of order, form, and structure in the Universe.

1. PARTICLE EPOCH

Simplicity Fleeting



WHAT WAS IT LIKE AT THE ORIGIN of the Universe? Exactly what happened at the instant of time's beginning? Can anything concrete be said about the precise start of the Universe itself or about the prevailing conditions during its first few moments? And how have those conditions changed to give rise to the Universe we see around us today?

These are surely fundamental questions. They are also hard questions. Yet they are among the most basic wonders that perhaps every thinking human being who has ever lived has contemplated in one way or another, at one time or another. Now, after more than ten thousand years of organized civilization, twenty-first-century science seems poised to provide some testable ideas regarding the ultimate origin of all things.

Solutions that scientists have devised should be considered qualified and provisional. Times long past are times long gone. It's difficult to be precise about events we cannot observe directly. To be sure, the early Universe was a decidedly unearthly domain, quite unlike anything encountered today. Neither stars nor planets, indeed not even atoms then existed; all was pure energy, the cosmic currency that makes change happen. Nonetheless, as noted in the prologue, Universe models can be constructed—mathematical sketches based on theoretical insights and lodes of data constraining the size, shape, and structure of the cosmos.

These models grant us some inkling of what the Universe was like well more than ten billion years ago, indeed surprisingly close to its very origin.

∞

To appreciate the earliest Particle Epoch of the Universe, we must be willing to think deeply about times long, long ago. We must strive to imagine what it was like well before Earth and the Sun emerged, even before any planet or star existed. Some people have trouble mentally visualizing such truly ancient times. Fortunately, a trick can help us comprehend the earliest moments of the Universe.

Physicists are mainly charged with the application of the laws of Nature to the present state of something in order to predict its future. Although, in recent years, a renewed respect for the role of chance has somewhat diminished our ability to predict outcomes in the old, mechanistic, Newtonian sense, we still like to try our hand at predicting general trends, if not the details. In the case of the whole Universe, that "something" is literally all things—nothing in particular, just everything in general. Hence, if we find it mentally hard to reverse time to appreciate the earliest epoch of the Universe, we can instead take advantage of the natural symmetry of a model Universe that will eventually contract and thereby predict the physical events destined to occur as a closed Universe nears its final phase of total collapse. This procedure is valid only because the mathematics describing contraction are a mirror image of those for expansion. In other words, the events that *will* occur just prior to the end of a contracting Universe mimic those that *already* happened just after the start of an expanding Universe. Not that time ever does reverse, as best we know. Rather, we can use some of the symmetry built into the laws of physics to estimate the final events of such a hypothetically closed Universe, thus gaining some inkling of the initial events some fourteen billion years ago.

Even if the Universe is not closed in this way and will never collapse to a singularity, astrophysicists employ closed models in order to understand theoretically some of the highlights of the earliest epoch of either a closed or an open evolutionary Universe. It's an example of how we can use symmetry and scaling arguments—to scale models up, scale them down, or in this case to scale them back in time—in order to recreate mentally places and times we could never visit physically.

"Numerical experiments" are needed to crank out the Universe models. These are essentially number-crunching exercises, utilizing mathematical knowledge of the laws of physics and sophisticated software running on powerful computers. The resulting simulations are ponderous and computationally intensive, incorporating much of what we know about the bulk features of the Universe—again, largely the generalities, minus the messy details. The objective is to determine the average density and average temperature for the whole Universe at any moment in time. The input numbers, such as mass, energy, and expansion rate, can be varied as the computer routines are run again and again, the idea being to match the resulting state of the model Universe with that of the currently observed, real Universe. In this way, the range of input values can be progressively narrowed, thereby converging on a description of the Universe that reasonably mimics reality.

Most computer models suggest that in the beginning, there was chaos! But, frankly, we are sometimes unsure if the chaos was in the Universe or is now in our computer codes. Again, the problem is the singularity at the moment of the big bang itself—a decidedly odd state about which mathematicians are currently perplexed. It's hard to imagine that science will ever be able to prove what happened at the exact moment of the bang—precisely zero time. That's why big bang cosmology, contrary to popular belief, is not a theory of the big bang per se. Rather, it's a cognitive map, or worldview, that aspires to explain events in the aftermath of the big bang.

Many theorists contend that the physical conditions can be approximated for extremely short times after the bang, well less than the first second of existence. For example, most models specify that a Universe younger than a trillionth of a trillionth of a second (i.e., 10^{-24} second) would have had an average density greater than a trillion trillion trillion trillion (or 10^{48}) grams per cubic centimeter and an average temperature greater than a billion trillion (10^{21}) degrees Celsius. By way of comparison, the average densities of water and lead are one and ten grams per cubic centimeter, respectively, and of atomic nuclei a trillion grams per cubic centimeter. Also, the present average density of all material objects in the Universe is roughly a million trillion trillion times *less* than that of water (or about 10^{-30} gram per cubic centimeter); this is the average density of everything—galaxies, stars, planets and life-forms, as well as mostly empty space. Similarly, water freezes at 0 degree Celsius and boils at 100 degrees Celsius, while the average temperature at the surface of an ordinary star is several thousand degrees Celsius. The present temperature of everything in the Universe, again on average, is only a few degrees above "absolute zero," some −270 degrees Celsius (or 3 degrees Kelvin).

As for the time just noted, it's nearly impossible to appreciate such youth; 10^{-24} second is the duration needed for light to cross a proton—the nucleus of the smallest atom. Such minute fractions of time—much quicker than a flash, literally—are as incomprehensible as the huge densities and extreme temperatures characterizing the early Universe. Yet these are the conditions specified by the laws of physics as a contracting Universe inexorably speeds toward its demise. They are thus, through the above symmetry arguments, the conditions thought to prevail in those violent moments shortly after the birth of the Universe.

The composition of the Universe at such extraordinarily early times is hardly describable. Surely, much energy must have existed as essentially pure radiation, along with exotic elementary particles of many types, but beyond that science can currently only speculate. The dominant action at the start of the Particle Epoch must have been nearly unimaginable. Undaunted, we shall return soon to take a stab at it, but first we pause to remind our minds.

∞

Here, we sidetrack for a brief review of the fundamental makeup of matter and a short note about the basic forces that govern it. By this we mean normal, or baryonic, matter, for we can hardly address dark matter more than we already have, given that there are so few clues about what it really is.

Exploration of the basic nature of matter is not new. At least as far back as ancient Greece, attempts were made to unravel the composition of all things. Although clearly great thinkers, the Greek philosophers were deeply in error; they presumed that thinking about Nature was better than looking at it. Still, their ideas prevailed for more than two thousand years, culminating in the witchcraft and magic that befuddled the efforts of medieval astrologers and Dark Age alchemists.

Only with the rise of logical, deductive reasoning during Renaissance times, and especially its heavy reliance on testing, did the technique of "experimental philosophy" become fashionable. At last, a proper balance between thinking and looking was achieved. The technique is simple: thoughts (theoretical work) are to be taken seriously only if confirmed by tests (experimental work). Modern science thereby emerged and with it the "scientific method."

Not that the scientific method is entirely objective, as confessed in the preface. Science is practiced by human beings, and scientists are no different from others who have subjective emotions and personal biases. Yet over the course of time, criticism, and debate, scientific issues eventually gain a measure of objectivity. By incessantly demanding tests and proven facts, the scientific community gradually damps the subjectivity of individuals and arrives at a more objective view among a community of critical thinkers. Skepticism and doubt are essential features of the modern scientific method.

In one of the greatest triumphs of the scientific method to date, physicists of a century ago were able to prove that atoms are not the most basic entities of Nature. All atoms of every different kind—that is, all elements—are made of negatively charged electrons whirling around positively charged nuclei. Each neutral atom has equal numbers of electrons and protons, as well as a similar number of neutrons. The protons and neutrons contain virtually all the mass of any atom, and together they constitute the atom's nucleus—so compact relative to the size of the larger atom as to be like a grain of sand floating alone amid a sphere the size of a football stadium.

For the first half of the twentieth century, electrons, protons, and neutrons, along with photons of radiation, were considered the very essence of matter. However, during the second half of the century, physicists also discovered a bewildering array of additional elementary particles. These newer particles are not likely any more "elementary" or basic than the better-known protons and electrons. Rather, each one seems to play its own role in the subatomic realm far from everyday familiarity. And that role is not always clear, as none of them can be seen directly; they can only be inferred when passing through laboratory apparatus.

More than two hundred elementary particles are currently known—which makes one wonder just how elementary, or fundamental, they really are. Some behave like lightweight electrons, whereas others resemble heavyweight protons. Still others display bizarre properties not yet understood. Many particles exist for only fleeting moments during fierce collisions induced in high-energy accelerators—vast underground laboratories where, typically, electrons and protons are boosted to velocities near the speed of light and then slammed together violently. The largest and most powerful of these machines are the Conseil Européen pour la Recherche Nucleaire (CERN), which extends for several kilometers across the French-Swiss border near Geneva, and the Fermi Laboratory, which spans a similarly large piece of real estate beneath a Chicago suburb. The new particles literally materialize from the energy of the collisions; no magic is involved, as this is a well-understood physical process. Usually, after a microsecond or so, the particles change back into energy, but not before leaving behind momentary traces on the accelerators' detectors.

The history of efforts to decipher the building blocks of Nature is full of false claims. Each time researchers thought they had discovered a truly basic component of matter, they have been proved wrong. With molecules now known to be made of atoms, and atoms in turn made of elementary particles, other questions naturally come to mind: How elementary are the new particles seen in the debris of accelerator collisions? Are these particles perhaps made of even more fundamental subparticles that have some identity or existence of their own? Current theory and some data do suggest another layer of fundamentality.

Popular consensus has it that protons and neutrons, among a whole menagerie of elementary particles (called "hadrons") with sizes of order 10^{-13} centimeter, are made of units called quarks, and together they make up more than ninety-nine percent of normal, baryonic mass in the Universe; the rest is made mostly of dimensionless electrons, which are not dividable into quarks or apparently anything else. Quarks (which derive their name from a meaningless word coined by the novelist James Joyce in his book *Finnegan's Wake*) are minute particles having only a fraction of the electric charge carried by a proton. For example, a proton consists of two "up" quarks (each with two-thirds charge) and one "down" quark (having a negative one-third charge); a neutron has one up and two down quarks. Over the past few decades, an intricately detailed yet remarkably successful theory, called quantum chromodynamics (or QCD, for short), has been refined around six quarks having the metaphorical names of up, down, top, bottom, strange, and charm, each variously bound by yet another elementary entity, the gluon. Despite its inherently fuzzy (quantum) nature and oft-intractable equations, this mathematically elegant theory, aspects of which have generated Nobel Prizes for more than twenty physicists, underlies many popular products in today's technological world, including televisions, lasers, computers, and a whole industry of electronic devices built on digital chips.

Originally, when the idea of quarks was first proposed several decades ago, they were mostly judged to be no more than a mathematical convenience—a mental bookkeeping system for describing quantum interactions—not real objects that could be studied tangibly. Nowadays, accelerator experiments clearly demonstrate the physical existence of the six different kinds of quarks, mainly by observing the way electrons deflect when fired at protons. These events involve violent, head-on collisions, a little akin to a hypothetical attempt to understand the makeup of a clock by smashing two of them together at high speed. Traces of all six quarks have now been found in the accelerator debris, and together they form the essence of the "standard model" of particle physics—a widely acknowledged description of submicroscopic phenomena, bolstered by accelerator experiments and the quantum theory of particles and forces. Yet no compelling reasons exist to prohibit Nature from having more of them, indeed the six quarks are thought to have six partners (called "leptons"), of which the electron is one. All of which suggests a conundrum: the very proliferation of quarks and their relatives threatens to topple the central idea that we have reached a truly fundamental realm of matter.

On paper (that is, in highly theoretical terms) and on very basic scales (that is, much smaller than even that of quarks)—thus well removed from anything testable—physicists are actively investigating nearly intractable mathematical models that seek to interpret particles not as minute points but as "strings." This approach, known as string theory, envisions matter on scales as small as 10^{-33} centimeter—twenty orders of magnitude smaller than a proton—as modes of vibration among ultra-submicroscopic items that, if we could see them, would resemble strings and loops vibrating in well more than four dimensions of spacetime. As complex as it sounds and as remote as it is from everyday practicality, string theory is considered "beautiful" and "elegant" by the experts pursuing it, many of whom feel that it offers the best hope to unify all the known forces of Nature. We shall return to reconsider strings later in this Particle Epoch.

As best we can tell, the behavior of normal matter on all scales—from elementary particles to clusters of galaxies—is ruled by just a few basic forces. Forces and the fields and energies they engender are the root cause of changes everywhere; they are *fundamental* to *everything* in the cosmos. In a sense, the search to understand the nature of the Universe is synonymous with the quest to understand the nature of these forces. Forces, fields, and energies are among the essential keys needed to unlock some of the most concealed secrets of the Universe.

Gravity is perhaps the best known force, binding galaxies, stars, and planets and also, of course, holding us on Earth. Like other forces, its strength decreases with distance from any object; in fact, it decreases as the square of the distance, and is said to obey the "inverse square law." However, that's only half of the law of gravity, as its strength is also proportional to mass. Thus, gravity is terribly weak near, for example, a puny atom, but enormously powerful near a huge galaxy. In fact, although gravity is by far the weakest of all of Nature's known forces, its effect can accumulate impressively over large volumes of space that contain mass. Nor can anything cancel the attractive pull of gravity; there is no such thing as antigravity that repels objects—at least not for normal matter. Even the peculiar stuff known as antimatter (discussed a few pages hence) has gravity, not antigravity. Consequently, the gravitational forces of all objects—even our own bodies—extend to the limits of the Universe, hence the reason why gravity is known as a "long-range" force. To be sure, on scales larger than Earth, gravity is the dominant force in the Universe.

The electromagnetic force is another of Nature's basic agents. Any particle having a net electric charge, like an atom's electron and proton, exerts an electromagnetic force. This force acts as the cement for most ordinary materials, including virtually everything in our homes, such as tables, chairs, books, even the kitchen sink. Because the electromagnetic force also binds the atoms within all life-forms, some biologists call it the "life force"—which, unfortunately, leads some to think that life is governed by some special "vitalism," which is wrong. Like gravity, the strength of the electromagnetic force decreases with distance according to the same inverse square law. But unlike gravity, it can repel (between like charges) as well as attract (between opposite charges). Such forces can then sometimes cancel one another, as when similar numbers of positive and negative charges neutralize the electromagnetic force, thereby diminishing its influence. For example, although human bodies are made of very many (about a billion billion billion, or 10^{27}) charged particles, they comprise almost equal mixtures of positive and negative charges; our bodies therefore exert hardly any net electromagnetic force. Overall, electromagnetism is much stronger than gravity on microscopic scales and smaller but is relatively unimportant on macroscopic scales where gravity rules.

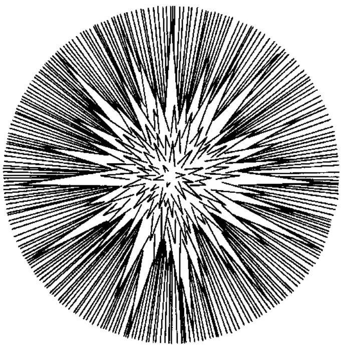
A third fundamental force is termed the weak nuclear force, as its effective range is less than the size of an atomic nucleus and its influence on matter much more subtle than any of the other forces. We shall not encounter it much in the course of describing cosmic evolution, except to note that the weak force helps to change one kind of elementary particle into another (such as the arcane neutrino particles released during nuclear reactions at the Sun's core). The weak force also governs the emission of radiation from radioactive atoms, which are useful in establishing dates that, in turn, reveal the tempo of cosmic evolution. Most scientists now agree that the weak force is not really a separate force at all; rather, it's probably another form of the electromagnetic force acting under peculiar circumstances. As such, we now often speak of the "electroweak force," an idea to which we shall return in the next section.

A stronger force than any of these is the nuclear force, mediated by the gluon particle that holds the quarks together. It glues—hence its name—protons and neutrons within atomic nuclei and, in effect, serves as the source of energy in the Sun and stars. Like the weak force, yet unlike the forces of gravity and electromagnetism, the nuclear force operates only at very close range; it's useless when matter is separated by more than a trillionth (10^{-12}) of a centimeter. But within this range, as for all atomic nuclei, it binds particles with enormous strength—stronger, in fact, than any other force known. Numerically, and in absolute terms independent of their most potent ranges, the nuclear force is 137 times stronger than the electromagnetic force, 100,000 times stronger than the weak force, and 10^{39} times stronger than gravity. Ironically and despite its extraordinary weakness, gravity is the only force that affects all things at all times on all scales.

And, then, there is dark energy, as noted (grudgingly) at the end of the prologue, which implies another, perhaps wholly new fifth force about which science is thus far truly in the dark.

∞

By most accounts, the Universe originated with the expansion of an unbelievably hot and dense "something"—hotter than the tens of millions of degrees Celsius in the cores of most stars, denser than the trillions of grams per cubic centimeter in the nucleus of any atom. Precisely what that state was, we cannot say for sure. And why it "exploded," we really don't know. At best, science contends that in the beginning a singularity released an outward burst of pure, radiant energy. *Why* the Universe suddenly began expanding more than ten billion years ago is a most intractable query—so formidable that scientists are currently unaware even how to formulate a meaningful question about it.



. . . the Universe originated with the expansion of an unbelievably hot and dense "something."

In the broadest sense, there are *what* questions, *how* questions, and *why* questions. Using astronomical telescopes and biological microscopes, among an arsenal of other experimental gear, researchers have employed the reductionist approach that has served science so well since Renaissance times to unravel both the macroscopic and microscopic nature of matter—namely, to tally fairly well *what* exists in the Universe—from atoms to galaxies and from cells to animals. Not that our inventory is complete by any means, for the nature of dark matter (let alone dark energy) remains unresolved. Yet, armed with a rather detailed inventory of what astronomers call "normal" matter—which is all that we perceive directly in the Universe—we are able to address the origin and evolution of that matter, in other words, *how* it got there initially and *how* it has changed ever since.

To inquire about the nature of the very beginning, however, requires us to address *why* questions, the most fundamental of all being, Why is there a Universe at all? To be honest, scientists don't know how to tackle why questions. These are outside the present fabric of modern science and probably always will be. In other words, when a ball is put in motion and Newton tells us that the ball "will remain in motion unless or until it is acted upon by some external force," we have no understanding *why* it does that. We do know what such balls do and also how they do it, but we have no clue why. No known procedure—not even the vaunted scientific method—enables us to investigate why, in the deepest sense of that word, the laws of physics and biology are as they are. We shall probably never know the answer. Nor do we know, or likely have any prospect of ever really knowing, why there is a Universe—or what might have preceded its origin.

The basic problem in attempting to discover the nature of what, if anything, existed prior to the very start of the Universe is simple: There are no data. None whatsoever. Sure, some people have hypotheses, but these hypotheses are not based upon data. They are in every case contingent largely on thoughts or beliefs, and while noble and comforting to many in society, they cannot be considered science. The methods used by scientists and those used by philosophers and theologians are as different as oil and water; they just don't mix. To be crass about it, if it's experimentally or observationally testable, then it qualifies as science; if it's not, then it's something else.

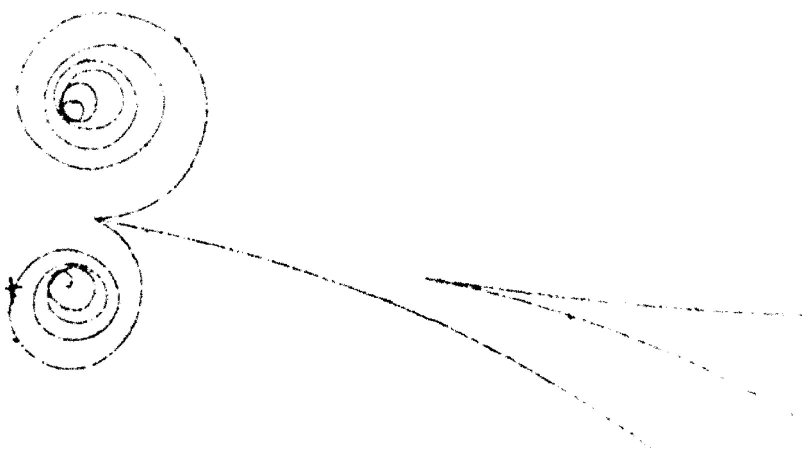
This is not a criticism of people who wonder about the start of the Universe, or even about what might have come before it. Long ago, Augustine related a popular fifth-century idea that before creating heaven and Earth, God made hell for those who worry about such issues. Augustine himself was more likely correct in thinking that the Universe was made *with* time, not *in* time. Even today's scientists occasionally ponder how to devise experiments to gather pre-Universal data. However, as things stand now, queries about the nature of whatever existed before the bang amount to inquiries less about the origin of the Universe and more about the origin of the origin. Not to be overly critical, what came before the big bang might well be a meaningless puzzle—like the popular medieval exercise of counting angels balanced on the head of a pin—since at the beginning of the Universe, matter, energy, space, *and* time probably all came into being. Resembling the "here-there-be-dragons" school of ancient cartography, time before the big bang did not likely exist. To most cosmologists, asking what happened prior to the big bang is akin to asking what lies north of the North Pole!

In what follows, we necessarily confine our discussion to events extant since the start of the Universe, based on our knowledge of its existence during the past fourteen billion years, and regardless of why the

Universe did originate. Indeed, cosmic evolution constitutes a broad synthesis of the whats and hows, and not at all of the whys.

Within a microsecond of its beginning, the fiery Universe was flooded with energy throughout every available niche. It was also peppered with a whole mélange of subatomic particles of matter, whizzing this way and that amidst great heat and blinding light. Whence did these particles come? From radiation, pure and simple. These particles "materialized"—a creation of sorts—as matter was literally fashioned from the energy of the primeval bang. Neither magic nor mysticism prevailed, just a well-known and oft-studied fact that the elementary building blocks of matter result from clashes among packets of energetic radiation. The interchangeability of matter and energy is proved daily in the underground bowels of particle accelerators around the world, the two obeying that most famous of all formulas noted in the prologue: $E = mc^2$.

Foremost among the particles made well within the first second of existence were the quarks and their gluon associates. A quark-gluon plasma, colloquially known as "quark soup," prevailed in the Universe just prior to the natural emergence of protons, neutrons, and other heavy



Evidence of particle production.

Tracks of elementary particles observed in high-energy accelerators often display particle-antiparticle creation. Here, a gamma-ray photon arrives from the right, suddenly yielding its energy to produce an electron-positron pair. The pairs of particles curve in opposite directions in the detector's magnetic field because of their opposite electric charges. *Source: CERN.*

elementary particles that are built of quarks. (Plasma, the "fourth state of matter" after solids, liquids, and gases, exclusively comprises charged particles, normally protons and electrons.) As bizarre as this stuff seems, quark soup was actually verified—that is, a whole new state of matter created—in one of the first notable accelerator experiments of the twenty-first century. Physicists did it by slamming together two gold nuclei at ninety-nine percent of the speed of light and then examining the thousands of particles that sprayed out from the ensuing meltdown. The superenergetic result was a seething concoction of free-roaming quarks and gluons—a miniature fireball of sorts—momentarily produced and controlled in the laboratory.

Soon thereafter, yet still only about a microsecond after the big bang, heavy, strongly interacting elementary particles such as protons and neutrons—those collectively called "hadrons"—became the most abundant types of matter. Such particles must have then existed as free, unbound entities, given the inferno prevalent in the Universe within its first second of existence. It was just too hot for these particles to have assembled into anything more ordered. Hadrons surely collided and interacted with one another as well as with other types of elementary particles, for the density then was also extremely high. Accordingly, the dominant action at this time was the inception and then self-annihilation of hadrons into radiation, which further fueled the brilliant fireball. Lacking a good understanding of elementary particles at the highest energies, physicists have only partial knowledge about this puzzling period in cosmic history.

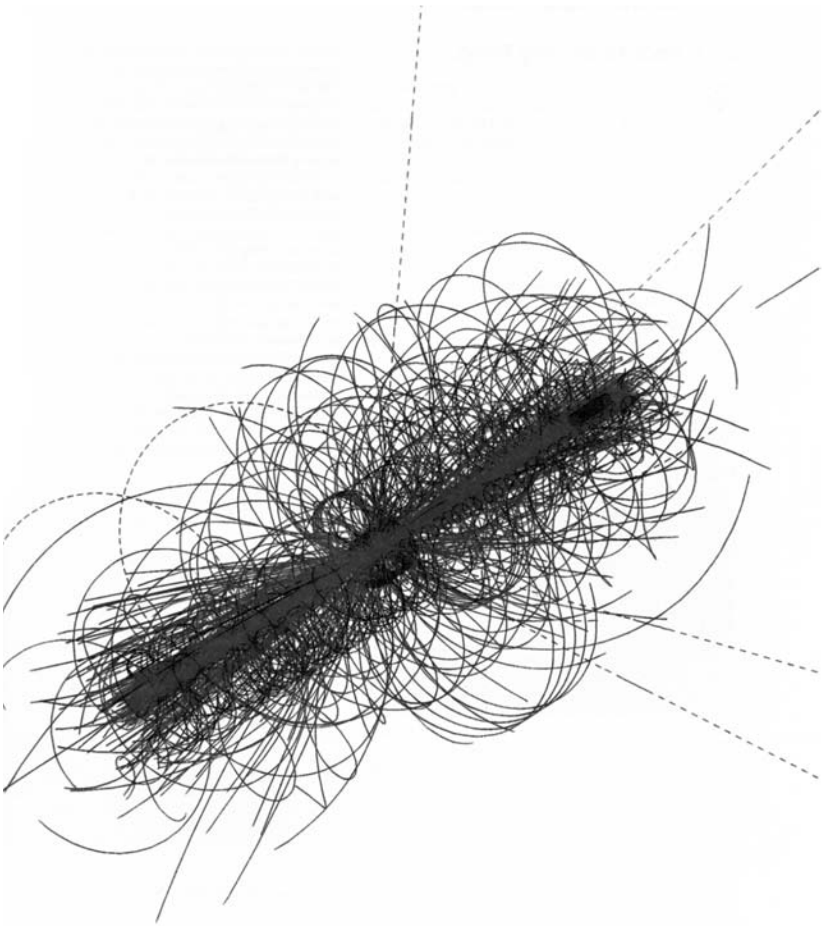
One fact we do know is that energy reigned supreme, vaporizing all but the smallest chunks of matter. Protons, neutrons, and electrons, as well as a veritable zoo of other submicroscopic particles were unable to cluster into more complex structures. No stars or planets existed at the time. Not even any atoms were tolerated. The environment was too energized, the Universe still too chaotic—the clear and frenzied aftermath of the biggest of all cosmic "bombs."

The basic stuff of the Universe continued to fly apart rapidly, cooling and thinning all the while. About a millisecond after the bang, the superhot and superdense conditions suitable for hadron creation had ceased, allowing a whole new class of particles such as electrons and neutrinos to come forth and dominate. Thus began another process of materialization whereby lightweight, weakly interacting particles—those called "leptons"—were fashioned from energy under an average density of ten billion grams per cubic centimeter and a temperature of about ten billion degrees Celsius. These physical conditions were still excessive by any earthly standards, but they had moderated greatly compared to the hugely dense and intensely hot values present a fraction of a second earlier. For, once the Universe began expanding, it did so extraordinarily rapidly, unhesitatingly dispersing its heat and its contents. By the time the first second had elapsed, leptons were being quickly made from radiation and many just as quickly destroyed back into radiation, much as had the hadrons earlier. In a kind of equilibrium between creation and destruction of subatomic particles, this cosmic fireball was still fueled by harsh radiation, such as X rays and gamma rays, as well as with (what we would now call) blinding light.

The density of radiation greatly exceeded the density of matter throughout these first few minutes. Not only did the photons of radiation far outnumber the particles of matter, but also most of the energy in the Universe was in the form of radiation, not matter. As soon as the elementary particles tried to combine into atoms, fierce radiation destroyed them. Structure, organization, and complexity did not yet exist; information content was minimal. Radiation was simply overwhelming, and for this reason much of the Particle Epoch is often called the Radiation Era. Whatever matter managed to exist at the time did so as a thin precipitate suspended in a glowing "fog" of dense, brilliant radiation.

As time elapsed, change continued. A few hundred years after the bang, the density had decreased to a value of about a billionth of a gram per cubic centimeter, while the average temperature had fallen to about a million degrees Celsius—values hardly different from those in the outer atmospheres of stars today. A principal feature of this latter part of the Particle Epoch was the steady waning of the original fireball; the annihilations of hadrons and leptons had ended. Even as the fireball faltered, though, a dramatic change began.

The first hundred centuries of the Universe saw radiation reign supreme. Radiation was in absolute and firm control, as all space was literally inundated with it. The cosmos remained a structureless and highly uniform blob; astrophysicists say that matter and radiation were intimately coupled to each other, in a sense equilibrated. As the Universe expanded, however, the radiation density decreased faster than the matter density. (That's because matter is diluted in proportion to the



Evidence of early Universe plasma.

By violently smashing the building blocks of matter and examining debris from the ensuing collisions, physicists infer knowledge about the basic forces and the structure of matter at very small scales and very high energies. Here, myriad particles emanate from the site where two gold nuclei collided, the result being "quark soup," which approximates the incredibly hot and dense conditions that likely prevailed within the first second of the Universe's existence. *Source: Brookhaven National Lab.*

volume increase, but radiation, being additionally affected by the Doppler effect, decreases more than with mere volume growth.) This imbalance ultimately caused the early sphere of blinding light to thin gradually, thus diminishing the early dominance of radiation. It was as though a luminous fog of energetic photons (like the gas inside a glowing neon sign) had begun to dim and then lift. Matter and radiation thereafter decoupled, their thermal equilibrium unraveling and their particle symmetry breaking as an evolutionary change of great importance began.

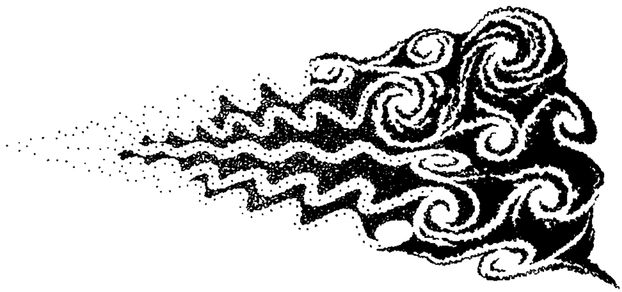
Sometime between the first few millennia and a million years after the bang—the exact moment cannot be pinned down much better, since the process was gradual—the charged elementary particles of matter began clustering into atoms. Their own electromagnetic forces pulled them together, sporadically at first and then more frequently. The weakening radiation could no longer break them apart as quickly as they combined. In effect, the authority of radiation had subsided as the previously charged matter (plasma) gradually became neutralized, a physical state over which radiation has little leverage. Matter had, in a sense, managed to overthrow the cosmic fireball while emerging as the principal constituent of the Universe. To denote this major turn of events, the latter portion of the Particle Epoch and all the remaining epochs that have occurred since are collectively termed the Matter Era.

Once the radiative fog dispersed and the Universe became nearly transparent, most photons traveled unhindered through space. Radiation had, in effect, become uncoupled from matter. And as the Universe expanded, that radiation simply cooled, eventually becoming the cosmic background radiation now perceived all around us, as described in the prologue. That last interaction of photons with matter occurred when the Universe was about a half-million years old. Thus, by observing the cosmic background radiation now, astronomers can probe conditions in the early Universe more than ninety-nine percent of the way back in time to the big bang.

The emergence of organized matter from chaotic radiation was the first of two preeminent changes in the history of the Universe. The Radiation Era had naturally and inevitably given way to the Matter Era. We shall not encounter the second of these truly fundamental changes—the rise of technological sentient life-forms—until the epilogue.

(Even if the mysterious "dark energy" noted at the end of the prologue does exist, indeed even if it now dominates the total cosmic density, its presence was not likely of much consequence in the early Universe. Dark energy, assuming it's real, is expected to grow in force and influence as the Universe expands. Only now, more than ten billion years later, would its large-scale effect have begun to manifest itself.)

The onset of the Matter Era saw the widespread creation of atoms; they were literally everywhere. The influence of radiation had grown so



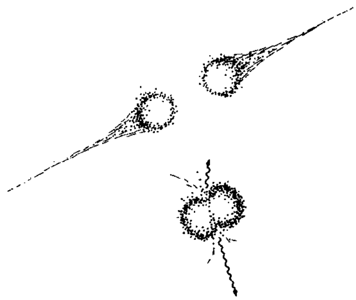
The Radiation Era naturally and inevitably gave way to the Matter Era.

weak that it could no longer prevent the attachment of hadron and lepton elementary particles that had survived annihilation. Hydrogen atoms were the first type of element to form, requiring only that a single negatively charged electron be electromagnetically linked to a single positively charged proton. Copious amounts of hydrogen were thereby made in the early Universe, and it is for this reason that we regard hydrogen as the common elemental ancestor of all material things.

Hydrogen (and its isotope, deuterium) was not the only kind of atom fashioned during the Particle Epoch. Before all the electrons and protons were swept up into hydrogen, atoms of the second simplest element, helium, began to form.

Heavy nuclei originate when two or more light nuclei fuse together. They do so by means of a dual process. First, a heavy nucleus of an atom is created whenever lighter ones collide violently enough to stick and fuse. Second, the newly formed positively charged nucleus then attracts a requisite number of negatively charged electrons, thereby yielding a neutral, albeit heavier, atom.

In the case of helium production, a temperature of at least ten million degrees Celsius is needed to thrust two hydrogen nuclei (protons) into one another. Each proton boasts a positive charge and at lower temperatures they would simply repel like identical poles of magnets. This minimum temperature ensures that the hydrogen nuclei collide with ample vigor to pierce the natural electromagnetic barrier that prevents them from interacting under ordinary circumstances. For a split second, the colliding particles enter the extremely small operating range of the powerful nuclear force. Once within about a trillionth of a centimeter of one another, the two hydrogen nuclei no longer repel. Instead, the attractive nuclear force seizes control, slamming them together ferociously and uniting them instantaneously into a heavier nucleus. Exactly the same process occurs now in the hearts of stars everywhere, as we shall see in the Stellar Epoch. And it's the same process that humans have made operational, though on a much smaller and uncontrolled scale, in the form of modern thermonuclear weapons, especially the hydrogen bomb.



. . . a heavy nucleus of an atom is created whenever lighter ones collide violently . . .

The superheat of the early Universe meant that the physical conditions were ripe for the creation of helium nuclei from protons of the primeval fireball. Thereafter, in the later stages of the Particle Epoch, pairs of electrons were electromagnetically attracted to each helium nucleus, thus fabricating neutral helium atoms. Given the rapid rate at which most models suggest the Universe expanded and cooled, only so much of the hydrogen could have been transformed into helium, leaving about a dozen hydrogen atoms for every one helium atom. That's a helium abundance of nearly ten percent by number, or twenty-five percent by mass.

By contrast, elements heavier than helium could not have been appreciably produced in the early Universe. (Nuclei of the third element—lithium—squeezed through the bottleneck, but only in smattering amounts fully a billion times less than helium.) The carbon-rich fibers composing this page of text, the oxygen and nitrogen in the air we breathe, as well as the copper and silver in the coins in our pockets were not made in the aftermath of the initial bang. Fusion of heavy elements, including all the way up to iron and uranium, for example, requires temperatures much higher than ten million degrees Celsius. Such syntheses also require lots of helium atoms. The trouble here is that even though helium production was in high gear during those first few years, both the density and temperature were quickly falling. Theoretical calculations suggest that, by the time there were sufficient helium atoms to interact with one another to manufacture heavier elements, the cosmic temperature had dipped below the threshold value needed for the mutual penetration of doubly charged helium nuclei. That threshold value is a hundred million degrees Celsius, for it takes even greater violence for multiply charged nuclei to collide, stick, and fuse. The Universe was still hot, but not quite hot enough anymore to make the heavies.

Contrary to the progressive cooling and thinning of the early Universe, the compact matter within stars, none of which had yet formed by that time, is perfectly suited for the generation of hotter temperatures, greater densities, more brutal collisions, and thus heavier elements. The hearts of stars, as examined later in the Stellar Epoch, are indeed where the heavies were eventually created, albeit long after the Particle Epoch had ended. It's where they are still being made today.

∞

An atom of ordinary matter is an invisible, submicroscopic entity comprising a positively charged heavyweight nucleus, usually several protons and neutrons, surrounded by one or more negatively charged lightweight electrons. All atoms found on Earth maintain this common structure—the essence of normal, baryonic matter. Furthermore, radiation received from extraterrestrial objects, near and far, is consistent with this same basic structure for all atoms everywhere.

Theorists nonetheless wonder about the possible existence of other kinds of atoms—not just additional elements yet undiscovered, but atoms built differently from the ones we know on Earth. How is it, for instance, that heavy nuclei always have a positive charge, relegating the negative charge to only the lightweight electrons? Some argue that the Universe would be more philosophically pleasing if its basic building blocks had more symmetry in their charge and mass. In other words, perhaps the Universe is also endowed with atoms made of negatively charged nuclei around which orbit positively charged particles.

Experimentalists did in fact discover, around the middle of the twentieth century, lightweight, electronlike particles having a positive charge.

These so-called antimatter particles are identical to ordinary matter particles in every way except charge. A particle called a positron, for example, has all the properties of an electron, except that its charge is positive. These same experiments also proved that when a matter particle and its antimatter opposite collide, the result is mutual annihilation and an explosion that releases pure energy of the lethal gamma-ray variety.

The reverse phenomenon can occur as well. Provided the temperature is extraordinarily great (in the range of billions of degrees Celsius), collisions among packets of gamma radiation can yield pairs of elementary particles, for instance, a matter electron and an antimatter positron. This sort of materialization (or "pair production") of matter and antimatter from energy still obeys the fundamental laws of physics; in this case, once again in accord with the formula, $E = mc^2$.

These and other kinds of conversion from energy to mass are precisely what theoretical models suggest happened in the earliest moments of the Universe. Yet we don't observe much antimatter around us now. Earth, the other planets, and the Sun all appear to be composed of ordinary matter. Exceptions include some particles produced in nuclear reactions known to be churning away inside stars, a small fraction of the baffling cosmic rays showering us each day, and minute fragments created during elementary-particle collisions in nuclear laboratories on Earth. Still, virtually all the mass in the Solar System seems to be of the matter variety, with little trace of naturally occurring antimatter. If matter and antimatter were created in equal amounts from primordial energy in the early Universe, then where has all the antimatter gone?

Note that antimatter does not imply antigravity. Particles of antimatter gravitationally attract one another just as do two or more particles of matter. The only property distinguishing matter from antimatter is charge; the mass of every matter particle is identical to that of its antimatter opposite, hence gravity invariably pulls while never pushing. Apart from the mysterious "dark energy," discussed in the prologue but not yet found, no such "antigravity" is known anywhere in the Universe.

Nothing, in principle, prohibits elementary particles of antimatter from combining into large clumps. Antihydrogen, antioxygen, anticarbon, and numerous other antiatoms could conceivably form antiplanets, antistars, and antigalaxies. The fact that we are unaware of such big groups of antimatter does not preclude their existence. Since atoms of antimatter emit and absorb precisely the same type of photons as do atoms of ordinary matter, astronomers have no way to determine if, for example, a distant star is made of matter or antimatter. Physicists like to say that a photon is identical to its own antiphoton. Radiation emitted by a clump of antimatter would equal that from a clump of matter; photons and antiphotons have no known differences. Accordingly, the nearby Alpha Centauri star system or the Andromeda Galaxy, for example, *could* be composed of antimatter—but it's doubtful.

Despite the fact that our Solar System is made mainly if not totally of matter, large pockets of antimatter may well exist elsewhere in the Universe. Provided clusters of matter remain separated from those of antimatter, then the two can coexist. As to where the primeval antimatter might be, we can only conjecture that it's wrapped up in large, distinct assemblages far outside the Solar System. Should similar matter and antimatter objects venture too closely together, however, they would mutually annihilate. Consequently, if our civilization ever attains an ability to travel beyond our Solar System, it will be important to dispatch automated probes before humans visit alien worlds. Should such an unmanned spacecraft suddenly evaporate in a puff of gamma radiation, we would be wise to visit elsewhere.

That said, scientists presently have no experimental evidence for macroscopic structures of antimatter beyond Earth. They are only theoretically inferred on the basis of symmetry arguments: the simplest cosmological models imply that equal quantities of matter and antimatter should have been created from energy in the Particle Epoch. We are not done with this dilemma.

∞

What about the very earliest moments of the Universe—those times well before the first second of existence had elapsed when all the forces of Nature are thought to have been merged into a single, grandly unified force that controlled everything? Going back even closer to the origin of all things, our quest to unify all the known forces has recently combined some aspects of the subjects of cosmology and particle physics. These efforts—including some of the most exotic work at the frontiers of science—have led to tentative advances toward a controversial "theory of everything."

What follows in this section is informed speculation, based on extensions of much better known phenomena akin to what we now witness in space, time, and energy. The closest time to the big bang that astronomers can observationally study physical phenomena directly is about a half-million years *after* the bang. This again is the cosmic background radiation that contains hints and clues regarding events in the earlier Universe. And the closest that physicists can experimentally study those earlier events is about 10^{-10} second after the bang—that's one-tenth of a nanosecond. These are laboratory simulations, done in quick bursts at the big accelerators, of the violence in a very young Universe impressively close in time to the bang, but currently no closer. Scientific descriptions of events earlier than a billionth of the first second of time are only reverse extrapolations—thought, at least by scientists, to be better than religious dogma, philosophical musing, or science fiction, but how much better is frankly unknown.

During the last quarter of the twentieth century, the electromagnetic force binding atoms and molecules and the weak nuclear force governing the decay of radioactive matter were merged into a single theory asserting them to be different manifestations of one and the same force—the "electroweak" force. Crucial parts of this theory have been confirmed at the world's most powerful accelerators at CERN and Fermi Lab, and concerted efforts are now under way to extend this unified theory to include the strong nuclear force that binds elementary particles within nuclei. Furthermore, though scientists are unsure at this time how, in turn, to incorporate into this comprehensive theory the fourth known force (gravity), we have reason to suspect that we are nearing the realization of Einstein's dream—understanding all the forces of Nature as different aspects of a single, fundamental force.

The intellectual synthesis of the macrodomain of cosmology (for gravity is a demonstrably long-range force) and the microdomain of particle physics is but a small part of the grand scenario of cosmic evolution. Yet it's an important part, for the newly emerging interdisciplinary specialty of "particle cosmology" could well provide great insight into a much earlier period of the Universe, namely, the time interval often colloquially labeled "chaos"—a temporal domain resembling the *terra incognita* parts bordering old maps of antiquity.

In brief, descriptive terms, this is the way the newly understood electroweak force operates: In submicroscopic (quantum) physics, forces between two elementary particles are exerted, or mediated, by the exchange of a generic particle, called a boson; in effect, the two particles can be imagined to be playing a rapid game of catch using a boson as a ball. In ordinary electromagnetism familiar to us on Earth, the boson is the usual photon, and for the strong nuclear force that boson is a gluon. Both types of bosons always travel at the velocity of light. The new electroweak theory includes four such bosons: the photon as well as three other subatomic particles with the innocuous names W^+, W^-, and Z^0. At temperatures less than a million billion (10^{15}) Celsius—the thermal range of absolutely all events on Earth and in the stars today—these four bosons split into two families: the photon that expresses the usual electromagnetic force and the other three that carry the weak force. But at temperatures greater than 10^{15} Celsius, these bosons work together in such a way as to make indistinguishable the weak and electromagnetic forces. Thus, by experimentally probing the behavior of this new force, we gain insight into not only the essence of Nature's building blocks but also some of the earliest periods of the Universe, especially the hadron period around 10^{-10} second after the bang.

This is where (or when) the experimental confirmations currently end, for humankind has not been able to build large enough accelerators capable of generating the even higher energies typifying the greater densities and temperatures prevalent at times closer to the big bang than 10^{-10} second. Even so, it's remarkable that science can manage to do that—to take that last demanding step in the scientific method and to test ideas thought pertinent to well less than the first second of time when anything at all may have existed.

To appreciate the nature of matter at temperatures exceeding 10^{15} Celsius, and thereby explore indirectly times even closer to the big bang, physicists are now researching a more general theory that merges the electroweak and the strong nuclear forces (but not yet gravity). Several versions of this so-called grand unified theory, dubbed GUT for short, have been proposed, though testing capable of determining which, if any, of these theories is correct has really only begun. Like the other known forces, this grand unified "superforce" is expected to be mediated by a boson elementary particle—for want of a better name, the X boson. It is, according to these unifying theories, the very massive (and thus very energetic) X bosons that are predicted to have played a vital role in the first instants of time.

For example, imagine a time equal to 10^{-39} second, when the temperature approximated 10^{30} Celsius. At that moment, only one type of force other than gravity operated—namely, the grand unified force just noted. According to the theory of such a superforce, the matter of the

Universe must have then exerted a huge pressure that pushed outward in all directions. (In classical terms, pressure is the product of density and temperature, so if, in the early Universe, each of these quantities was large, then the pressure must have been vast.) The Universe must have responded to this pressure by expanding and dropping its temperature. As time advanced from 10^{-39} to 10^{-35} second, say, the Universe grew by another couple of orders of magnitude and the temperature fell to about 10^{28} Celsius.

According to most grand unified theories, this temperature—10^{28} Celsius—is special, for at this value a dramatic change occurred in the expansion of the Universe. In short, when matter cooled beyond this temperature, the X bosons could no longer be produced; at times after 10^{-35} second, the energy needed to create such particles was too dispersed because of the diminishing temperature. So as the temperature fell below 10^{28} Celsius, the disappearance of the X bosons is thought to have caused a surge of energy roughly like that released as latent heat when water freezes (an event that often contributes to the bursting of closed containers.) After all, the energy that was no longer concentrated enough to yield X bosons was nonetheless available to enhance the general expansion of the Universe—in fact, to cause it to expand violently, or "burst," for a short duration just after the demise of the bosons.

The youthful Universe, though incredibly hot at the time, was quite definitely cooling and thus experienced a series of such "freezings" while passing progressively toward cooler states of being. As perhaps the most impressive of all such transitions, the rapid decay of the X bosons caused a tremendous acceleration in the rate of expansion. This period of exponentially fast expansion has been popularly termed "inflation." Each tiny patch of space doubled in size at least a hundred times, such inflation enlarging the Universe's volume from a trillionth that of a proton to that of an acorn—a huge difference. In well less than the blink of an eye represented by a mere 10^{-35} second, the Universe inflated some 10^{50} times, smoothing out (by stretching) any irregularities existing at the outset, much as crinkles on a balloon vanish as it's inflated. This is why the Universe seems so accurately described by flat, Euclidean geometry, despite all the curving and warping of spacetime near massive objects. We apparently now see only a tiny part of the whole Universe, and that part seems flat to us, much as an ant on the surface of that rapidly expanding balloon would see less and less of it while it seemingly grew flatter.

At the conclusion of the inflationary phase, at about 10^{-35} second, the X bosons had disappeared forever and with them the grand unified force. In its place were the electroweak and strong nuclear forces that operate around us in the more familiar, lower-temperature Universe of today. Physicists describe such an event as "broken symmetry," with the strong and electroweak forces, previously one, having then become separate entities. With these new forces in control (along with gravity), the Universe resumed its more leisurely expansion. Later, at around 10^{-10} second, when the cosmic temperature had decreased to some 10^{15} Celsius, a second symmetry breaking occurred, enabling the electroweak force to reveal its more familiar electromagnetic and weak natures, which guide almost everything we currently know about on Earth and in the stars.

Can we test this unified force idea, including its implied and spectacular inflationary phase change? The answer is a qualified yes, for we can do so only indirectly. After all, even the biggest accelerators on Earth are barely able to create, and then only for the briefest of instants, conditions approximating 10^{15} Celsius, sufficient to confirm the electroweak theory. By contrast, the grand unified theories become operative at much higher temperatures, in fact greater than 10^{28} Celsius, which physicists will likely never be able to simulate on Earth. To boost subatomic particles to the absolutely immense energies needed to test the grand unified theories would require a particle accelerator spanning the distance between Earth and the Alpha Centauri star system some four light-years away—a truly cosmic machine that would require for *each second* of operation an altogether unreasonable expenditure of power equal to the annual U.S. gross national product! So, while we have successfully mimicked in the laboratory the physical conditions characterizing the lepton period (approximately 10^{-6} second) and parts of the earlier hadron period (approximately 10^{-10} second), scientists have concluded that the earliest chaos period is likely to remain forevermore "too hot to handle."

One especially attractive aspect of the grand unified theories is that they seem able to account for the observed excess of matter over antimatter and thus potentially solve that dilemma noted earlier. It so happens that the decay of the X bosons at times earlier than 10^{-35} second lacked symmetry; their decay is expected to have created slightly greater numbers of protons than antiprotons (or electrons than positrons). Specifically, calculations suggest that for every billion antiprotons (or positrons), a billion and one (i.e., $10^9 + 1$) protons (or electrons) were created. The billion matched pairs subsequently annihilated each other, leaving a residue of ordinary matter from which everything—including ourselves—emerged. If this imbalance is true—another example of broken symmetry—then the matter extant today is just a tiny fraction of that formed originally.

This prediction can be tested in a straightforward way, for if protons can be created they can also be destroyed. Protons might not be the immortal building blocks once thought, and the grand unified theories can be used to estimate the proton's average life expectancy. That lifetime turns out to be 10^{32} years—a hundred quadrillion quadrillion years—which is much greater than the age of the Universe! This extremely long lifetime guarantees that although all matter might ultimately be destined to disappear, the probability of decay in any given time span is exceedingly small. Nonetheless, given that Nature is largely governed by statistical physics, any one proton is in danger of decaying at any one moment. In fact, since water is an abundant source of protons, theory predicts that roughly one proton should decay per year in each ton of water. Alternatively expressed, a typical human body is expected to lose only about a single proton in an entire human lifetime. Experiments are now in progress attempting to detect such events in huge water tanks stored in deep underground mines at several places on Earth (thus shielding them from spurious effects triggered by cosmic rays reaching Earth's surface from outer space). Furthermore, a statistical measurement of a proton's lifetime should enable us to discriminate among the various grand unified theories, further refining our "approximations of reality." Alas, the simplest of these theories has apparently been ruled out, as no proton decays have been found in the last few years in several tons of water. Perhaps protons, like diamonds, are forever, and it's Nature that's not so simple. Some physicists take this as a bad sign, for the history of science has taught us that, more often than not, theoretical complications usually indicate that we are on the wrong track.

An intriguing cosmological implication of the inflationary concept is that, if correct, it must have put the Universe into a state precariously balanced between infinite expansion and ultimate collapse. Recall from the prologue that for this to happen, the correct model is one for which its density equals precisely the critical density—namely, the case for which its accumulated gravitational effect exactly offsets its rate of expansion and the resulting geometry is flat. Since astronomers have observationally demonstrated that the density of normal, baryonic matter is only a few percent of this critical density, we surmise that more than ninety-five percent of the Universe is made not of normal matter but of some unorthodox, dark-matter form such as massive neutrinos and exotic particles (black holes won't do it, as they are made of normal matter), or of some entirely new kind of energy not yet known in physics.

What about even earlier phases of this, the earliest of all periods ("chaos"), at times prior to 10^{-35} second? Can we probe, even theoretically, any closer to the start of all things at the celebrated "t = 0" mark? Efforts are currently hampered because doing so requires the gravitational force to be incorporated into the correct grand unified theory. Yet no one has managed to develop a self-consistent, super–grand unified theory (or super-GUT), as this is tantamount to inventing a quantum theory of gravity—a towering intellectual achievement that would ostensibly merge Heisenberg's uncertainty principle (which guides submicroscopic phenomena) and Einstein's relativity theory (which describes macroscopic scales). Whoever does achieve this holy grail of physics gets a free, all-expense-paid trip to Stockholm, courtesy of the Nobel committee.

Our current knowledge of the strong gravitational force implies that such quantum effects very likely become important whenever the Universe is even more energetic than we have yet contemplated. Specifically, at a time earlier than 10^{-43} second (known as the "Planck time," after Max Planck, one of the creators of quantum theory), when the temperature exceeded 10^{32} Celsius, the four known basic forces are thought to have been one—a truly fundamental force operating at energies prevalent during the earliest parts of the chaos period. There and then, with all the matter in the Universe theorized to have been unimaginably compacted and a trillion trillion times hotter than the core of a hydrogen-bomb explosion, the curvature of (Einsteinian) spacetime and the dimension of (Heisenbergian) uncertainty both equal 10^{-33} centimeter (the "Planck length"), inside which relativity theory is no longer an adequate description of Nature. Only at lesser energies (i.e., at times after 10^{-43} second) would the more familiar four forces begin to manifest themselves distinctly, though in reality all four are merely different aspects of the single, fundamental, supergrand force that ruled at or near the big bang.

In a potentially related advance, attempts to understand force unification have driven theorists toward the fascinating concept of "supersymmetry." This extends the idea of symmetry among forces to particles. Accordingly, all elementary particles are reasoned to have so-called supersymmetric partners—exotic particles (sometimes called "sparticles") that exist alongside their normal counterparts readily observed in the everyday world of our human senses. Of particular interest to astronomers, these particles would have been produced in great abundance in the early aftermath of the big bang and should still be around today. Since they are thought to be very massive (at least a hundred times that of a proton), these supersymmetric relics are among the leading candidates for dark matter within and beyond the galaxies. However, none of these suspected particles has yet been experimentally detected, so the theory's validity remains uncertain.

Ironically, with the physicists unable to build equipment on Earth sufficiently energetic to reproduce cosmic chaos, and thus perhaps to recreate in the lab some of the bizarre elementary particles likely created in the very early Universe, it's the astronomers who, by studying the macrorealm, are beginning to provide tests, albeit indirect ones, of the grand unification of the microrealm. This is another example of how interdisciplinary efforts are so richly rewarding, in this case the newly emerging subject of particle cosmology bringing together the very largest and very smallest scales in Nature.

In another possibly important development noted earlier, some physicists have recently become enamored of a radical idea originally proposed several decades ago. Variously called "string theory," "superstrings," or, mysteriously, "M-theory," this idea aspires to unite all the laws of physics into a single mathematical framework, perhaps even to discover one equation less than a few centimeters long that can explain all things—the so-called theory of everything! New and provocative terms—such as strings, curls, and membranes—derive from the notion that the ultimate building blocks of Nature might not be pointlike particles at all, but tiny, vibrating, extended objects. If this novel view is correct, it means that the protons and neutrons in all matter, from our bodies to the farthest star, are fundamentally made of strings or superstrings shaped like loops. Alas, no one has ever seen, or has much prospect of seeing, such strings since they are predicted to be more than a billion billion times smaller than a proton—in fact, 10^{-33} centimeter, again the Planck length. Depending on the mode of vibration, separate particles of matter can be made from such subatomic strings, much the way violin strings can resonate with different frequencies, each one creating a separate tone of the musical scale. Disconcertingly, the theory of superstrings works only if the Universe began with (usually) eleven dimensions of spacetime, seven of which (somehow) collapsed or otherwise became "hidden" near the time of the big bang. To some physicists, such a revolutionary idea borders on science fiction (or even religion), whereas for others, it possesses breathtaking mathematical elegance and perhaps the best hope of avoiding a whole host of thorny problems on the road to quantum gravity. Regardless, the science journals are littered with mathematically beautiful theories that apparently have no basis in physical reality. And although the theory of superstrings is now causing great excitement in the physics community, to date not a shred of experimental or observational evidence supports it.

Any theory purporting to penetrate even closer to the very beginning of time is currently hardly more than conjecture. Given our current knowledge of physics, it makes little scientific sense to talk about times earlier than 10^{-43} second. Time intervals smaller than this are not yet part of the lexicon of science, and notions of space and time earlier than this border on the meaningless.

That said, many researchers suspect that once a proper theory of quantum gravity is in hand, our understanding might automatically include a *natural* description of the original creation event itself. It's even conceivable that the primal energy emerged at zero time from essentially nothing, uncannily in accord with the structureless singularity described by the time-honored poetic expression, "without form and void, with darkness upon the face of the deep." Even in a perfect vacuum—a region of space containing neither matter nor energy—particle-antiparticle pairs (such as an electron and its antiparticle opposite, the positron) constantly appear and disappear in time spans too short to observe. Although it would seem impossible that a particle could materialize from nothing—not even from energy—it so happens that no laws of physics are violated because the particle is annihilated by its corresponding antiparticle before either one can be detected. Furthermore, for such events not to happen would violate quantum physics, which cites, via Heisenberg's principle again, the impossibility of determining the exact energy content of a system at every moment in time. Hence, natural, quantum fluctuations of energy must occur in empty space, *even when the average energy present is zero.*

In this way, the Universe may well have been a case of *creatio ex nihilo* by means of an energy change that lasted for an unimaginably short duration—a "self-creating Universe" that erupted into existence spontaneously, the result of a random quantum fluctuation! The net energy was then, is now, and forever shall be zero; all of gravity and its myriad negative, attractive, potential energies would perfectly balance all other known positive energies (including heat, light, mass, and so on), making our vast Universe seem like "something for nothing," yet it really isn't. The Universe arose from a quantum fluctuation large enough that energy, matter, time, and space all sprang into being. Could this be the solution to the time-honored philosophical query, "Why is there something rather than nothing?"—the answer being, ostensibly, that the probability is greater that "something" rather than "nothing" will happen. This sort of "statistical" birth of the Universe from a kind of nothingness has been sacrilegiously dubbed "the ultimate free lunch"—an extreme manifestation of the longstanding quip that Nature abhors a vacuum.

That some of these latter ideas are speculative is putting it mildly. Skeptics would say that they are not real science at all, for they violate one of the central tenets of the modern scientific method: many of these concepts are virtually impossible to test experimentally. But they do illustrate how the world of science has itself changed at the start of the new millennium, as its scope now encompasses, for the first time, a model for the very origin of the origin. Might this be the beginning of a meaningful merger of science and religion into a truly profound interdiscipline, or might it signal renewed warfare between these two great institutions, as science treads on sacred turf where it's not quite gone before?

Should this quantum scenario, or some revised version of it, prove to be a correct description of the birth of the Universe, then our Galaxy, our Sun, our Earth, and ourselves are a direct consequence of a series of random events, albeit ones obeying physical laws, that occurred during an unimaginably short period of time some fourteen billion years ago. Even if not a valid understanding of the ultimate creation event itself, such submicroscopic fluctuations in density, enhanced by inflation and thereafter guided by expansion, might well have eventually grown into today's large-scale, macroscopic structures encountered throughout the remaining epochs of this book. Clearly, the development of a quantum-gravitational description of events at literally the origin of time, none of which attempts has thus far met with much consensus, is the foremost challenge in the subject of physics today.

∞

By the end of the Particle Epoch, the Universe had evolved dramatically. The spectacularly bright fireball identified as the origin of all things had subsided. Energy, which drives all changes in the Universe, had itself changed with its dispersal over time. The physical conditions of temperature and density had undergone extraordinary change. Atoms, mainly hydrogen and helium, had been synthesized. And matter had wrested firm control from radiation, heralding a whole new era.

Major events in the Universe would thereafter occur less frequently. Change continued, though at a more relaxed pace. Key transactions between matter and radiation may well have occurred posthaste immediately after the bang, and especially in the first few minutes of the Universe. But these interactions eventually lessened, becoming few and far between by the end of the Particle Epoch. The average density fell enormously throughout this initial epoch, plummeting below a billionth of a billionth of a gram per cubic centimeter before the epoch had ended—less than a million years after the Universe began. The average temperature of the cosmos had also dropped to a relatively cool thousand degrees Celsius, a pale, sluggish remnant of the intense heat prevalent at creation.

With time, the Universe had grown thinner, colder, and darker. It was destined to evolve much more slowly in later epochs, but it evolved nonetheless. The average physical conditions were on their way to becoming a billion times still less dense and a thousand times still less hot—tenuous and frigid conditions now present more than ten billion years after the bang—the fossilized grandeur of a bygone era.

The history of the early Universe presented here represents the prevailing view among cosmologists. Most share this general outline, though consensus is lacking regarding the fine details. Scientists agree on events as far back as the first nanosecond of existence, but the earlier we explore beyond that, the more unsure our statements become. Accordingly, the temperature and density in the first instants of the Universe are quite obscure, mainly because their values depend upon poorly understood interactions among the heaviest elementary particles at some of the highest conceivable energies. This uncertainty should not surprise us, for the primordial moments of the Universe are long gone with cosmic expansion, forever lost to the march of time. We can fathom the most ancient realms of Nature only indirectly, aided by crutches of abstract formulae and logical symbols.

What is surprising is that science can address any of this at all, modeling times and events that are very much over and done, perhaps never to occur again. And what we find, in virtually all models that are based on real data, is an early Universe reasoned to have been exceedingly hot and dense, growing cooler and thinner with time, and basically changing in ways to set the stage for the successive appearance of galaxies, stars, planets, and life.

2. GALACTIC EPOCH

Hierarchy of Structures

DESCENDANTS OF OUR CIVILIZATION MAY never become advanced enough to journey far enough from our Milky Way Galaxy to look back and witness the full grandeur of our extended home in space; the finite speed of light is too limiting, the Galaxy too vast. A literal picture of our resident swarm of a hundred billion stars floating proudly and silently in the void of space may forever elude us. Yet, from our Earth-based vantage point in the suburbs of our Galaxy—nearly thirty thousand light-years from its hub—astronomers study the variety and spread of other colossal star systems well beyond our own Milky Way. Many of these distant galaxies launched their light—some of the very light in tonight's nighttime sky—long before Earth and the Sun even emerged in the firmament.

Deep space harbors myriad objects looking strangely unlike stars. Photographic exposures, taken with even small, backyard telescopes, often reveal fuzzy, lens-shaped images resembling disks more than the bright, round points typical of stars. The eighteenth-century German philosopher Immanuel Kant regarded these blurry blobs of light, like so many flattened yet luminous puffs of cotton, as individual "island universes" far outside the confines of our Milky Way. We now know that labeling each of them a universe—"the totality of all things"—presents a clear semantics problem, but he was correct in thinking that these peculiar patches of light reside way beyond the well-known stars that constitute the familiar constellations.

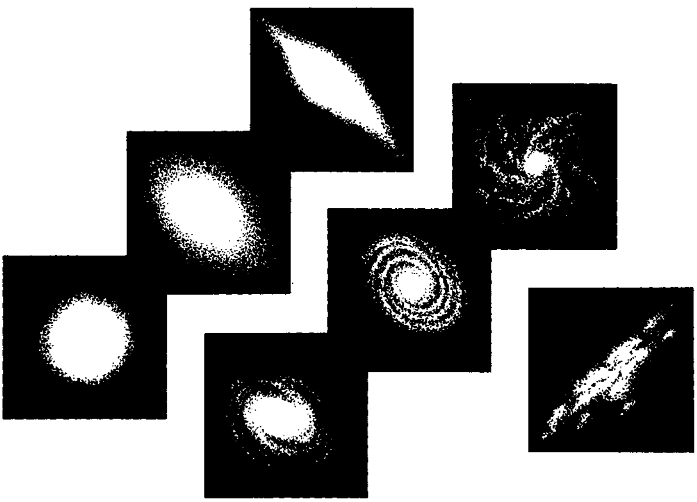
Large, modern telescopes have since revealed these remote beacons to be entire galaxies, each a huge collection of matter comparable to our Milky Way, measuring some hundred thousand light-years across, or a billion billion kilometers. Replete with hundreds of billions of stars bound loosely by gravity, each galaxy harbors more stars than all the people who have ever lived on Earth. Silently and majestically, galaxies twirl in the deep reaches of the Universe—vast hordes of radiation, matter, and perhaps life—simultaneously granting us a feeling both for the immensity of the Universe and for the minuteness of our position in it.

That position, when internalized, often resembles a boat adrift at sea. For there are as many galaxies in the Universe as there are stars in our Galaxy—all told, probably as many stars in the observable cosmos as grains of sand on all the beaches of the world—some 10^{22}, to be numerical about it. Yet organized patterns abound—grand dynamical processions like the pinwheeling of individual galaxies and the recessional of many more galaxies—provided we are willing to ponder the big and the broad.

Objects identified as galaxies in photographs often display spiral shapes much like our Milky Way or the neighboring Andromeda Galaxy. (Andromeda is the one distant galaxy that our naked eyes can see, perhaps best with averted vision, as a faintly glowing oval amid the constellation of the same name.) Each has a central bulge from which sport thin spiral regions, or "arms," chock-full of stars. The apparent prevalence of spiral galaxies is just that—apparent—largely the result of whorled patterns easily noticed among the many other patches of light in the nighttime sky. In reality, galaxies have an array of morphologies, and spirals are not the most common type of galaxy in the Universe.

After decades of effort in the mid-twentieth century led by the American astronomer Edwin Hubble, inventories of extragalactic matter beyond our Milky Way are now nearly complete for that relatively local part of the Universe in which we live. At least for normal, baryonic matter, that is, for we are still unsure what to make of the issue of "dark matter" raised in the Particle Epoch as strongly inferred but not yet seen. The most abundant galaxies are shaped like footballs and officially called elliptical galaxies. Some are less elongated, akin more to very large beach balls. Others resemble fat cigars and even thin cigars, but that's probably due to their tilted perspective. None of the ellipticals exhibits internal structure of any kind, and they are notably devoid of any spiral arms. Regardless of their shape, most elliptical galaxies harbor the usual hundred billion or so stars, typically spread across a hundred-thousand-light-year domain. A minority of them measure up to ten times that size and contain trillions of stars, but otherwise elliptical galaxies are undistinguished, albeit monumental, throngs of stars.

Spiral arms are not the only trait that elliptical galaxies lack. Hardly any cool gas drifts among the stars of these galaxies; they have little or no interstellar matter. This implies that all the elliptical galaxies are old. Stars, which originate from interstellar matter, apparently did so in these galaxies long ago, leaving hardly any loose gas for the continued formation of future generations of stars. Analysis of the radiation emitted by individual stars within the ellipticals further proves that those stars are also old. Evidently, nearly all their interstellar matter was used up eons ago, thus quenching the star-formation process.



Silently and majestically, galaxies twirl in the vast reaches of the Universe . . .

The infrequence of star formation in elliptical galaxies contrasts sharply with the abundance and activity of interstellar matter within spiral galaxies. Here, there are also a variety of shapes, though all the spirals are basically flattened disks resembling double sombreros clapped brim to brim. Some spiral galaxies have a large central bulge, mostly made of intact stars and diffuse gas, around which the spiral arms are tightly wrapped. Others have a more open pattern of arms emanating from an intermediate-sized central region. Still others have a rather small center from which long, stringy arms protrude, often making it hard to recognize these as spirals.

Spiral galaxies are known to contain lots of gas, and dust, too, mixed throughout the vast spaces among their stars. The oldest stars extend into the galaxies' spherical halos, but the youngest stars are found exclusively in the thin disks. Furthermore, observations during the past few decades have shown that stars are still forming, most of them in the arms. Unlike the old elliptical galaxies, spiral galaxies have a good deal of vitality—which doesn't necessarily mean that the spirals are young. Rather, they are simply still rich enough in interstellar gas to provide for ongoing stellar birth.

Some spiral galaxies sport a peculiar feature that has astronomers puzzled—namely, a linear "bar" of stellar and interstellar matter passing through their midsections. For these so-called barred spirals, the arms stem from near the ends of bar rather than from the central bulge itself. The puzzle concerns the way the bars form, evolve, or maintain themselves. Even our own Milky Way is now judged to perhaps have a small one passing through its galactic nucleus.

A final, catchall class of galaxies groups all those termed irregular galaxies, the type most widely seen in the Universe. These are oddly shaped structures of stars, gas, and dust whose visual appearance prevents their placement in any of the other categories. The irregulars clearly have much interstellar matter yet no organized structure, spiral arms, or central bulge. By and large, irregulars tend to be a bit smaller than other types of galaxies, so some astronomers call them dwarf galaxies. They often do seem to be dominated by the larger spiral or elliptical galaxies near which they are usually found. In fact, irregular galaxies seldom exist alone in space; they are mostly allied with larger "parent" galaxies of the spiral or elliptical variety.

Their proximity to the big galaxies is probably telling us something: The irregulars might be severely distorted regular galaxies that have experienced close encounters, and thus great tidal disruptions, with their parent galaxies. Or, they might be leftover building blocks of the larger galaxies into which they have not yet fallen. Some observations do hint at possible bridges of hydrogen gas connecting parent and irregular galaxies, suggestive of interactions between them. We shall return to these issues when later discussing mergers and acquisitions among galaxies rife with evolutionary change.

Our Milky Way Galaxy has a few small, companion irregular galaxies, most notably the Large and Small Magellanic Clouds, so named for the sixteenth-century Portuguese voyager Ferdinand Magellan whose round-the-world expedition first brought word of these great fuzzy patches of light to Europeans living in the Northern Hemisphere. Resembling dimly luminous atmospheric clouds and seen easily with the naked eye, they can be viewed only from locations south of Earth's equator, making them spectacular targets for first-time northerners traveling south, though they have undoubtedly served as celestial wonders to residents of the Southern Hemisphere since the dawn of civilization. Though one is slightly larger than the other, each is roughly a hundred times smaller than our own Milky Way system—meaning that they house "only" a billion or so stars—and both reside not quite two hundred thousand light-years away. The Magellanic Clouds probably orbit our Galaxy, just as Earth orbits the Sun or the Moon orbits Earth. The periods of these irregular galaxies are long by human standards, however, and their orbital paths have not yet been firmly established.

Actually, these famous celestial objects (at least to residents "down under") might not deserve the term galaxy at all, not even irregular galaxy. Though they do contain about a billion solar masses and do measure some ten thousand light-years across, they reside at a distance from our Galaxy that is only fifty percent again of its disk's typical extent. Thus, if our Milky Way system does harbor matter (dark or otherwise) in an extended halo—as many astronomers now suspect—then the Magellanic Clouds may well be nothing more than rich regions of star formation in the halo of our own Galaxy. Perhaps all such dwarf structures now classified as irregular galaxies will turn out to be residents of the outer realms of larger, well-categorized galaxies—and therefore not genuine galaxies at all.

∞

Planet Earth is finite; beyond it stretches the scant flimsiness of interplanetary space. Our Solar System is also finite; beyond it lies the near vacuum of interstellar space. And our Galaxy, in turn, is finite; beyond it exists the absolute material void of intergalactic space. Perhaps beyond even that, then, the arrangement of galaxies in space is also finite. Which brings to mind the obvious question: How are the galaxies spread throughout the expansive tracts far from the Milky Way? Is there some boundary, or terminus, beyond which galaxies are no longer seen? Or do they reside everywhere, all the way out to the limits of the observable Universe?

Within the "neighboring" realm of a few million light-years, astronomers know of a few dozen galaxies. Giant spirals, such as our own Milky Way and Andromeda, are found among small ellipticals and many irregulars, such as the Magellanic Clouds. Surprisingly, some nearby galaxies have been discovered even as recently as the past decade, such as the Sagittarius Spheroid, a newly found dwarf galaxy only eighty thousand light-years distant yet mostly obscured by our Galaxy's central bulge. Evidently, these two-score galaxies are bound together by their own mutual gravitational attraction—a mammoth version of the same natural phenomenon that holds stars in galaxies, planets around stars, and people on Earth. In all, these "local" galaxies are clustered within a volume whose diameter is some five million light-years. Including our Milky Way, the whole bunch is known as the Local Group. It constitutes our extended neighborhood in space.

Several million light-years comprise a significant chunk of cosmic real estate. Do note two important things about it. First, we have suddenly made a large jump in spatial dimensions, from the hundred-thousand-light-year size of our Milky Way to this five-million-light-year size of our Local Group. Galaxy clusters represent a distinctly higher level of hierarchical structure in the Universe—structure well beyond that of individual galaxies. Second, the Milky Way doesn't lie at the center of this cluster of galaxies. Not only is Earth not the center of our Solar System and the Sun not the center of our Galaxy, but our Galaxy is also not the center of the much larger Local Group. Though we might like to think so, humankind is not at any special, unique, or privileged location in the gargantuan, perhaps infinite, Universe.

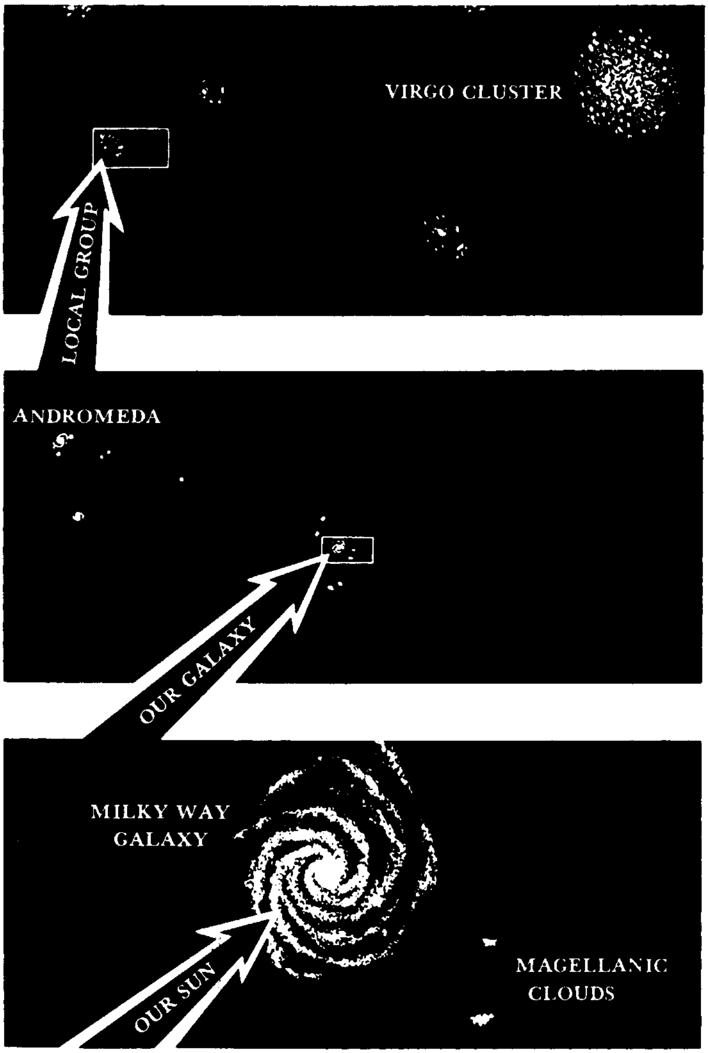
Many more than a few dozen galaxies reside in the Universe. Time exposures made with large telescopes reveal thousands of galaxies within virtually any small field of view. In all, astronomers estimate that some forty billion other galaxies inhabit the observable Universe. And virtually all of them are much farther away than even the distant members of our local galaxy cluster. For millions of light-years beyond the edge of the Local Group, there appears to be nothing—no galaxies, no stars, no gas or dust—just empty intergalactic space.

Strive to appreciate the far recesses of deep space outside the Local Group. Searching a seemingly interminable void, we occasionally sight a "field" galaxy scattered lonely here and there. Not until we reach a distance of some sixty million light-years away do we find another galaxy cluster, an unmistakable volume of space brimming with galaxies. This cluster is especially rich, containing not just forty galaxies as does our own Local Group; the so-called Virgo Cluster harbors nearly three thousand galaxies. Try to visualize in mind's eye thousands of individual galaxies all clustered in a swarm, each one housing about a hundred billion stars. No wonder most people have trouble appreciating the immensity of matter, space, and time in the Universe. Astrophysicists are no different; we, too, share the burden of trying to fathom such humongous sizes and scales, including astronomical numbers of astronomical objects.

Galaxy clusters like these populate the Universe throughout. They are not figments of our imagination. Their existence is fact, as hundreds have now been mapped and cataloged through numerous observations. In much the same way that galaxies are collections of stars, galaxy clusters are collections of galaxies. And beyond them, in turn, galaxy superclusters (or clusters of clusters) apparently also exist on colossal scales of typically hundreds of millions of light-years across. Both the Local Group and the Virgo Cluster are mere members of such a larger system—perhaps. These truly titanic structures occupy a most lofty rung—the greatest established to date—in the hierarchy of material assemblies within the Universe: particles, atoms, molecules, dust, planets, stars, galaxies, galaxy clusters, and now galaxy superclusters.

When contemplating the congested confines of rich galaxy clusters—such as Virgo, with its thousands of members, or the appropriately named Hercules Cluster, with its estimated hundred thousand galaxies—it's hard to avoid the impression that galactic traffic jams must be common. Just as atoms collide when confined in a closed container or hockey players in a enclosed rink, the random motions of galaxies within a galaxy cluster could conceivably induce phenomenal collisions among these huge material constructs.

Galaxies do indeed collide. A good deal of observational evidence proves that they do so, and quite often. Numerous celestial images show



. . . humankind is not at any special or unique location in the gargantuan, perhaps infinite, Universe.

two or more galaxies interacting, some of them tearing each other apart. While, in many photographs, galaxies lie along the same line of sight yet are really far separated in space, others are physically near one another, especially those within the galaxy clusters. Whether galaxies are colliding head-on or only experiencing close encounters cannot often be easily determined, for detectable motions among the distant galaxies typically take millions of years—which is why no human has ever witnessed the full panoply of a galaxy collision, as much as we note its effects virtually frozen in time.

At first thought, collisions among giant galaxies might be expected to create a mind-boggling crunching of matter, complete with spectacular explosions and superlative fireworks. Surprisingly, that doesn't happen much at all. Such collisions, in fact, are rather quiescent. The stars in each galaxy more or less just glide past one another as the two galaxies slide through each other. That's because stars themselves hardly ever collide; they are, after all, small objects by cosmic standards. While astronomers have plenty of direct photographic evidence for galaxy collisions, no one has ever witnessed or imaged a collision between two stars—not even in our own Milky Way, which we can see more closely and clearly.

This oddity occurs because galaxies within most clusters are bunched fairly tightly. The distance between adjacent galaxies in a given cluster averages a million light-years, which is only about ten times greater than the size of a typical galaxy. This doesn't really give them much room to

Evidence of galaxies colliding.

Like majestic ships passing in the night, these two spiral galaxies (called NGC 2207 and IC 2163) are experiencing a close encounter. They might eventually even suffer a head-on collision. Representative of many such galactic collisions seen in deep space, scenes like this one are common in the Universe. In roughly a billion years, after several more interactions, these spiral galaxies will probably merge into a single, colossal, elliptical galaxy. *Source: Space Telescope Science Institute.*

roam around without crashing. By contrast, stars within a galaxy are spread out much more thinly. The average distance between stars in a galaxy is roughly five light-years, millions of times greater than the size of a typical star. Said another way, if our Sun were the size of an apple, its closest neighbors would be some two thousand kilometers away. Hence, stellar collisions are extremely rare within any one galaxy, with the possible exception of their central regions. When two galaxies collide, the population density of stars merely doubles, leaving ample space for the stars to meander without sustaining much damage. To be sure, the interstellar contents, and perhaps the stars as well, of each galaxy are likely rearranged by the tidal forces induced by gravitational interactions, but no spectacular explosions result, even if the collision is head-on.

That's not to say that the pushing, shoving, and shocking of the interstellar gas doesn't cause any change. In fact, among the loose gas they wreak relative havoc! Bursts of star formation erupt in interacting galaxies like hurricanes in a pas de deux. In recent years, astronomers have imaged numerous "starburst galaxies," where the internal gas of colliding galaxies has been disturbed and rearranged enough to trigger sudden episodes of new stars in the disks of both. Additionally, already formed stars appear agitated, oddly orbiting like frenzied moths around a lamp, while other stars seem to be ejected along streamers stretching as much as a hundred thousand light-years from the site of the collision. So, although dramatic fireworks following direct hits among stars are most unlikely, computer simulations do show that the ensuing commotion causes the galaxies to glow about fifty percent brighter for a hundred million years or so owing to their many newborn stars. Later in this Galactic Epoch, we shall address the evolutionary implications of mergers and acquisitions as galaxies collide, mutually attract, and agglomerate.

Completing our inventory of the large-scale spread of matter in space, we naturally pose the next obvious question: Are there even greater assemblies of matter in the Universe, or do galaxy superclusters top the cosmic hierarchy? At present, astronomers are unsure. Some data imply clusterings of galaxy superclusters—or at least nonrandom arrays of galaxies on the largest scales yet observed—but this evidence is shaky and subject to debate. If correct, though, this would mean that our Local Group along with several other galaxy clusters embody a galaxy supercluster centered near the rich Virgo Cluster and that, in turn, these tens of thousands of galaxies form part of an even larger structure on the order of several hundred million light-years in diameter—which is more than a thousand times larger than the size of the Milky Way itself.

What is most clear from the latest cosmic maps of matter on the largest scales probed thus far are the irregularities: galaxies seem to be arranged in networks of filaments, or sheets, surrounded by relatively empty regions of space known as voids. A colossal sponge might be a good visual image, or perhaps an immense bubble bath. The so-called Great Wall, a lengthy arc of several thousand galaxies extending some half-billion light-years across the sky nearly three hundred million light-years away, is the nearest and most prominent of these giant features. An even bigger "wall," sporting nearly a hundred thousand galaxies about one-and-a-half billion light-years long and about a billion light-years away, is currently the largest known structure in the Universe. The bright galaxies' locations resemble spider webs or the neural structure of the human brain, whereas the dark voids, often measuring hundreds of millions of light-years wide, are almost completely absent of any galaxies. The most likely interpretation of these maps—the largest ones ever made—is that individual galaxies, and even whole galaxy clusters, are spread across the surface of vast "bubbles" in space. Much like soapy water, the gigantic bubbles ostensibly fill the entire Universe, whereas the voids are the interiors of those bubbles. Furthermore, the galaxies seem distributed like beads on strings only because the observed two-dimensional maps are actually crossectional cuts through the real three-dimensional bubbles. The densest of the galaxy clusters and perhaps the superclusters apparently lie in regions where several bubbles meet—at intersections and nodes of vast cosmic filaments, that is, at some of the great crossroads of the Universe. The observed, "frothy" patterns of galaxies in deep space might be telling us something about our origins, for those patterns are probably traceable to the earliest parts of the Particle Epoch.

All told, individual galaxies contribute little to the large-scale architecture of the Universe as a grand cosmic system—but they are key to unraveling that architecture. Each galaxy is essentially a passenger on an expanding, foamy framework, much like humans who have little bearing on the overall tectonics of Earth yet ride along with the drifting continents. On the other hand, galaxies can be used to probe the framework of the Universe, in much the same way that geologists probe the structure of Earth. Metaphorically, galaxies resemble billiard balls whose motions can help determine the size and shape of a playing table, or, better yet, golf balls that can survey the curved topology of a putting green. Cosmologists thereby analyze the radiation emitted by distant galaxies to unravel the very fabric of the Universe, a vitally important endeavor for any full appreciation of the cosmos.

We have reached the limits of telescopic exploration—at least as pertains to the size and scale of organized structures. We have also broached the realm of conjecture at the upper end of those structures. Let's pause for a moment to recapitulate the mental picture before us: We live on planet Earth, which orbits the Sun. The Sun, in turn, is just one of hundreds of billions of stars in the immense Milky Way. Our Galaxy is moreover only one of many residents of the Local Group, which, in turn, is merely an undistinguished galaxy cluster near the periphery of what might be an even larger galaxy supercluster. And so on, among the filaments, voids, and potentially greater structures in the Universe.

At every level in our inventory, nothing seems special about our Earth, our Sun, our Galaxy, our Local Group. Evidently, mediocrity reigns throughout.

Such is our niche in the Universe.

∞

Astronomers have charted normal galaxies out to several billion light-years. Many galaxylike objects are also known to exist beyond this galaxy horizon, but their fuzziness makes it tough to place them into any of the normal galaxy categories. More importantly, the basic character of many of the most distant objects differs from those nearby. By and large, objects more than several billion light-years away are more "active," to a certain extent more violent. Overall, the radiative powers of the active galaxies are much greater than those of the nearby spirals and ellipticals. Furthermore, the active galaxies emit copious amounts of different kinds of radiation—for example, X rays from the interior cores of those galaxies and radio waves from the environments well beyond their cores.

The adjective "normal," used to describe the elliptical, spiral, and irregular galaxies, conveys that those objects radiate the accumulated light of large numbers of stars. Much of their emitted energy is of the visible type, supplemented by lesser amounts of radio, infrared, ultraviolet, or X-ray radiation. That's because stars, too, emit mostly in the visible part of the spectrum. But this is not true for the active galaxies. Some of them are completely invisible to us, undetectable with even the world's largest optical telescopes. Their presence is sensed and studied by radio and infrared telescopes on Earth and by orbiting satellites capable of capturing higher-energy photons. Radiation from the active galaxies, then, is largely inconsistent with the summed emission of myriad individual stars. To be blunt about it, astrophysicists are unsure if active galaxies really have many stars.

The abnormal power and odd character of these mostly distant astronomical objects imply that the Universe was once more robust than it is today. They confirm the idea, described in the previous Particle Epoch, that the first few billion years of the Universe must have been a tumultuous period, quite unlike the more tranquil state surrounding us now in space and time. Since physical conditions were undoubtedly different in the earlier Universe—and recalling that probing great distances in space equals searching far back in time—it shouldn't surprise us that remote objects seen in their youth differ from nearby, older ones. What is enigmatic—in fact, downright astounding—is the enormous amount of energy radiated by some of the most powerful active galaxies. Their total release of energy often stretches astrophysical understanding to its limits.

To gain some perspective, an average star such as our Sun emits in any one second the equivalent of about a billion-megaton nuclear bomb—an impressive feat in and of itself. Yet our Galaxy is a hundred billion times more powerful because, after all, it contains that many more stars. By contrast, an active galaxy is generally a hundred to a thousand times more energetic than that. Active galaxies can launch *in one second* as much radiation as the Sun emits in about a million years.

Now imagine the equivalent of a hundred normal galaxies all packed into the space usually occupied by one. This is the crux of the problem encountered while trying to fathom the monstrously active galaxies. Decades ago, it was fashionable to suggest that these objects were the sites of spectacular galaxy collisions. However, as noted above, computer simulations now show that even such collisions would not produce energy in the amount required nor much explosiveness at all.

The fact that active galaxies often emit more invisible than visible radiation implies that these objects differ fundamentally from normal galaxies. What's more, some of the active galaxies' cores are extremely luminous while others sport huge lobes that resemble wings, all of which further exacerbate their many oddities, making them among the hardest objects in the Universe to decipher. Perhaps we shouldn't even be calling them galaxies.

The gross emission features of some active galaxies can be explained by invoking a distinctly nonstellar mechanism. Called the "synchrotron process" after the laboratory accelerators (sometimes called synchrotrons) used to study subatomic particles, this nonthermal action describes the emission of radiation when charged elementary particles interact with magnetic fields. No stars are involved, nor is any heat per se, hence the term "nonthermal." The radiation arises simply from fast-moving particles, especially electrons, traveling through magnetized regions of space.

Magnetism presumably pervades all things, not just Earth, the Sun, and the Solar System but also entire galaxies. Although the magnetic forces in typically diffuse galaxies are some millions of times weaker than on Earth, magnetism can still play a significant role, especially when its effects mount across an entire galaxy. For many active galaxies, especially a subclass known as radio galaxies, the emitted radiation arises from a pair of oppositely aligned and hugely extended lobes that often span a million light-years; that's a single object equal to some ten times the size of our Milky Way, in fact comparable to the Local Group of galaxies. Fortunately, images of a handful of these objects—most notably one of the closest (at three billion light-years!) of the active galaxies, known only by its catalog name of 3C273—also reveal a kind of Rosetta Stone: a jet of high-speed matter fired from the core of the galaxy out into the intergalactic medium, thus "feeding" the extended lobes farther away. The velocity of the outflowing matter in the jets typically measures fifty thousand kilometers per second, or nearly two-tenths of the speed of light, and some of the most energetic ones surpass half of light's speed. The jets themselves not only point toward the huge lobes from which most of the invisible radiation arises, but, more tellingly, they also point back to the central nucleus where the energy is actually produced.

Laboratory experiments have proved that when charged particles, particularly electrons, are injected into a magnetic field, they spiral around much like the needle of a compass thrown spinning through the air. Magnetism slows the particles, causing some of their kinetic energy to be changed into radiant energy (which is why the process is technically termed "nonthermal *bremsstrahlung*," or breaking radiation). The amount of radiation emitted from a single encounter of an electron and magnetism is not terribly large in the laboratory. But in the case of a huge galaxylike object, the radiation can mount fiercely because of vast numbers of electron encounters. Furthermore, the emitted radiation is theorized to be of the radio variety, in accord with what is observed.

That said, the details of the emission mechanism within many active galaxies remain enigmatic, even assuming repeated injections of fast and numerous electrons into the galaxies' lobes. Although the synchrotron process gives us an inkling of the type of abnormal event responsible for the emission of such intense radio power, active galaxies also display a kind of explosiveness that requires continual acceleration of electrons to speeds close to that of light itself. Moreover, large clumps of plasma are observed and occasionally tracked moving outward, forming the extended lobes so characteristic of many of these active galaxies. The implication is that fast-moving matter is violently ejected in opposite directions by extraordinarily energetic events at the cores of these galaxies.

What might be the source of such great energy? Can any known means explain such outbursts on truly galactic scales? Somewhat ironically, black holes can—or so astronomers think. But before encountering these denizens of Nature, do note that the active galaxies are still not the most energetic objects in the Universe. An additional, extraordinarily luminous class of active astronomical objects has been monitored for several decades now—objects so puzzling that they sometimes seem to defy the currently known laws of physics. These are the innocuous-looking, though inordinately powerful, quasi-stellar sources—quasars for short. Resembling common stars, the quasars' very great distances mean that they not only rival the energy emission problems of active galaxies; quasars actually exacerbate those problems. Here's why:

Not only are the quasars the most energetic objects in the known Universe, but their radio and optical radiation is highly variable, often displaying variations from week to week, occasionally from day to day. The implication is straightforward: Galaxy-sized objects could never synchronize their front-to-back emission to produce such rapid and coherent time variations; otherwise, the intensity of those variations would be blurred and not as sharp as observed. Expressed another way, cause-and-effect arguments demand that no object can flicker more quickly than radiation can cross it. Thus, daily variations imply that quasars cannot be much larger than a light-day across, or roughly the diameter of our Solar System. The enormous power of the quasars, ranging from a hundred on up to a million times that of our Milky Way Galaxy, must then arise from a region much smaller than our Milky Way, in fact tiny by cosmic standards. All of which drives us further toward the idea of compact black holes as candidates for the quasars' central engines.

Quasar emission mechanisms—whatever they really are—must operate, again by comparative cosmic standards, within an almost unbelievably small realm of space, conceivably well less than a single light-year. Try to imagine the equivalent of a hundred or more normal galaxies all packed into a region comparable to the Solar System. That's an indication of the anomalous state of affairs needed to appreciate the Herculean quasars, certainly among the most baffling objects in all the Universe.

∞

Black holes. Although perhaps best treated in the context of stars in the next Stellar Epoch, the most massive black holes likely arose during the Galactic Epoch, roughly a billion years after the Matter Era began. Observations made during the 1990s imply that black holes reside in the hearts of most galaxies—relatively dormant holes at the cores of normal galaxies and extremely energetic ones powering the active galaxies. So, rather than sidestepping this important issue—a central topic in much of astrophysics today—let's consider the phenomenon of black holes now.

A black hole is a region containing a huge amount of mass occupying a relatively small volume. It's not an object per se so much as a hole, and one that's dark to boot. Such a hole still exerts gravity, to be sure exceptionally strong gravity, great enough to warp spacetime severely in its vicinity. Its two main features—large mass and small size—guarantee an enormously strong gravitational force. Why? Because one-half of the law of gravity states that its force directly relates to the mass in question. The other half dictates that gravity is inversely proportional to the square of the distance over which the mass is spread—the inverse-square law again, as noted in the Particle Epoch. And because the distance term is squared, the gravitational force grows spectacularly when distances separating any two parts of an object decrease, which is exactly what happens for a compressed object like a black hole.

Gravitational theory—either Newton's or Einstein's, they both make this prediction—stipulates that when any object having a mass of about three times that of the Sun is no longer countered by an countervailing force (such as heat in a star, or rotation in a cloud), that object will collapse indefinitely, crushing matter to the dimensions of a point. It implodes catastrophically without limit; apparently nothing can stop it.

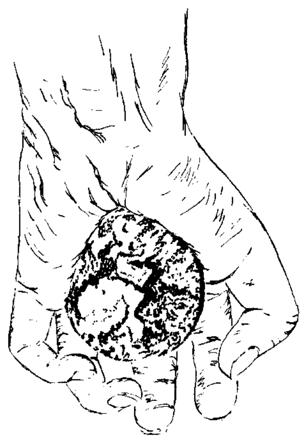
Can anyone possibly grasp such a seemingly ridiculous phenomenon? How can an entire star (or larger) shrink to the size of an atom (or smaller), while presumably on its way to infinitely small dimensions? Does this make any sense? Well, detailed mathematical studies do predict that, without some agent to compete against gravity, massive objects are expected to instantaneously shrink to singular points of infinitesimal volume—singularities, much as posited in the Particle Epoch regarding the origin of the Universe—which is why some researchers consider black holes as "laboratories" in which to explore aspects of the big bang itself. Strange as these statements may seem, observational evidence mounts daily in good agreement with theory. Black holes apparently really do exist.

Though the messy mathematics needed to understand black holes intimately are beyond the scope of this book, we can still explore a few qualitative aspects of these extremely dense and bizarre regions of spacetime. The details are sketchy, precisely because the behavior of matter at extreme densities is not well understood. Magnetism and rotation are also tricky to model for highly compressed objects; the laws of physics here are clearly incomplete. Whoever manages to decipher those details will surely become famous.

Consider first the concept of escape velocity—the speed needed for one object to escape from the gravitational pull of another. For any relatively small piece of matter—a molecule, baseball, rocket, whatever—that velocity is proportional to the square root of an object's mass divided by its radius. For example, on Earth, with a radius of about sixty-four hundred kilometers (or four thousand miles), the escape velocity equals about eleven kilometers per second (or seven miles per second). To launch anything away from the surface of our planet, it must have a velocity greater than this, explaining why typical speeding bullets, fired at about two kilometers per second, return to Earth's surface. Also, the Space Shuttle, for example, orbits Earth at a speed of about eight kilometers per second, but the interplanetary probes, such as *Voyager* that went to Jupiter or *Viking* to Mars, required a boost to eleven kilometers per second to physically escape Earth's gravitational pull.

Consider now a hypothetical experiment for which the apparatus is a gigantic, three-dimensional vise. Imagine the vise to be large enough to hold the entire Earth and, as awful as it sounds, for Earth to be squeezed on all sides. As our planet shrinks under the assault, its density rises because the total amount of mass remains constant inside an ever-shrinking volume. Accordingly, the escape velocity increases.

Suppose that our planet is compressed to one-quarter its present size, thus doubling the escape velocity. Anything attempting to escape from this hypothetically compressed Earth would then need a velocity of at least twenty-two kilometers per second. Imagine compressing Earth still more. Squeeze it, for example, by an additional factor of a thousand, making its radius hardly more than a kilometer. Now its escape velocity increases dramatically, to many hundred kilometers per second.



. . . a hypothetical experiment for which the apparatus is a gigantic, three-dimensional vise.

And so it goes: as an object of any mass contracts, the gravitational force grows stronger at its surface, mostly because of increased density. In fact, if this frightful vise were to compact our home planet hard enough to crush it to merely a centimeter across (about half an inch), then the escape velocity would reach three-hundred thousand kilometers per second (or 186,000 miles per second). And this is no ordinary velocity; it's the velocity of light, the fastest velocity allowed by the laws of physics as we now know them.

So if, by some fantastic means, the entire planet Earth could be shrunk to the size of a pea, then its escape velocity would have to exceed the velocity of light. And since that's impossible, the compelling conclusion is that nothing—absolutely nothing—could get away from the surface of such a compressed "Earth." There is simply no way to launch away a rocket, a beam of light, or anything whatsoever. Furthermore, no exchange of information would be permitted with such an astronomical object. It would have become invisible and uncommunicative, making the origin of the term "black hole" clear. For all practical purposes, such an ultra-compact object has disappeared from the Universe!

The above example is of course hypothetical. It's likely (and fortunate) that no such cosmic vise exists that is capable of squeezing the entire Earth to centimeter dimensions. But in massive stars and galaxies, such a vise does in fact exist—the force of gravity. The relentless pull of gravity is truly strong enough to compress dead stars and galactic cores to extraordinarily small dimensions. The gravitational force in massively compact objects is not at all hypothetical; it's real.

Gravity cannot crush Earth in this way because our planet simply lacks enough mass. The collective gravitational pull of every part of Earth on all other parts of Earth is not powerful enough. However, as we shall see in the next Stellar Epoch, when the nuclear fires have ceased at the end of a star's life, gravity can literally crush a star on all sides, thereby packing a vast amount of matter into a very small sphere.

When stars have more than several solar masses, the critical size at which the escape velocity equals that of light is not, as for Earth, of centimeter dimensions. For typically massive stellar core remnants, this critical size is comparable to kilometers. For example, a ten-solar-mass star would have a critical size of about thirty kilometers. This critical size below which astronomical objects are predicted to disappear is given a special name. Astronomers call it the "event horizon," a region within which no event can ever be seen, heard, or known by anyone outside. Accordingly, the event horizons of Earth and of a ten-solar-mass star equal one centimeter and thirty kilometers, respectively.

We might then claim that magicians really could make coins and rabbits disappear provided they squeezed their hands hard enough. Even people could disappear if compressed to a size smaller than 10^{-23}

centimeter! In English units, that's a trillionth of a trillionth of an inch. Gravity won't naturally do it to us, though, again because we are just not massive enough. The collective gravitational pull of all the atoms in our bodies falls far short of the force needed to compact us to this minuscule size. Nor does any technological device presently known come close to doing so.

The important point here is the following: Should no force, or counteracting agent of some sort, be capable of withstanding the self-gravity of a celestial object having several solar masses or more, then such a hulk will naturally collapse of its own accord to an ever-diminishing size. Theory demands that the infall of such a massive object will not even stop at its event horizon. An event horizon is not a physical boundary of any type, just a communications barrier. The massive object shrinks right on past it to smaller sizes, presumably on its way toward becoming an infinitely small point—singularity again. We say "presumably" because physicists are unsure if any undiscovered forces can halt the catastrophic collapse somewhere between the event horizon and the point of singularity. This, again, is the realm of the as-yet-unconceived subject of quantum gravity, the holy grail of the previous Particle Epoch.

Black holes are products largely of Einstein's relativity theory, although a logical extrapolation of Newton's law of gravity does permit their existence. Whereas the Newtonian theory of gravity describes many other odd phenomena in the Universe, only the Einsteinian theory of spacetime can properly account for the truly bizarre properties of black holes where matter becomes extraordinarily dense. Of particular interest, and to make a connection with the spacetime concepts of the prologue, the mass contained within a black hole is expected to warp greatly both space and time in its vicinity. Close to the hole, the gravitational force becomes overwhelming and the curvature of spacetime extreme. At the event horizon itself, the curvature is so severe that spacetime folds over onto itself, causing objects within it to become trapped and disappear.

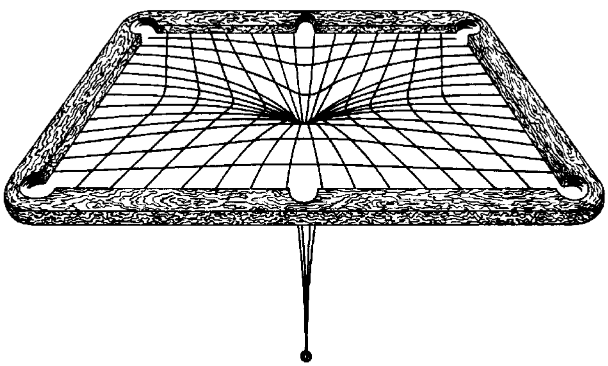
Several props can help us visualize the curvature of spacetime near a black hole. Each way is, however, only an analogy. The problem here, as earlier in the case of the whole Universe, is our inability to deal conceptually with four dimensions. Here's one such fanciful analogy designed to elucidate the formation of black holes and the spacetime warp caused by them:

Imagine a large group of people living on an enormous rubber sheet—a gigantic trampoline of sorts. Deciding to hold a reunion, all except one person converge on a given location at a given time. Their reunion is to be an event in spacetime. The one person remaining behind can still keep in touch by means of "message balls" rolled out to him along the rubber sheet. These balls are the analogue of radiation traveling at the velocity of light, while the rubber sheet mimics the fabric of spacetime itself.

As the people converge, the rubber sheet sags under their growing weight. Their accumulating mass in a small place creates an increasingly large spacetime curvature. The message balls can still reach the lone person residing far away in nearly flat spacetime, but they arrive less frequently as the sheet becomes progressively more warped.

Finally, when enough people have gathered at the appointed spot, their mass becomes too great for the rubber to support. The sheet breaks and compresses them into a bubble, sending them into oblivion and severing communications with the lone survivor outside. Regardless of the speed of the last message ball, it cannot quite outrun the downward-stretching sheet.

Analogously, a black hole is theorized to warp spacetime completely around on itself, thereby isolating it from the rest of the Universe.



. . . a black hole is theorized to warp spacetime completely around on itself, thereby isolating it from the rest of the Universe.

Two important caveats pertain to black holes. The first is that they're not cosmic vacuum cleaners; they don't cruise around interstellar space, sucking up everything in sight. The movements of objects near black holes mimic those of any object near a region of concentrated mass. The only difference is that, in the case of a black hole, objects skirt or orbit about a dark, invisible region, where nothing at all can be seen. Neither emitted nor reflected radiation of any sort emanates from the position of the black hole itself.

Black holes, then, don't go out of their way to drag in matter, but if some matter does happen to infall via the normal pull of gravity, it will be unable to get out. Black holes are like turnstiles, permitting matter to flow in only one direction—inward. Swallowing matter, they constantly increase their mass as well as their event horizons, for the region of invisibility also depends on the amount of mass trapped inside. Those black holes that really do exist in space are probably enlarging their mass and size, some more than others, all of them apparently gulping, eating, growing.

A whirlpool is an apt analogy for the grip that black holes have on matter. Whirlpools of water, for example, tend to have attractive affects on nearby fish. Since the speed of the water is greater closer to the center of the whirlpool, fish entering an area where the water speed is faster than the fish can swim will be sucked inward. Those closest in will never make it out.

Another notable point is that strong gravity near black holes causes great tidal stress. An unfortunate person, falling feet first into a black hole, would find himself stretched enormously in height, all the while being squeezed laterally. He would, moreover, be literally torn apart, for the pull of gravity would be stronger at his feet than at his head. He wouldn't stay in one piece for more than a fraction of a second after passing the event horizon. Similar distortion and breakup apply to any kind of matter near a black hole. Whatever falls in—gas, people, robots, whatever—is vertically stretched and horizontally compressed in the process of being accelerated to high speeds. The upshot is numerous and violent collisions among the torn-up debris, causing much heating of the matter that plunges into the hole.

This rapid destruction of infalling matter by tides and collisions is so efficient that, prior to submersion below the hole's event horizon, even matter outside a black hole can be effectively converted to heat energy while falling inward. Although radiation ceases to be detectable once the hot matter dips below the event horizon, regions just outside black holes are expected to be energetically emitting, mostly in the X-ray part of the spectrum since the matter is so hot. To distant observers, contrary to popular belief, black holes can then appear as bright points of radiation and prodigious *sources* of energy.

With this in mind, and being only partly facetious, perhaps black-hole research may eventually result in practical applications after all. Through some marvel of technology, our descendants might someday learn how to compact garbage to an almost incredibly small size—after which it would disappear! Not only that, the crushed garbage would emit copious amounts of energy in return. Maybe black holes are just what the doctor ordered for technological civilizations long on pollution and short on energy. An ability to tap this energy safely may be a major milestone in the history of any long-lived civilization.

Of much interest is the obvious question, What lies within the event horizon of a black hole? What's it like deep inside? The answers are simple: No one knows.

Some researchers maintain that the inner workings of black holes are irrelevant. In situ observations could conceivably be done by robots sent "down under" to test the nature of space and time beneath the event horizon, but that information could never reach the rest of us outside. Apparently, no theory offered to explain the hidden recesses of black holes could ever be put to the experimental test. Anyone's guess seems as valid as anyone else's. Perhaps the inner sanctums of black holes then represent the ultimate in the unknowable. For that very reason, though, other researchers argue that it's of utmost importance to unravel the nature of black holes, lest we someday begin to worship them. Sounds ridiculous, but whole segments of humankind have often revered the unknowable, venerating that which cannot be tested experimentally. Come to think of it, many still do in twenty-first-century society.

What sense are we to make of black holes? The basis for these outlandish objects is the relativistic concept that mass curves spacetime—an admittedly weird phenomenon, yet one that has been partially tested locally in our Solar System. The larger the mass concentration, the greater the warp, and thus the stranger the observational consequences. Perhaps. Some theorists are convinced that relativity is incorrect, or at least incomplete, when applied to black holes. It does seem nonsensical to claim that very massive astronomical objects will collapse catastrophically to infinitely small points. Not even our wildest imaginations can visualize such phenomena; science-fiction stories fall short, mathematicians are baffled. Maybe the current laws of physics are inadequate in the vicinity of a singularity; precisely *at* the point of singularity, general relativity is probably absurd. On the other hand, perhaps matter trapped in black holes never does compress all the way down to that mathematically arcane singularity. Perhaps matter just approaches this most bizarre state in all of science, in which case relativity theory may still hold true.

This is where in many accounts, even by leading scientists, writers often launch into discussions of parallel universes, multiple universes, hyperspace, warp drive, worm holes, time travel, other dimensions, and a host of other "possibilities," both remote and fanciful. But these and other like-minded speculations are not within the scope of this book. Here, we strive to stay on reasonably solid ground, appealing to empirical findings and acquired data while admitting our ignorance wherever it lay. And when it comes to the secluded sanctorum of black holes, the honest answer is that scientists just don't know what to make of them—nor will we likely ever learn much until the frontier subject of quantum gravity is realized and mastered.

Despite their freakishness, black holes do seemingly populate Nature. In addition to the "smallish," stellar black-hole candidates best assessed in the next Stellar Epoch, many astronomers contend that the much larger galaxies display convincing evidence for black holes. Particularly intriguing are the centers of galaxies, including the core of our own Milky Way, some thirty thousand light-years from Earth. Our Galaxy's midsection should provide us with a stunning view, given that it's teeming with so many billions of densely packed stars. But we don't see its brilliance because its midst is completely obscured by dust, denying studies with optical telescopes; even the largest such devices can visually see only about a tenth of the way toward the galactic center. Fortunately, longer-wavelength, radio and infrared observations are possible, enabling us to probe more deeply into the heart of the Galaxy (much like radar cuts through thick fog on Earth). And what was found in the innermost few hundred light-years initially yielded spectacularly unexpected results; now, in retrospect some two decades later, the findings seem typical of the black holes probably lurking in the hearts of galaxies everywhere.

At the Milky Way's core, infrared sensing shows thousands of stars swarming per cubic light year—a stellar density more than a million times greater than in our solar neighborhood. Giant nebulae tens of light-years across, rich in gas and embedded among even bigger clouds loaded with dust, reside in a ringlike structure more than a thousand light-years across, the whole formation housing some tens of thousands of solar masses and rotating at the fast clip of a hundred kilometers per second (or more than two hundred thousand miles per hour). And in the center of the ring is an intense radio source—the dynamical nucleus of our Galaxy.

On even smaller scales, high-resolution radio maps show an inner ring of gas less than ten light-years across, rotating even more rapidly (at more than a thousand kilometers per second) and resembling a colossal whirlpool at the very center of our Galaxy. This remarkable realm, quite unlike anything near Earth, has been closely monitored ever since it was first found some twenty years ago, including recent outbursts at X-ray wavelengths that imply the presence of a spinning, white-hot accretion disk of million-degree-Celsius gas right in the middle of it all. Magical and mysterious, yet not mystical, the enshrouded nature of this most perplexing piece of galactic real estate so far and foreign is slowly being deciphered.

Frustrated late one evening at the Harvard Observatory, some colleagues wandered to Cambridge Common, where we perched ourselves on a bench near the edge of the park. Straining to fathom the locations of crosswalks, benches, and trees, we gained some insight into the problem of trying to map the Milky Way while stuck inside it. Barring ourselves from walking, bicycling, or otherwise sauntering about, we soon discovered that charting the park's layout is no easy task. Any resulting map would likely be subject to distortion, obscuration, and incompleteness. Statues and signposts—and especially the grand monument near the common's center—seemed especially strange and intriguing from a distance, for they resembled none of the familiar shrubs and benches near the outer part of the park. And so it is with our Milky Way Galaxy. Relegated perhaps forever to the galactic boondocks, we strain to unravel the spread of stars, gas, and dust in that part of the Universe we call home.

Models capable of accounting for most of the galactic-center observations to date stipulate that a rapidly rotating halo of thin, hot, ionized matter surrounds a furiously spinning vortex of even hotter, denser matter. This swirling mess of stars, gas, and dust is apparently orbiting—and here's the punch line—a tremendously compact object housing a few million solar masses, all packed into a region comparable to our Solar System. Such an enormously massive and compact blob is needed to give the maelstrom some structural integrity—to prevent the whirlpool of gas from dispersing into the outer regions of the Galaxy. Fast rotations doubtless produce strong centrifugal forces and, unless a huge mass were gravitationally pulling back, the gas would be flung away like mud from the edge of the spinning bicycle wheel. The implication that millions of stars are compressed to planetary-system dimensions follows from simple, well-understood physics, even if the result borders on surrealism.

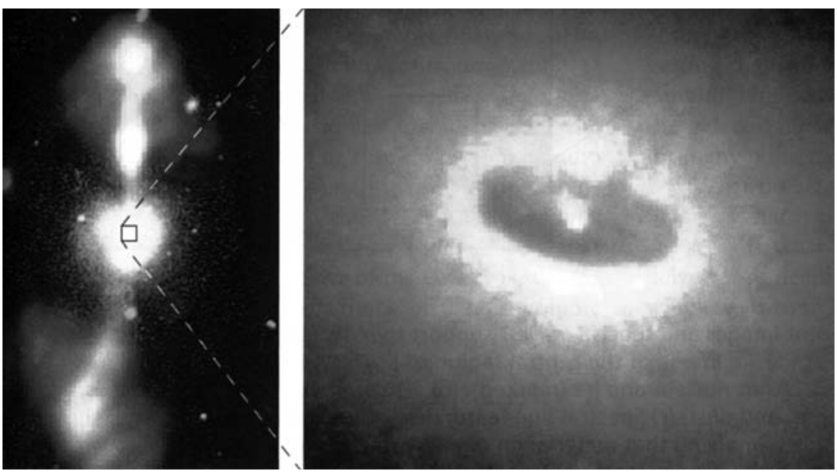
Though the details are controversial, a consensus now seems at hand that a supermassive, ultracompact "something" resides at the hub of our Milky Way Galaxy. As best we can tell, that something can be only one thing—a black hole. Not to worry, the hole currently seems rather quiescent, if not dormant, and in any case is more than two billion times farther from Earth than is the Sun.

Our Milky Way isn't alone in having a troublesome core. Recent observations imply the presence of supermassive objects in or near the middle of many other galaxies. The evidence here is much the same as for our own Galaxy, with gas and stars in the innermost regions of several normal galaxies, including perhaps nearby Andromeda, observed to be rapidly whirling—apparently, again, centered on black holes of millions of solar masses. And although the active galaxies cannot be seen as well owing to their greater distances, observations of them also suggest that highly compact regions lurk in their hearts, usually housing even more gyrating matter than at the cores of the normal galaxies. Astonishingly, for some of the most active galaxies, several billion—not million, but billion—solar masses are implied, all within a region less than a few light-years across. Perhaps these central whirlpools are remnants of the primordial eddies that gave rise to the galaxies, as noted below.

Astronomers now sense that the center of virtually every galaxy is inhabited by a supermassive black hole. Normal galaxies such as our own probably have relatively small holes of "mere" millions of solar masses, most of them, as in our Galaxy, now relatively inactive for lack of fuel. Those galaxies considered more active have larger black holes, often on the order of billions of solar masses, as betrayed by their more intense radiation. It is this enhanced emission of energy that makes them "active," largely because we see many of the distant, active galaxies in their youth, when fuel was more plentiful.

As antic as this scenario seemed when first proposed some twenty years ago, astrophysicists now generally agree that the great energetics of the active galaxies are naturally explained by matter perishing within the clutches of supermassive black holes. Thus, we discern one of the greatest paradoxes in science, as forewarned earlier in this section: black holes trigger some of the brightest objects in the Universe—all of it caused by matter being gobbled, distorted, accelerated, and heated before disappearing below their event horizons. How the enigmatic jets perpendicular to a black hole's accretion disk manage to launch away matter despite the powerful gravity of the hole, however, remains one big puzzle.

Not inconceivably, the most energetic objects in the Universe—the innocuous-looking yet powerful quasars—might be ruled by hypermassive black holes that regularly consume whole stars. Roughly ten stars devoured per year would do it for typical quasars; a thousand stars per year, or therefore a few per day, would be needed to explain the brightest of them. If true, then black holes in quasars might be even more massive, more compact, and more abnormal than the billion-solar-mass objects implied for the active galaxies. This idea, however hard to swallow intellectually—since it's so foreign compared to the more mundane events near us in space and time—can seemingly ex-



Evidence of a supermassive black hole.

At left is a combination of an optical photo and a radio image of a giant elliptical galaxy (called NGC 4261). Its visible part is the blob at center, whereas the invisible radio-emitting lobes at top and bottom extend hundreds of thousands of light-years beyond. At right is a close-up photo of the galaxy's core, revealing a whirling disk of hot gas surrounding a bright hub that likely harbors a black hole containing several million times the mass of our Sun. *Source: National Radio Astronomy Observatory/Space Telescope Science Institute.*

plain most quasar observations. It also has the added advantage of resembling the process thought to power smaller-scale yet still violent regions, such as normal galactic centers and stellar X-ray sources within galaxies, implying that unifying principles may be at work on many scales in Nature.

Clearly, a complete understanding of the powerhouse galaxies lies partly buried deep in their cores, awaiting future explorers to discover, unravel, and share their secrets. There, their central engines are both the instigators of change and the recipients of change; again akin to biological events broadly considered, black holes drive change and adapt to it. Yet the timescales for noticeable change differ so markedly—in biology on the order of thousands to millions of years for species change, in astronomy easily millions to billions of years for architectural change. Astrophysicists are still learning to decipher the clues hidden within invisible radiation emitted by alien environments near hugely massive and totally invisible black holes. We are only beginning to appreciate the full magnitude of these strange new realms deep in the hearts of galaxies.

Some final words of caution regarding black holes, large and small: Forces may yet be discovered capable of withstanding the relentless pull of gravity, even that near exceedingly massive and compact astronomical objects. Magnetism and rotation have not yet been fully incorporated into black-hole theory, and no one knows what to expect regarding the behavior of gravity on deeply submicroscopic, quantum scales. Massive clusters of dark stars and ultradense pools of elementary particles have been proposed as alternatives to black holes, as have queer and inventive groupings of more exotic kinds of dark matter. That these are all terribly hard problems to solve is an understatement, so much so that some of the best minds confess ignorance as to how to go about even attacking them. Serious research regarding realistic models of black holes is only beginning at many observatories around the world.

Skepticism is healthy in science. Unless astrophysicists can find direct, or compellingly indirect, evidence for the existence of black holes, neither of which is currently at hand, then the whole concept of black holes may well turn out to be no more than a whim of human fantasy—another case of mathematics gone awry without the check and balance of tested physical law. The nature of matter, energy, space, and time deep down inside event horizons may be no more significant than a challenging and amusing academic problem devoid of reality.

On the other hand, the Universe did emerge from what seems to have been a naked singularity some fourteen billion years ago. Of all the amassments now known or suspected to be part of our cosmic inventory, black-hole singularities might just be the keys needed to unlock an understanding of the creation state from which the Universe arose. By theoretically studying the nature of black holes, and especially by observationally probing their physical behavior, we shall perhaps someday be in a better position to address *the* most fundamental problem of all—the origin of the Universe itself.

∞

Regardless of how galaxies populate space or how they emit their radiation, an even more basic question comes to mind: Where did the galaxies come from? How did the grandest of material structures arise from an early Universe comprising a uniform mixture of hot matter and intense radiation? Do galaxies form by engorging already-made stars, or do stars gestate in already-made galaxies? Which came first, stars or galaxies? Not least, how do galaxies evolve, once formed?

Fortunately, we can address these and other questions pertaining to the Matter Era with more assurance than the rather uncertain events of the Radiation Era previously described in the Particle Epoch. Even here, though, substantial puzzles remain about the details of the galaxy-formation process. Astrophysicists are now tackling the issue of galaxy origins and have identified its main problems, but they have not yet solved them.

Lack of good observational knowledge of the galaxies themselves creates the basic enigma. Galaxies can be classified according to their gross morphology and their energy budgets, as done above. But we have thus far no explanation for the observed properties of all the galaxies in terms of, for example, the simple gas laws that describe our rather detailed knowledge of stars, a topic of focus next in the Stellar Epoch. Not surprisingly, it's hard to fathom how galaxies emerge and change when we don't quite know what they are.

When put under bright lights and interrogated, astronomers admit they know only a few grand and mutual facts about galaxies. Together, these common denominators are helping us begin to understand the events that produced these most majestic of all objects in the Universe.

First, the galaxies are out there. That's a telling datum, for we do know that the galaxies exist. And our civilization should be mightily proud of that fact; no other life form on Earth knows, or ever has known, of the presence of the galaxies. Yet their mere existence doesn't help us much to decipher their origins. Given the galaxies' expanse and magnificence—in vast numbers, in any direction, as far as our best telescopes can see—we are left perplexed and wondering: Just how did these awesome structures come into being?

Second, there are now no young galaxies. Said another way, no galaxies seem to be originating at the present time. Some may be still growing and developing as they accrete more matter, but none seems to have emerged within the past ten billion years or so. Since all normal galaxies contain some old stars, and since most active galaxies are far away in space (and thus in time), the bulk of the observed galaxies must have come forth long ago. Whatever the seminal mechanism, it was surely widespread in the early parts of the Matter Era. But if the galaxies originated so prolifically in the younger Universe, then why aren't they still doing so now?

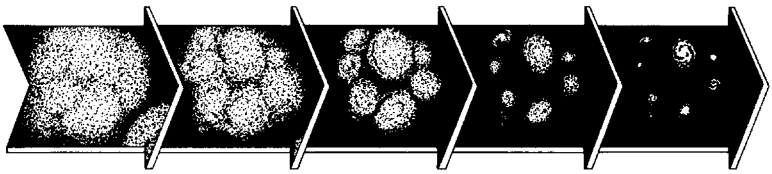
Yet another common factor derives from the finding that most galaxies house comparable amounts of matter. The capacity of virtually every individual galaxy thus far measured ranges between a billion and a trillion solar masses. Normal galaxies appear to have about that many stars and, as best can be determined, active galaxies also include roughly this much matter in some form. No known galaxy is much smaller and none much larger. They all seem to average a hundred billion stars, or their equivalent, much like our own Milky Way Galaxy, give or take a factor of ten for giant ellipticals or dwarf irregulars. Why should Nature's grandest intact assemblies have such a narrow range of sizes? What precludes the construction of galaxies containing, for instance, a quadrillion stars?

To summarize, reiterate, and clarify: There is no evidence that galaxies are originating at the present time, nor have any done so within the past many billions of years. Galaxies do seem to be evolving currently, as noted toward the end of this Galactic Epoch, including additional growth as new matter occasionally falls into already established galaxies. But if new galaxies were emerging only now, astronomers should have spotted some objects having sizes and morphologies somewhere between well-defined galaxies and sheer empty space. We know of no such nearby, amorphous, "half-baked" objects. Furthermore, the regions beyond the galaxy clusters—the intergalactic voids—don't seem to contain much matter, if any at all. Whenever and however the galaxies did form, they apparently did so very efficiently, sweeping up almost all the (normal) matter available and leaving little behind for further assembly.

What's more, key theoretical ideas presented in the next Stellar Epoch strongly suggest that stars ought to be forming now within galaxies. The bulk of most galaxies most likely formed first, yielding environmental conditions ripe for the later formation of the stars we now see richly populating galaxies. These ideas have been handsomely verified in the past couple of decades by splendid observations of widespread locations throughout our Milky Way, where stars are known to be originating slowly but surely from the galactic hodgepodge of loose gas and dust. Recent stellar census implies that roughly ten new stars now form in our Galaxy each year.

To address the issue of galaxy formation, imagine a giant cloud of hydrogen and helium atoms embedded in a weakening sea of radiation, some hundreds of millions of years after the big bang. This giant cloud should not be regarded as filling the entire Universe; rather, think of only a small sector of the cosmos, yet one still millions of light-years across. Although universal expansion continued apace, such a huge clump of mass would not have indefinitely expanded; local gravity had slowed the cloud to a maximum size, after which it began to fall back on itself. The cosmic temperature and density had dropped greatly since the onset of the Matter Era. Radiation was no longer sufficiently intense to shatter atomic matter, as fully formed hydrogen and helium atoms were then numerous enough to exert a collective influence of their own. Electromagnetic and nuclear forces bound elementary particles within atoms, while gravity in turn bound the atoms within the giant cloud. All the known forces that now direct the evolution of matter were already operating well enough to grant the cloud some structural integrity of its own. Vast parcels of matter were becoming distinguishable from one another, each isolated in a fragmenting cosmos, a state of affairs strongly contrasting with the uniform mixing and chaotic violence of the earlier Radiation Era.

Despite its growing stability locally and its steady recession globally, the initially homogeneous cloud would have surely experienced occasional fluctuations—small irregularities in the gas density that came and went at random. No cloud, whether a terrestrial fluffy cloud in Earth's atmosphere, a tenuous interstellar cloud in our Milky Way, or the primordial cloud visualized here in the young and formative Universe, can remain completely homogeneous indefinitely. Eventually, as one atom somewhere in the cloud accidentally moved closer to another, that part of the cloud became just a little denser than the rest. The atoms might have then separated, dispersing this density fluctuation, or they might have acted together to attract a third atom to enhance it. In this way, small pockets of gas arose anywhere in the cloud simply by virtue of random atomic motions. Each pocket of enhanced matter was a temporary condensation in an otherwise rarefied medium. The whole process is not unlike the billowing clouds of a terrestrial thunderstorm, collecting, growing, and eventually forming condensation nuclei that give rise to rain.



Vast parcels of matter were becoming distinguishable from one another . . .

Provided some density fluctuations further developed by gravitationally attracting many more atoms, they could have conceivably grown into clumps of matter that became the seeds of galaxies. Theoretical calculations support the idea that such chancy gas fluctuations could have been the forerunners—protogalaxies—of today's galaxies. But—and this is an important *but*—these same calculations suggest that, given the slow rate of chance encounters, the protogalaxies would only just be forming at the present time. Yet, as noted, astronomers have no evidence whatever that galaxies are now orginating; we have found few, if any regions caught in the act midway between full-fledged galaxies and intergalactic nothingness.

Extremely long times—typically several tens of billions of years—are needed for enough randomly moving atoms to coalesce into a large pocket of gas that can be rightfully called a galaxy. This lengthy duration is not surprising given the absolutely gargantuan quantity of atoms in a typical galaxy—namely, nearly a million billion billion billion billion billion billion billion atoms. How do we know that? Well, each galaxy houses about a hundred billion stars, each star averages a million billion billion billion grams, and each gram has a million billion billion atoms, all of which adds up to a very big number. That's why astronomers prefer scientific notation, in which case the number is some 10^{68} atoms—clearly an awful lot of atoms to collect regardless of the notation used. Consequently, it takes a great while for Nature to do it at random.

But—and this is an even bigger *but*—no scientist ever said that galaxies were built by random events, by chance and chance alone. Some philosophers of science or historians of science or others who, like postmodernists, tend to criticize the methodology of science yet have never practiced science themselves, have occasionally made such claims to champion the cause of pure chance. By contrast, few natural scientists have ever argued that chance and only chance plays a role in any physical phenomenon. Rather, Nature almost surely operates by combining chance with necessity, randomness with determinism—a basic, unifying issue to which we shall return several times in this book, especially when describing the origin and evolution of life in later epochs.

A time of several tens of billions of years is of course well longer than the current age of the Universe—meaning that no galaxies should now exist. So, despite the likelihood that random density fluctuations in an otherwise homogeneous gas could have *eventually* produced galaxies, it's unlikely that the galaxies we now see emerged strictly in this way. Chance cannot be the sole factor governing the origin of these truly immense cosmic systems. Still, the idea of naturally arising spots of different gas densities remains a powerful concept, for it's a reasonably well understood process not requiring any unknown forces or unique conditions. If we could find an agent or mechanism, some means or another, that would accelerate the growth of the gargantuan number of atoms needed to form a galaxy, then we might begin to understand their origins.

To clarify the oft-misunderstood issue of chance versus necessity: Chance surely does play some role in galaxy formation, especially as the initial trigger that starts the fragmentation of primordial clouds. Other, more deterministic agents in the early Matter Era, such as turbulence or shocks, likely enhanced the growth of the inhomogeneities so that myriad galaxies could have formed in a timescale shorter than the age of the Universe. Or, perhaps the seeds of the galaxies were sown at the quantum level much earlier, in the chaotic events of the Radiation Era as proposed below. Whatever it was and however it worked, the enhancement process must have been surprisingly effective since observations clearly imply that the bulk of virtually all galaxies formed long ago, apparently within the first few billion years after the big bang.

The problem of galaxy formation is currently a tough one for astrophysicists. Its solution has exasperated many brilliant minds and is still not yet in hand. The origin of galaxies appeals to theorists with fertile imaginations (to visualize conditions so long ago) and computing skills (to keep track of all those atoms), and especially to those willing to make unorthodox assumptions. This is one of the trickiest areas of astronomy to appreciate, for few hard facts are known about galaxies, and fewer still about the physical events that formed them long ago.

One hard fact that is clearly known, however, is the first one noted above: galaxies do exist. They populate the Universe in great abundance. Somehow they came into being, and somehow they got to be where and when we find them now in space and time. Let's consider in greater detail some of the specific galaxy formation scenarios recently proposed by theoreticians.

Astrophysicists today seek to identify ways that random gas fluctuations might have been enhanced earlier in the Universe. If some factor could be found that might have speeded the growth of the density irregularities, the galaxy-formation problem might be solvable. One such possibility assumes that the Universe was quite turbulent long ago—a not altogether unreasonable idea since turbulence involves the inevitable "confusion" or disordered motion of matter (the gas) within a rapidly moving medium (space itself).

Once the Matter Era dawned, all the atoms within the vast primordial clouds were set into motion not only from the expulsion of the bang but also from the heat of the fireball. The gas then had some "directed" kinetic energy—outward, from the ordered expansion of the Universe. It also had some "undirected" thermal energy—random, from the disordered aftermath of the blazing inferno. Intact pockets of gas undoubtedly surged this way and that, whirling here and shearing there amid collisions with one other, in addition to the disarrayed agitation of each of the individual atoms. In particular, turbulence probably aided the growth of spinning eddies at those places where density fluctuations had already become established in the early Universe.

Turbulent eddies of this sort can be visualized by watching water swirl down the drain of a bathtub. In a sense, the swirling eddies themselves *are* turbulence. Even better examples can be created by moving your hand gently through water, or a teaspoon through coffee; swirling eddies naturally emerge in the wake of this turbulence. Water flowing past rocks in a stream also gives an appreciation for the whirlpools that naturally arise in its aftermath.

Probably the best examples of the effects of turbulence are the fluffy clouds of Earth's atmosphere. Especially vivid in photographs of the tops of clouds, taken with Earth-orbiting satellites, kilometer-sized eddies can be seen as density enhancements of the atmospheric gas. Such whirling eddies are known to become more pronounced whenever air currents are particularly turbulent. Should they grow, in this case by accumulating additional amounts of moisture, the eddies may well form stormy depressions and occasionally even hurricanes hundreds of kilometers across.

Here is a case, then, where studies of a terrestrial phenomenon—Earth's weather—may help us understand one of today's most vexing extraterrestrial problems. Planetary hurricanes roughly mimic the overall morphology, the pancake shape, the differential rotation, and the disposition of energy within spiral galaxies. Though these two systems are entirely unrelated and of vastly different sizes, their many resemblances might teach us something about galaxy formation via the study of hurricane formation. In particular, since most meteorologists agree that some sort of turbulent "priming" is needed to trigger hurricanes, the early stages of such storms could conceivably be used by astronomers to extract clues about the turbulence-enhanced density fluctuations that gave rise to protogalaxies long ago.

It's worth pursuing this idea a little further. Despite the inevitable cooling caused by the expansion of the Universe, each localized eddy within a large gas cloud must begin to heat. It can't avoid it. Eddies are sites not only of turbulence but also of rising heat within a steadily cooling cloud. The heat results from friction caused by frequent collisions among the increasingly dense collections of atoms within each eddy. The process is a simple one, not unlike the heat derived by rubbing our hands together on a cold wintry day.

Eventually, individual eddies must rid themselves of some of this newly acquired energy, much as the Sun or any other heated object needs to unload energy, lest it blow up. The eddies in the protogalactic cloud did it by radiating away some of their heat. In this way, a large cloud containing lots of eddies can cool even faster than would the normal, homogenous clouds of the expanding Universe. As it cools, the entire cloud contracts a little, thereby increasing the density and hence the heat within each eddy. Both the individual eddies and the whole cloud simultaneously radiate some of this newly gained energy into space, thereby allowing further contraction of the parent cloud and its smaller eddies. On and on, this cycle of contracting, heating, radiating, cooling, and contracting proceeds—all fundamentally driven by gravity. The cycle may operate at different speeds for each eddy, particularly since some eddies will be more successful than others at sweeping up additional gas from the parent cloud.

It's easy to conceptualize a cluster of galaxies forming in this way, with each eddy becoming a member galaxy within that cluster. Alternatively, perhaps only one or a few galaxies formed within each of the vast primordial clouds of the early Matter Era, after which gravity gradually swept the galaxies into the very much larger galaxy clusters now seen scattered throughout the Universe. Either way, fragmentation models of this sort resemble a "top-down" approach to galaxy birth whereby huge clouds give rise to litters of young galaxies—a process known in the trade as "monolithic collapse."

As nice as this galaxy-formation scenario seems, it, too, runs into some serious problems once mathematics are applied to it. Calculations show that timing is once again an issue but not, as above, because the eddies take too long to form. Here, it's more a case of competing timescales between physical events affecting the eddies: the time needed for capture and contraction of the gas in a turbulent eddy is longer than the typical time for the random dissipation of that eddy. In other words, eddies tend to break up long before they have a real chance to bind tightly. Turbulent eddies do enhance the random gas fluctuations, but they don't last long enough to form galaxies.

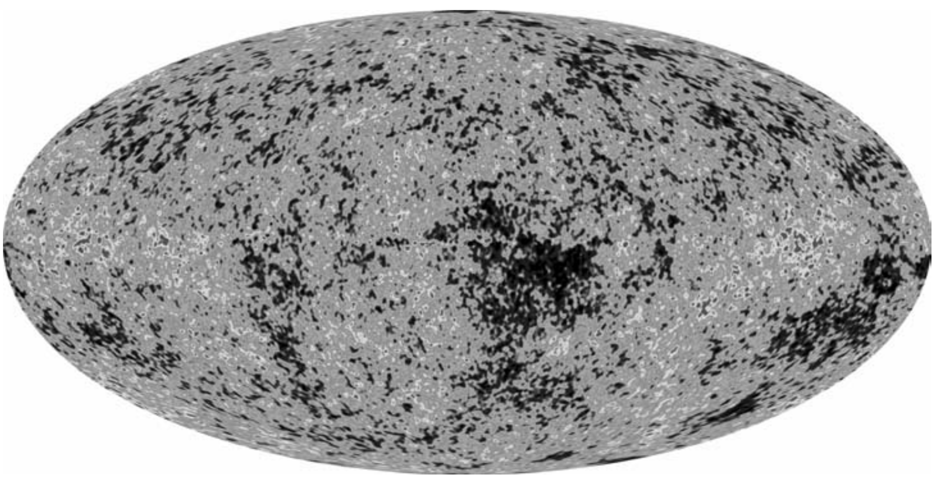
Any kind of eddy, then—in the bathtub or in the early Universe—comes and goes in iffy sorts of ways, all governed by the laws of statistical physics. Eddies appear, disappear, and reappear at different parts of either a terrestrial atmospheric cloud of moist air or an extraterrestrial galactic cloud of primordial gas. Occasionally, a terrestrial eddy does indeed grow to form a flourishing hurricane, or a primordial eddy, presumably a genuine galaxy. But the expected rarity of their rapid growth implies that turbulent eddies cannot be the sole solution to the problem of the ancient formation of the galaxies or of galaxylike objects.

Mainstream astrophysicists prefer to avoid radical theories of galaxy formation—such as a weird one postulating the ballooning of compact, primordial blobs (called by some "white holes") for which there's no evidence whatever. They head back to first principles and embrace once again the basic notion of random gas fluctuations developing into something bigger—a "bottom-up" approach that groups smaller chunks of matter to build galaxies. Still, some additional means must be found to speed the growth of such fluctuations in the gas-radiation mix of the early cosmic fireball. Current research therefore centers on other ways that might have enhanced, or accelerated, the growth of simple gas fluctuations.

The general scenario now favored by the astronomical community—an idea known as hierarchical clustering—postulates an early Universe that was not homogeneous. Instead, it's imagined to have been peppered, even in the Particle Epoch, with minute density clumps. In other words, the eddies got a head start even in the Radiation Era and thereafter acted as seeds for the growth of galaxies early in the Matter Era. These already-formed pockets of gas would then have developed during the Galactic Epoch to fabricate at least the essential features of galaxies seen today. Although this idea initially sounded like a cop-out to many astronomers, observational evidence for these truly primeval inhomogeneities was marvelously confirmed in the first few years of the twenty-first century, allowing theorists to breathe a sigh of relief that they might be on the right track.

Our only direct probe into the early Universe is the cosmic background radiation noted near the end of the prologue and again briefly in the previous Particle Epoch. Launched at the end of the Radiation Era, some half-million years after the start of all things, the radio photons now engulfing us grant some inkling of the wild physical conditions prevailing at the time. Briefly explained, radiation is influenced by the gravity of growing clumps of matter, so that as the density of the clumps varied from place to place in the early Universe, the observed radiation—then launched and now observed—ought to show slight temperature variations from place to place on the sky. Such "ripples" in the temperature of the background radiation have indeed been spotted, though only weakly, at the level of parts per million. That is, given that the average temperature of the fossil radiation is about 3 degrees absolute, or −270 degrees Celsius, the minute thermal variations that have been detected by radio receivers aboard Earth-orbiting satellites, most notably the Wilkinson Microwave Anisotropy Probe (WMAP), are only on the order of millionths of one degree. Yet they are in accord with those expected for a wide range of theoretical models of galaxy formation, including the superclusters, voids, filaments, and bubbles observed all across the firmament.

Here, in a nutshell, is the basic idea of hierarchical clustering, considered more of an ongoing process than a single event: Extremely small-scale fluctuations in the matter density present before the time of inflation—an inevitable consequence of quantum physics operating in the very early Universe well less than a second old—would have been stretched and amplified by inflation to a size and scale typifying whole galaxies and even larger. The subsequent growth of those gravitational instabilities, already established when the Radiation Era gave way to the Matter Era, probably led to the gradual formation of self-gravitating collections of matter. Should this idea be correct, then the vast assemblages of matter we see today as galaxies, galaxy clusters, and even the gargantuan galaxy superclusters are the progeny of subatomic quantum effects prevalent when the Universe was a mere 10^{-35} second old.



Evidence of galactic origins.

This is a map of temperature variations in the cosmic background radiation measured across the entire sky, much more sensitive than the one shown in the prologue. Here, the thermal changes are minute, amounting to mere millionths of a degree Celsius, yet they display clear departures from an otherwise uniform sea of radiation dating back to about a half-million years after the big bang. These variations, shown here as shades of grey and implying clumps of enhanced density, were probably the "seeds" from which galaxies began forming in the earlier Universe. *Source: Wilkinson Microwave Anisotropy Probe.*

The accepted mechanism of galaxy formation, still only roughly understood, is then a familiar one to experts of star formation, as will be examined in detail in the Stellar Epoch. Nature quite naturally selects mass-density fluctuations that gravitationally induce cycles of contracting, heating, radiating, cooling, and eventual flattening into disk-shaped objects. But, just when we feel good about getting closer to grasping reality, another complication sets in. For galaxies, unlike for stars, these events didn't likely involve only normal matter; dark matter has been implicated to some (unknown) extent, and that clearly confuses things.

Given the prevailing conditions in the early Universe, specifically at the interface of the Matter and Radiation Eras, only regions of higher-than-average density containing more than a million times the mass of the Sun would have begun to contract. However, if galaxies grew long ago exclusively from the fluctuations within normal matter (in the absence of dark matter), those density fluctuations should manifest themselves now as a clear observable imprint of *large* temperature variations in today's cosmic background radiation; that imprint is not observed.

Instead, if dark matter was involved, it might have acted as that long-sought agent, or gravitational scaffolding, to help normal matter clump earlier in the Universe. The reason is that dark matter—whatever its true nature—interacts only weakly with normal matter and with radiation. So, its natural tendency to gravitationally infall (for dark matter still exerts gravity) was neither hindered by radiation, nor expected to leave a large signature on the cosmic background radiation. Accordingly, dark matter, being ten times more abundant than normal matter, probably clumped first and then acted as an accelerant to draw normal matter into the regions of highest density. This scenario explains why so much dark matter seems to reside in the vicinity of the visible galaxies. That's where the dark matter initially concentrated, thus attracting the normal matter that became the galaxies now so luminously seen. The brightly lit galaxies resemble the visible tips of mostly hidden icebergs, or the illuminated bulbs on an otherwise dark Christmas tree.

Of course, all this modeling is a little shaky given that astronomers don't yet know the nature of that dark matter. To be honest, some of the uncertainty is welcome, allowing theorists much freedom in choosing dark matter's properties while seeking to match galaxy-formation models with observed structures in the sky—and that, in turn, might imply valuable information about the dark matter. To be just as honest, the theorists may be—to make a bad pun—whistling in the dark, as their models depend on vast quantities of abnormal matter that is only inferred and has never been detected.

In any case, the seeds of galaxy formation were likely sown in the very early Universe when small density fluctuations in the primordial matter began to grow. The initial masses of these pregalactic blobs were quite small by galactic standards—perhaps only a few million, yet more likely a few billion, solar masses, comparable to those of the smallest, irregular galaxies. Those irregulars now seen scattered all around the edges of galaxy clusters may well be the building blocks of galaxies—the so-called baby galaxies. As we shall see in the final section of this Galactic Epoch, a growing consensus champions the idea that today's big galaxies formed by the repeated merging and accumulation of smaller objects. This is indeed a "bottom-up" scheme, but not one that begins with objects as small as stars and planets, rather, with million-to-billion-solar-mass blobs that emerged near the start of the Matter Era.

Support for this hierarchical scenario is moderate and derives from two fronts. Theoretical backing is provided mainly by computer simulations stipulating how normal (baryonic) and abnormal (dark) matter might have interacted with radiation during the Universe's first few billion years. These models demonstrate that merging was a viable phenomenon in the Galactic Epoch and could have conceivably led to the formation (and evolution) of the many varied galaxies observed today. Although the models have wide latitude among their input parameters, while at the same time suffer from computer codes obviously not as robust as the real cosmos, no "showstoppers" have yet intruded—nothing in the theoretical analyses that leads us to believe we are not on the right track, finally.

Observational support derives from the finding that some of the most remote galaxies (namely, those seen in their youth) appear distinctly smaller and less regular than those found nearby. Deep, long-exposed images acquired by the world's most powerful telescopes—such as the Hubble Space Telescope in orbit, the Keck Observatory in Hawaii, and the Very Large Telescope in Chile—show evidence for distant and distorted spheroids containing a million to a billion solar masses (but no distinct stars) in regions typically a few thousand light-years across—roughly the size and scale expected for pregalactic building blocks. We seem to be seeing these blobs as they were some twelve billion years ago, perhaps poised to merge into larger, galaxy-sized objects. Alas, not all astronomers buy this interpretation, as the data are sketchy, the images fuzzy, and the modeling simplified. It remains unclear if anyone has yet seen a genuine "baby" galaxy or any luminous object caught in the act of galactic birth—another of science's unachieved grails.

∞

Which came first: black holes or galaxies? In other words, did supermassive black holes form initially and then accumulate around them matter that eventually became genuine galaxies, or did the galaxies form much as we see them now, after which they gave birth to holes at their cores as early matter migrated toward their centers? This is the first of several chicken-or-egg conundrums encountered in cosmic evolution, most of them unresolved or at least not solved satisfactorily to date. That's probably because nothing in Nature is black or white, few solutions are clean and clear; rather, reality, and especially our models of it, possess shades of gray throughout.

Favoring the "inside-out" idea, whereby black holes form first, is the notion that in any gravitationally bound system the densest things tend to infall early on, followed by the host galaxy taking shape around the central hole. Some data bolster this idea demographically: given that quasars are more abundant than galaxies as we probe farther back in time—and where there are quasars, there are likely supermassive black holes—it would seem that black holes led the way.

By contrast, the radiative effects of the really big black holes might have actually hindered the formation of host galaxies, meaning that the galaxies probably formed first. If so, then the process was more "outside-in," whereby the galaxies came first, at least in rough form, after which the stars, gas, and dust later trickled toward the cores to create the huge black-hole engines. Computer simulations do imply that powerful jets associated with young, massive holes would have blown away surrounding material, possibly preventing the formation of galaxies at all. Furthermore, many supermassive black holes are still actively accreting matter, implying that the process of creating them is actually quite slow, perhaps taking many billions of years to settle at the cores of already formed galaxies.

The answer, citing those shades of gray again, likely mixes aspects of both models—that is, massive black holes and enveloping galaxies may have formed together. Astronomers have discovered recently that the mass of the central black hole is proportional to the bulge of their host galaxy, so the construction of both might well have been tightly wedded and coeval in Nature.

We can pose this unsolved riddle in another related way: Did the galaxies precede the stars or was it the other way around? The answer is important for the cosmic-evolutionary scenario since, as told here, the Galactic Epoch precedes the Stellar Epoch. Is this justified? Most modern arguments do favor early origins for galaxies, followed by later formation of stars and then planets within those galaxies. But the latest data are beginning to soften that view or at least to muddy the waters a bit.

Recent findings suggest that some star formation must have occurred early in the Galactic Epoch, since traces of heavy elements, such as carbon, silicon, magnesium, and iron, are observed to have been present ten billion years ago. We know this because quasars, being typically a hundred times brighter than normal galaxies, act like thin-beam cosmic flashlights, illuminating that part of intergalactic space between the quasars and Earth. And in the quasars' spectra—when their light is split into its component colors—is clear evidence for minute amounts of heavy elements (about a hundred times less than those in the Sun), implying that at least some stars lived and died back then, for stars are the only places known where heavy elements are made. (Astronomers take the "heavies"—sometimes also called metals, to the dismay of chemists—to mean any element more massive than helium.)

The idea that *some* massive stars preceded the galaxies is bolstered by evidence (from quasar spectra) that early in the Matter Era the Universe was reionized, separating atoms everywhere into ions and electrons, much as had been the case in the earlier plasma-rich Radiation Era. This would have been a relatively brief period, probably less than a billion years following the "cosmic dark ages" when no luminous objects anywhere—no quasars, stars, or any other kind of light-emitting bodies whatever—had yet graced the cosmos. All was completely and totally dark, from the first half-million years—when the Universe became neutralized and cosmic expansion redshifted the background radiation out of the visible and into the infrared part of the spectrum—to roughly a half-billion years after the bang, when gravity finally but only locally overcame expansion enough to begin clumping matter into spherical structures. As the first glowing objects—almost surely the building blocks of fledgling galaxies—began emerging from those dark ages, a renaissance of light began to flow through the Universe. The details are murky, but a consensus has emerged:

Surely quasars and possibly massive stars formed in the young Universe, starting no more than a billion years into the arrow of time. Objects smaller than a million solar masses would not likely have clumped, given the rapidly expansive conditions at the time; the thermodynamics in that warm environment would tend to dissipate smaller clumps. The quasars largely lit up (and ionized) matter near the start of the Galactic Epoch, possibly aided by ultraviolet radiation from the earliest stars. In addition, those "first stars" quickly created some heavy elements through the same kind of nuclear-fusion events that occur in stars today, as later explained in the Stellar Epoch—so quickly that all these first massive stars are now long gone, having dramatically expired as supernovae (or having been eaten by black holes—it's possible!) within those first few billion years. Although we do see plenty of quasars in the earlier Universe, not a shred of observational evidence exists for those first stars—which means either that they did all disappear somehow (if theory is right) or that they never existed (if theory is wrong). Nor have astronomers ever found any stars with zero heavy-element content within them, as would be the case for any celestial objects among the first genuine stars. Perhaps the quasars themselves did all the reionizing, without the need for any early stars.

The upshot is that mostly big, million-to-billion-solar-mass blobs likely took shape early in the Galactic Epoch. These were the building blocks of galaxies—almost surely quasars and their black-hole engines, and possibly massive star groupings that resembled today's globular star clusters, which still linger in the haloes of many nearby galaxies. The quasars were clearly there then, probably thousands of times more populous than now; our telescopes spy on them in the distant past, a few billion years after the big bang, when the number of quasars peaked. (None of them resides near us in space or time; the closest quasar is more than two billion light-years distant, the last of a dying breed.) As best we can explore those truly ancient times, the primordial blobs are mostly gone, presumably having merged together quickly to build the galaxies. Those blobs, either lit with stars or not, must have repeatedly merged to make virtually all the galaxies within the first few billion years—which is probably why, even in our own Milky Way, most globular star clusters in the halo average twelve billion years in age and none is younger than nine billion years. To what extent stars were already up and running in those formative blobs, or whether the stars originated mostly after the fledgling galaxies had formed, is frankly unknown.

Astronomers are closely examining the latest data from both stars and galaxies, struggling to get the timing and sequencing correct. The task is nontrivial, for we are looking *way back in time*, trying to decipher events long over and done. To date, these data imply major assembly of the galaxies from smaller blobs mostly within two to four billion years after the bang; later formation was not as robust, if only because universal expansion was continuing to carry those building blocks and the young galaxies away from one another, reducing the number of interactions. By contrast, stars' formation rates peaked some five to ten billion years after the bang; this we know by tracking back in time their ultraviolet radiation—the hallmark of newborn stars. In the main, then, the origin of most galaxies definitely preceded that of most stars. Although star production has been declining during for the past several billion years, stars still do now originate—which means, again in the main (for these are averages over all the details), that the Galactic Epoch preceded the Stellar Epoch.

When did galaxy formation stop, or has it? Astronomers are divided on this issue, too, which may be more semantics than astrophysics. Some contend that at a fairly well-determined time in the past—given by the age of the globular clusters in our Galaxy, for example—most galaxy formation was over. If true, then all galaxies are old, in fact nearly equally old and on the of order twelve billion years. Other astronomers demur, citing evidence that many galaxies seem to experience repeated collisions and mixing with dwarf satellite galaxies over extended periods of time—perhaps even up to the present day. If so, then galaxies might be said to be still forming today.

What constitutes an origin in contrast to evolution? Most experts have reached a tentative consensus that billion-solar-mass protogalaxies became established in some form or another relatively early on, probably within the first couple of billion years of the Matter Era. Virtually all galaxies originated contemporaneously long ago as simply structured yet distinct objects. Their emergence was clearly the dominant feature of the Galactic Epoch. In addition, ongoing mergers, interactions, and rearrangements within and among the galaxies ever since are regarded as evolution—developmental changes that further bulked up the galaxies with each successive merger. To be sure, astronomers have ample evidence that galaxies have evolved in response to external factors, indeed that evolution continues among the galaxies today.

Much of the fascination felt by workers studying the subject of galaxy formation derives from our inability to disprove many contending theories. An array of ideas remains possible, there being only meager experimental data to discriminate among the details. However, this state of affairs will not likely last long when new data begin pouring in at rapid pace as telescopes of the twenty-first century become powerful "time machines" designed to probe the far away and long ago. New and ambitious projects—to name just one, the Sloan Digital Sky Survey now underway—are expected to map accurately several million galaxies in the northern sky in the next few years. Until galactic data become demonstrably better, however, researchers familiar with the sophisticated mathematics of notoriously tough subjects such as fluid mechanics, turbulent physics, and magnetohydrodynamics will continue to justify their interests by tinkering with the problem of the galaxies' origins. For despite heroic efforts of the past few decades to unlock the secrets of galaxy formation, the specifics of a tested, plausible process have thus far eluded discovery.

∞

However galaxies might have originated, either their formative stages or subsequent evolutionary events led to the myriad galaxies now seen in the nighttime sky. We observe loose and tight spiral galaxies with mixtures of old and new stars, large and small ellipticals containing only old stars, dwarf irregular, and explosively active galaxies, let alone the baffling quasars whose central engines may not house any stars at all.

With such a zoo of galaxylike objects littering the Universe, we naturally wonder if any overall pattern or evolutionary scheme interrelates all the various types of galaxies. The answer is, none discerned presently. As best we know, no identifiable physical mechanisms underlie all the galaxies and no clear developmental bonds relate one type of galaxy to another. Whoever does discover strong evolutionary links among the galaxies, akin to those connecting stars as discussed next in the Stellar Epoch, not to mention the elaborate relations among life-forms described later in the Biological Epoch, will get their names in textbooks forever.

Astronomers decades ago proposed an evolutionary progression among normal galaxies, starting with the nearly spherical ellipticals that gradually became squashed ellipticals, eventually changing into closed spirals, followed by open spirals, and finally culminating in irregular galaxies. The central idea here is that galaxies originate with a more or less spherical shape and, as they grow older, their rotation tends to flatten them, first producing some ellipticity and then some spiral arms, prior to their breaking up as aged irregular galaxies. However, therein lies a problem: this type of evolutionary notion requires all elliptical galaxies to be young and all irregular galaxies old—which isn't the case at all. Observationally, elliptical galaxies are not young. They are populated with only old stars, nearly depleted of interstellar gas and dust, and display no evidence of active star formation.

On the other hand, given that the elliptical galaxies are so clearly old, then perhaps the evolutionary sequence runs in the opposite sense. Maybe irregulars are young and, having formed first, gradually evolve into ellipticals. It's easy to imagine loose spiral galaxies wrapping up into tighter spirals and eventually becoming elliptical galaxies. But troubles abound here, too. Apart from the obvious puzzle of how beautiful spirals might have emerged from the contorted irregulars, it's hard to reconcile this idea with the abundance of old stars observed in the irregular and loose-spiral galaxies. Simply put: If irregular and loose spiral galaxies are the starting point in a scheme of galactic evolution, then all of them should be young. But they're not. Virtually all irregulars and spirals contain a mix of old and new stars. The existence of old stars is inconsistent with the nature of a youthful galaxy. The fact that astronomers know of no "dead galaxies" doesn't help our understanding either.

Alas, normal galaxies do not likely evolve directly from one type to another. Spirals don't seem to be ellipticals with arms, nor do ellipticals appear to be spirals without arms. No unambiguous parent-child relationships connect these huge cosmic systems—other than the idea that all galaxies are cousins that trace their birth to the same grandparent, namely, the turbulence of the gases in the aftermath of the big bang. Indeed, all galaxies' dispositions probably result partly from the intrinsic physical conditions extant in the gas clouds from which they originated more than ten billion years ago and partly from environmental interactions with other galaxies ever since.

Frankly, this contrast between intrinsic and surrounding influences is not much different from the way that biological species evolve, combining aspects of their internal genes with those of their external environment. In the above paragraph, we could replace the word galaxy with the word organism and still be reasonably correct. Apparently, the nature-versus-nurture struggle extends beyond the living world. All through this book, we shall be confronted with the issue of whether systems change inherently or in response to external events. The answer for astronomical galaxies, as for biological life, is probably, both. And much like human life, wherein genes are estimated to influence well less than half of human behavior, environmental effects probably dominate changes among the galaxies too.

Astronomers do have ample evidence that galaxies change in response to external, environmental factors, long after the first pregalactic fragments originated. As already noted, given the size, scale, and groupings of galaxies, collisions and interactions among them are commonplace events. This is especially true for the dark-matter halos surrounding many spiral galaxies, including our own, and probably those around all galaxies. Computer simulations performed during the past decade show that these dark halos are strongly involved in, and influenced by, such galactic interactions.

As galaxies orbit or encounter one another, halo material from one galaxy can become stripped by tidal forces exerted by the other. The freed matter often ends up in a common envelope surrounding both galaxies; occasionally it's lost entirely to (that is, flung out of) the system. In this way, even small galaxies can severely distort larger ones, depending upon the angle and proximity of interaction and the energy transferred between them. In some cases, over the course of hundreds of millions of years—a span of cosmic time that powerful computers can model in minutes—the simulations illustrate how close encounters between galaxies can cause spiral arms to appear where none existed before. The pinwheeling arms are literally drawn out of one or both galaxies, as they pass by in each others' wakes like giant ships at sea.

Such environmental factors may be the sole source of galaxies' spiral arms, implying that "arms" are evolutionary appendages, not products of birth. If so, then even our home Milky Way plausibly got its arms by interacting with another galaxy at some time in the past. Our Galaxy's stellar census surely does contain evidence that it has feasted on its neighbors, now seen as remnant, elongated clumps of elderly stars captured into the Milky Way's halo and disk billions of years ago. Perhaps the culprits were systems as small as the Magellanic Clouds now orbiting in the halo of the Milky Way, or the Sagittarius dwarf galaxy now being torn apart and subsumed by our Galaxy on its far side, opposite the Sun. Previous, long-ago encounters with a larger, comparable galactic system, such as the nearby spiral galaxy, Andromeda, is another possibility. Andromeda does currently have a component of motion toward us, meaning that our two giant galactic systems are destined for a close encounter that could cause both to become tidally disrupted and eventually more elliptical. Even more dire (or spectacular, depending on one's viewpoint), these two grand spirals might merge together during their next encounter—the result often glibly called Milkyomeda—though that won't happen for another several billion years.

Mergers and acquisitions may well be common among galaxies in clusters, triggering changes in shape well after their initial formation. To appreciate such evolutionary events, however, we need to contemplate extremely long durations of time. And that's where computer simulations again come in handy. The simulations clearly show that interacting galaxies occasionally tend to gravitate toward one another, eventually merging. What's more, those simulations imply that giant elliptical galaxies probably grew via generations of mergers with spiral galaxies, potentially explaining why the big ellipticals reside near the core of galaxy clusters and the somewhat smaller spirals toward their perimeter. Colloquially termed "galactic cannibalism," or "galaxy gobbling," these are the cases in which the galaxies experience very close encounters, often, in fact, direct collisions. Still, the interactions are sluggish, their explosiveness muted. The last big impacts seemingly occurred eight to ten billion years ago, after which most galaxies, dispersed somewhat by cosmic expansion, have enjoyed a relatively peaceful existence.

Despite these fanciful terms, astronomers have acquired remarkable observational support for such cannibalism, as actual imagery shows smaller galaxies at or near the central regions of large galaxies, apparently in the process of being "digested" as the larger galaxy gobbles them up and consumes them. Such cannibalism may also explain why supermassive galaxies—those having roughly ten times more mass than typical galaxies—are often found near the centers of rich galaxy clusters. The relatively nearby Virgo cluster of galaxies, some sixty million light-years distant, offers a prime example. There, a titanic, trillion-solar-mass galaxy known as Messier 87 resides in the middle of this cosmic archipelago, ostensibly ruling the cluster's dynamics. Having dined on its companions, this supermassive galaxy now lies in wait, patiently awaiting more "food" to fall into the gravitational grip of its three-billion-solar-mass black hole. The other, smaller galaxies swarming around in the outskirts of this and other galaxy clusters like it are almost surely destined to be someday integrated into the swelling central "beasts" at the heart of their evolving systems.

Nothing in this area of research is clear cut. The above ideas represent frontier thinking, which is itself evolving with each generation of astronomers. Puzzles abound at every turn: Some isolated elliptical galaxies reside in the "field" well outside clusters, which would seem hard to explain as the result of mergers. (Perhaps they have already gobbled up everything around them.) Spiral galaxies often populate the outskirts of galaxy clusters where encounters would seem to be rare and thus not conducive to the growth of spiral arms. (Perhaps they are in wide orbits about the cluster core, obeying Kepler's laws and spending most of their time far from the center.) And the irregular galaxies don't seem to fit into any evolutionary scheme—unless, ironically, they are the larger galaxies' building blocks staring us right in the face. (Perhaps those irregulars that still exist are the survivors, having so far managed to avoid extinction.)

Simply stated, owing to their distance and therefore their dimness, galaxies are hard to observe and the observations even harder to interpret. Many galactic secrets still lurk within them, awaiting new probes and new insights by future generations of astronomers eager to solve one of the great unresolved riddles in all of science—the origin and evolution of normal galaxies, abundantly and ubiquitously scattered through the Universe.

Evolutionary links between normal galaxies and active galaxies are more robust, though they, too, are hotly debated. A time sequence starting with quasars and proceeding to active galaxies and finally to normal galaxies, implying a continuous range of cosmic energy, has been bolstered in recent years. Adjacent objects along this sequence are almost indistinguishable from one another, meaning that all galaxies, regardless of type, might have similar "engines" at various stages of activity—such as supermassive black holes, which virtually all galaxies do seem to have at their cores. For example, weak quasars have some commonality with the most explosive of the active galaxies, whilst the feeblest active galaxies often resemble the most energetic members of the normal galaxies. Such a chain of cosmic verve suggests that galaxylike objects originated as quasars some twelve billion years or so ago, after which their emissive powers gradually declined, becoming active galaxies and eventually normal galaxies. This continuity among all galaxies has been strengthened recently as astronomers have become convinced that the black-hole energy-generation mechanism can account for the luminosities of quasars, active galaxies, and the central regions of most normal galaxies.

This unifying idea maintains that the quasars are actually ancestors of all (or most of) the galaxies. Consistent with the observed fact that quasars were more common in the past than they are today, galaxies do seem to have been more active long ago than they are now. Far too remote for us to resolve any individual stars within them, the quasars are detectable at great distances only because of their tremendously energetic central engines. Precisely because of their great distances, we perceive them as they once were in their blazing youth. As their core activity decayed with time, quasars assumed forms closer to those of more familiar and nearby galaxies. They essentially “wound down” while running out of fuel to feed their central black holes, eventually becoming the relatively quiescent normal galaxies now observed closer to us in space and time.

Should this view be proved correct, then maybe even our Milky Way Galaxy was once a brilliant quasar. Most ironic, if true. For decades, as-

A time sequence starting with quasars, then active galaxies, finally normal galaxies . . .

tronomers have struggled to decipher the Herculean quasars, especially their prodigious energy emission, only to find, perhaps, that we live inside an old, burned-out one—a time-tamed version of a quasar that once lit up the far away and the long ago.

For this quasar-evolutionary idea to hold, we ought to be able to see the vague outlines, however far away, of the more normal galaxies surrounding the quasars. Until quite recently, astronomers were hard-pressed to discern any galactic structure whatever in quasar images. However, the Hubble Space Telescope has done yeoman service since the mid-1990s by indeed finding "host" galaxies around some of the distant quasars. The evidence is in the form of very dimly glowing "fuzz" now seen to be faintly enveloping a few dozen of the brighter quasars studied to date. The quasars really do seem to be residents within the centers of normal galaxies, rich in ordinary matter beyond their bright cores; the fuzz is apparently the accumulated soft emission of innumerable unresolved stars or stars-to-be. Some of the deepest, long-exposure quasar images even show suggestive evidence for spiral arms.

Although attractive, this quasar → active galaxy → normal galaxy evolutionary sequence has its drawbacks. Not all astronomers have yet embraced the idea, arguing that evolutionary links may not exist at all. They suggest that the powerful quasars are merely extreme manifestations of the explosive phenomena seen in virtually all galaxies. After all, even the center of our own Milky Way is known to be expelling matter and radiation. The same can be said for active galaxies and quasars, though on vastly larger scales. Perhaps all these objects are part of the same family without there being any evolutionary sequence linking its members, just as evolutionary changes cannot be said to bridge different races within the human species. Each galaxy type or human race is distinctly different. One race of humans doesn't evolve into another, and similarly one type of galaxy might not necessarily evolve into any other. Instead, all the galaxies might be quite ordinary galaxies that formed long ago, though some were endowed with especially explosive central regions. Those able to exercise their explosiveness more than others for some still unknown reason are called quasars, while those hardly able to fire up their cores much at all are called normal galaxies.

Why the quasars emit radiation so prodigiously, even violently, is also unknown, though there is the notion that more fuel was available at earlier times. And for how long the quasars endure in their bright phase, adequately supplied with fuel, is also unknown; certainly they cannot do so indefinitely, lest their central black holes consume their whole being. The answers presumably lay buried within the relatively uncharted centers of galaxies, including the startling idea, now subject to heated debate, that quasars originally formed and regularly flare as supermassive black holes themselves merge, especially during the Galactic Epoch within a few billion years after the big bang.

Future research focused on the cores of galaxies will probably provide the best insights for deciphering the secrets of the bright and shining quasars, whose troubling properties of huge energy yet small size once threatened to topple the laws of physics. Even if their details are sorely lacking, their main issues now seem reasonably solved and the laws of physics intact. Rather than jeopardizing our knowledge of the cosmos, these violent objects have become an integral part of the thread of understanding that binds our own Galaxy to the earliest epochs of the Universe in which we live.

∞

Our knowledge of the galaxies, especially their origin and evolution, is inadequate. How each of them materialized, endowed with peculiar shapes and prodigious energies, remains largely unsolved. Their enigma is deepened by the fact that astronomers cannot find any galaxy unambiguously in the act of formation. Parts of all of them seem almost as old as the Universe itself, their youthful exuberance still beyond the clear reach of our best telescopes. Furthermore, even when galaxies do evolve, their changes are so agonizingly slow, compared to the duration of our technological civilization, as to make them appear immutable. If our understanding of galaxies seems sketchy, that's because it is; in some ways, galaxy research is only now coming into its own.

Currently, the origin and evolution of galaxies pose more problems than the formation of stars, which we can observe directly; than the evolution of stars, which we can decipher clearly; than the origin of life, which we can test in our laboratories; than the evolution of life, which we can study in action; even than the origins of intelligence, culture, and technology, all of which we can probe tangibly with fossils and artifacts unearthed from layers of historical rubble. Practically everything else discussed in this book is on firmer ground than the origin and evolution of galaxies. Exempt those eternal perplexities about the origin of the Universe itself, the subject of galaxy formation is the foremost missing link in the scenario of cosmic evolution.

Galaxies, though, are so very important. Apart from the creation of atoms, the formation of galaxies (perhaps along with some massive, extinct stars) was the first great accomplishment of the Matter Era. Until we learn a great deal more about how gravity leverages even slight initial gas irregularities into conspicuous density contrasts, our understanding of galaxy origins, and hence of cosmic evolution, will remain incomplete and unsatisfactory. Yet the promise is great, the potential payoff even greater. With the physicists unable to build accelerators on Earth sufficiently energetic to reproduce the earliest instants of time, it is the astronomers who, by studying the macrorealm of galaxies and their large-scale structure, are beginning to provide tests, albeit indirect ones, of the grand unification of particles and forces in the microrealm.

Astronomers now stand on the threshold of a golden age of galaxy research, much of which is a century or so behind stellar research if only because the galaxies are so dim and distant and therefore tricky to study. The equipment scheduled to debut during the early years of the new millennium will have greater sensitivity to collect more radiation as well as higher resolution to clarify the spread of that radiation, thereby almost surely advancing our knowledge of the origin and evolution of galaxies. Over the entire range of phenomena—from the earliest onset of density fluctuations in the primordial Universe, through the emergence of activity in the centers of galaxies, and on to the slow conversion of galactic gas into stars and planets—observations with novel instruments on the ground and in orbit are poised to provide a wealth of new and exciting data that hold clues to nothing less than some of the most profound and ancient cosmic secrets.

3. STELLAR EPOCH

Forges for Elements

STARS ARE GLOWING BALLS of gas, tenuous and hot on the outside, dense and hotter on the inside. Sized midway between the smallest and largest of all known objects, stars are bigger than atoms by roughly the same factor of a billion billion by which they are dwarfed by galaxy clusters.

Except for their shape, stars do not resemble hard, rocky planets in any way whatever. Normal stars are immensely larger and tremendously hotter than planets, and they experience changes in a completely different manner. They have no real surface, let alone any hard, solid matter as has Earth. Stars are simply composed of loose gas held intact by the relentless pull of their own gravity, which, at their cores, manages to compact that gas enough to trigger thermonuclear fusion.

This same gravity forces the gas to take on an austere geometrical configuration—a sphere. Wherever gravity dominates, it compels matter to adopt a round shape, which is Nature's minimum energy configuration for objects of sufficiently large mass. All the known stars, planets, and moons are spheres, or very nearly so.

Gravity is not the only force operating in stars. Otherwise, the inward pull of this all-pervasive force would shrink stars to such a small size that, as black holes, they could not radiate any heat and light. Competing against gravity in a star is the pressure of its heated gas, which, pushing outward, tries to disperse the star into space. The result is a structural balance, or stable condition: gravity in, pressure out. That's the simple prescription for a star—any star.

The star we know best is the Sun. Ole Sol is an average star whose properties lie in the middle of the observed ranges of mass, size, brightness, and composition for all known stars. Its very mediocrity is what makes the Sun so interesting to astronomers—it's "typical." Its proximity to us is also useful, allowing us to see this star "up close." Only eight light-minutes away, the Sun is some three hundred thousand times closer than our next nearest neighbor, the Alpha Centauri star system some four light-years away. Accordingly, we know far more about the Sun than about any of the other distant points of light in the Universe. Our parent star is a benchmark against which we compare many other objects in the cosmos.

Stars are fascinating for many reasons, though two prevail. First and foremost, stars play essential roles in the heating and lighting of any nearby planets. For us, the energy of our Sun is critically important not only for the origin of life on Earth, but also for the continued maintenance and further development of that life. Without a nearby star, Earth would be a frozen, barren wasteland—a boulder so unimaginably hostile that life as we know it could not possibly exist.

Second, stars are the furnaces where heavy elements are forged. Colliding viciously, the light nuclei of hydrogen and helium fuse into the more complex nuclei at the cores of more than a hundred chemical elements such as carbon, nitrogen, oxygen, silicon, and iron. "Better living through chemistry" is surely a theme (and not just an industrial slogan) prominent later in our cosmic-evolutionary scenario, and the building blocks of chemistry began in the stars. Without the heavies, nothing around us—not the ground, not the air, not much of Earth itself—would exist.

Stars, then, are key in the evolution of both matter and life; they might be absolute prerequisites for any kind of life. Stars themselves participate in the great, ongoing process of change in the Universe, the so-called stellar evolution involving aspects of adaptation and selection—though not as dramatically so as for the biological evolution of life-forms encountered later in the Biological Epoch. "Generations" of billions of stars were "born," have "lived," and have "died" since our Galaxy originated some twelve billion years ago. By forming in galaxies everywhere and by providing heat, light, and heavy elements for planets and life to follow, stars of this Stellar Epoch segue nicely from the previous Galactic Epoch to the next Planetary Epoch. Stars are a pivotally important, integral part of the cosmic-evolutionary story.

During the second half of the twentieth century, astrophysicists learned a great deal about how stars pass through phases of youth, maturity, and aging—through developmental and evolutionary paces. Although they appear immutable in the nighttime sky—secure in their remote and steady brilliance night after night—stars do actually change their appearance over great durations of time. The giant red star Betelgeuse in the constellation Orion and the dwarf-white companion to Sirius, the Dog Star, among myriad yellowish stars like our Sun, are not really different types of stars. Rather, each is at a distinct stage in the changing "life cycle" of nearly all stars.

As with all aspects of evolution broadly considered, change is central, though among the stars it is woefully slow. Some stars are old and bloated, some young and luminous. Others are long gone, having literally run out of fuel and perished eons ago. Still others are only now emerging from the interstellar hodgepodge of surging gas and dust—those vast and dark regions in and among the stars of our nighttime sky. Much of this change is unobvious because stellar life cycles are astronomically longer than human life spans—often billions of years compared to hardly a hundred, or millions of millennia compared to less than a century. Throughout all of evolution, we usually see change in snapshots—here and there in Nature, quick and sometimes dirty—each nonetheless helping us depict that broad and big picture of natural history writ large.

∞

Stars are of fundamental scientific significance. Few things in science are more basic than the way stars shine, or in astronomy than the way they form. Do astronomers really understand the origin of stars? Can we describe the specific formative stages in the changing galactic matter that eventually produce stars? What about the evidence—is there any experimental support for our ideas? Questions like these are now under intellectual attack at observatories around the world. The answers are not yet crystal clear, but remarkable progress has been made in the past few decades. As it stands now, our knowledge of star formation is a robust combination of theoretical insight and observational fact.

The Stellar Epoch offers a better description of matter on scales smaller than galaxies than does the Galactic Epoch on scales larger than galaxies. In other words, we know much more about the origin and evolution of stars than we do about galaxies. Gravitational instabilities, invoked with partial success for galaxies, can be modeled more effectively to understand the formation of stars within those galaxies, regardless of how the galaxies themselves arose. Much as was the case in the early Universe, chance mixes with necessity to affect change. At the outset, random fluctuations often occur at various parts of large gas clouds within any already-formed galaxy. Although such chancy fluctuations alone prove insufficient to cluster huge amounts of matter into galaxies, calculations imply that the process should work much better—and more quickly—to assemble smaller clumps of matter into stars. Swirling eddies of loose gas in interstellar space are cooler and denser than those of the primordial fireball, hence well suited to collecting enough matter to mold individual stars or groups of stars, after which they contract, heat, and eventually ignite their nuclear fires.

Astrophysicists have built intricate models of the stages through which gas clouds evolve to become genuine stars. These models, like those of the early Universe, are essentially "number-crunching experiments" performed on powerful, high-speed computers. But here, in the Stellar Epoch, we have many data with which to test the models. The computational factors include mass, heat, rotation, magnetism, elemental abundances, and a few other physical conditions typifying a changing interstellar cloud. These factors resemble the ingredients of an elaborate recipe, yet in this case the recipe is mathematical and teems with symbolic equations. And as is true for any new recipe, although the types of ingredients are known, the amounts of each are often uncertain.

Huge computer programs, built during the past twenty years and containing as many as a million lines of code, enable theorists to use trial-and-error routines for this multifaceted problem of star formation. Though computers do nothing more than calculate numbers rapidly, they can do this basic task more agreeably than humans, adjusting and readjusting the many ingredients to best match the theoretical predictions of the models with the observational findings of actual stars in the Milky Way.

The accuracy of these models is presently mediocre, for it's tricky to take that third step of the scientific method and test them experimentally. To stress an oft-repeated quandary, no one has ever seen an interstellar cloud or a genuine star parade through all of its evolutionary paces. The lifetime of a human being, or even the duration of our civilization, is very much shorter than the time for a cloud to contract and form a star. Since about thirty million years (or about a million human generations) are needed to concoct a star such as our Sun, no one person can realistically expect to observe any celestial object proceed through its full pageant of star birth.

The stellar models are not without observational support, however. Telescopic monitoring of various gas clouds at many stages of their evolutionary trek helps refine our knowledge of star formation. Modern technology enables astronomers to peek at interstellar clouds and nascent stars for hints and clues about their embryonic development. Studies of invisible radio and infrared radiation emitted by cool, tenuous galactic regions have proved especially useful, though we are still learning to grope in the dark where young stars emerge; like mammals, the bright stars incubate in total darkness. By studying numerous interstellar clouds, often at unrelated places along the Milky Way, we can now piece together an observational understanding of many key stages of prestellar evolution.

Current efforts of astronomers and astrophysicists resemble those of anthropologists and archaeologists, who unearth bones and artifacts at many unrelated locales strewn across our planet's surface. Not having lived at the time of our ancient ancestors, the social scientists sift the scattered rubble and ponder the myriad remains, trying to decipher how all of it can be pieced together into an overall mosaic of human evolution. Likewise, space scientists observe a panoply of celestial objects in many disparate parts of our Galaxy, seeking to fathom how each one fits into the larger scheme of stellar evolution. The terrestrial bones and extraterrestrial objects are much like segments of a puzzle. The picture becomes clear only when each piece is found, identified, and fitted properly relative to all the others.

∞

Imagine a large plot of interstellar real estate somewhere in the Milky Way. By definition, *inter*stellar matter is that which exists beyond each of the stars—in short, matter scattered throughout the black and vast expanses among the myriad stars in our nighttime sky. Most people think nothing exists there, for, sure enough, a clear night shows only darkness among all the minute points of glowing starlight. But the darkness of outer space only affirms the limits of our human vision.

Not much matter resides in any one interstellar region, but some most surely exists. Interstellar matter is a trillion trillion times less dense than that in either stars or planets, in fact thinner than the best vacuum achievable on Earth by pumping all the air out of a cylinder. (Such laboratory vacuums still contain some million atoms in each cubic centimeter.) Even so, interstellar space is so huge that small amounts of matter here and there can accumulate to play a significant role. It's not unlike the prospect of becoming a multimillionaire by collecting a mere penny from every person in North America. Even minute quantities can add up to extremely large amounts, given enough space and time. All told, roughly as much mass resides in the immense realms of interstellar space as in the stars themselves.

The interstellar medium, then, includes the mostly invisible and rarefied regions from which all stars arise at birth—in any galaxy. In our own Galaxy, it forms a disk nearly a thousand light-years thick that extends for the full hundred-thousand-light-year width of the Milky Way. We also now realize, as noted later in this Stellar Epoch, that interstellar space is the very same domain into which many stars explode at death. It's one of the busiest crossroads through which matter passes anywhere in the Universe.

Interstellar matter is largely a mixture of gas and dust. Much of the gas is made of thinly dispersed atoms (mostly hydrogen, H, and a little helium, He), though frequently clusters of atoms—molecules (mostly diatomic hydrogen, H_2)—are evident. The interstellar gas density averages a single atom per cubic centimeter, except in those places where it clumps into richer groups of atoms sometimes reaching a thousand to a million times greater density. And it is in those denser "clouds" that interesting things happen, such as star formation. In general, though, the interstellar medium is so sparsely populated that harvesting all the gas in a region the size of Earth would yield barely enough matter to make a pair of dice.

As thinly spread as is the gas, interstellar dust is even more so; only one dust particle lurks in the darkness for every trillion atoms of gas. That's much like a single dust grain residing in a volume of interstellar space equivalent to that housed in the New Orleans Superdome. By dust, we mean solid particles made mostly of heavy elements, not terribly unlike the fine chalk dust that settles on blackboard ledges or domestic dust that lurks under beds and in closets; tiny particles in a terrestrial fog or cigarette smoke might be even better examples. The dust was, and still is, probably manufactured in the cool, outer atmospheres of old stars. Still, the vastness of space grants dust a role; an imaginary cylinder one square meter in cross section and extending from Earth to Alpha Centauri would contain more than ten billion billion dust particles.

We can also think of the dust in this way: by enlarging a solid dust particle about a billion times (or by collecting a billion of them in one place), it might resemble a rocky asteroid; a trillion times, perhaps the core of primitive Earth. Small parcels can accumulate impressively in realms as expansive as the Milky Way.

Despite their rarity, dust particles make interstellar space a relatively dirty place. If we were able to capture such a parcel of interstellar matter and compress it to the typical density on Earth, the resulting gray fog would be so thick that we wouldn't be able to see our hands in front of us. Pound for pound, space is heavily "polluted" with dust, but that dust is normally sprinkled throughout enormous tracts of galactic territory. By comparison, Earth's atmosphere is about a million times less dusty. So, place humanity's pollution problems into perspective; compared to the Galaxy in general and on a fair scale, Earth is a relatively clean place.

If the gas and dust of interstellar space had remained evenly dispersed forever, neither stars nor planets, and certainly not life, would have ever formed. The sky would be absolutely dark and no one would exist to know it. Fortunately, the interstellar medium is not immutable. Like everything else, it changes its disposition.

Theory suggests that matter contained within the dark regions of space will naturally fluctuate in density and eventually fragment into larger clumps typically spanning tens to hundreds of light-years. Because these dark regions are just that—dark—they have always been difficult or impossible for astronomers to visually inspect. Quite frankly, there's literally nothing to see in a dark region—which, by the way, partly explains why humankind has been, until relatively recently, virtually ignorant about star formation since the birth of astronomy thousands of years ago.

Dark and dusty regions of interstellar space are inaccessible to study by optical means; they simply emit no light. Even stars behind these regions are invisible because dust diverts their radiation from reaching

Earth, hopelessly scattering it like automobile headlights in fog. That doesn't mean that the murky galactic recesses are totally impenetrable, however. Marvels of modern technology, such as parabolic-dish radio telescopes and heat-seeking infrared satellites, permit the sampling of invisible regions for their long-wavelength emissions, which are able to penetrate the debris of interstellar space. In the same way that soldiers use infrared sensors to locate the enemy at night, and for the same reasons that airport radars operate properly in the worst winter weather, infrared and radio astronomers can detect invisible radiation from the utter darkness of interstellar space.

Analysis of the radiation emitted by interstellar matter has now confirmed theoretical predictions that parts of the near-void among the stars of any galaxy are clumped into large gassy clouds. Their overall morphology tends to resemble the irregular, fluffy clouds of Earth's atmosphere, but there the resemblance ends. Interstellar clouds are billions of times larger than the entire Earth. They also amass and disperse, that is, come and go, billions of times more slowly than terrestrial clouds.

Radio and infrared observations have proved that interstellar clouds are not only tenuous but cold as well, often containing no more than a hundred atoms per cubic centimeter at temperatures hovering close to absolute zero. This density, though enhanced somewhat due to a cloud's bulk, is still extremely low, in fact still lower than that of the best vacuums attainable in physics laboratories around the world; for comparison, the normal density of air on Earth is more than a billion billion atoms per cubic centimeter. Typical cloud temperatures, some −250 degrees Celsius, are also extremely low, for the lowest possible temperature (at which atomic motion virtually ceases) is −273 degrees Celsius. We can thus fairly visualize an interstellar cloud as a wispy, frosty entity, but even that is an understatement.

Now imagine a small portion of an interstellar cloud, for instance a parcel of gas and dust calved from a larger cloud and much less than a light-year across. Given the cloud's flimsiness, such a parcel doesn't house many atoms. Yet unless the cloud is as cold as physically possible, each atom will still have some random motion owing to its heat, however minute. And each atom will be slightly influenced by the gravitational force exerted by all the other neighboring atoms, however small the mass of each atom. If only a few atoms coalesced accidentally for a moment, their combined gravitational pull would be insufficient to bind them permanently into a distinct clump, which would then disperse as quickly as it formed. The effect of heat, even for the frigid interstellar atoms, wins this battle with gravity.

Suppose we now widen our sights to include more than just a few atoms. Instead, consider fifty, a hundred, or even a thousand atoms. Would a group of that many atoms exert a net gravitational force strong enough to prevent the clump from dispersing as in the previous example? Just how many atoms are needed for gravity to bind them into a tight-knit assembly?

Answers to these questions cannot be obtained from a simple study of gravity alone; nor can they be found among the solutions at the back of any science textbook. Correct solutions depend not only on gravity but also on several other physical factors noted earlier such as heat, rotation, magnetism, and turbulence. These additional agents tend to influence the evolution of an interstellar cloud, for, although they should not be regarded as antigravity, they do compete against gravity.

Take heat, for example. Most of the slight warmth of interstellar clouds derives from rare yet inevitable collisions among the atoms. More frequent collisions mean greater friction and thus more heat, just as rapidly rubbing our hands together generates more warmth than doing so sluggishly. Heat gives a cloud of gas some buoyancy that tends to offset gravity. Heat is, in fact, the main reason that the Sun doesn't collapse; the outward pressure of its heated gas counteracts the inward pull of its gravity. The amount of heat contained within an interstellar cloud is, of course, small by solar standards—which is why bright stars are lit up and dark clouds are not. Consequently, thermal effects that compete strongly with gravity once stars form do not really play a large role until after interstellar clouds contract and become hotter.

Rotation—that is, spin—can also compete with gravity. A contracting cloud having even a small spin tends to develop a bulge around its midsection. This bulge is a sure sign that some of the matter is trying to defy gravity and thus disperse. As the cloud compresses on its way to becoming a star, its spin necessarily increases, just as a figure skater rotates faster with her arms retracted—a prescriptive principle of physics known as "conservation of angular momentum." Any rapidly rotating object exerts an outward force; the faster the spin, the greater the force, as anyone can feel while bearing the brunt of a circular ride at an amusement park. In the case of an interstellar gas cloud, atoms near its edge are especially vulnerable to escape if the pull of gravity is insufficient to retain them. Should a contracting cloud increase its spin so much that gravity can no longer bind it, then the cloud would simply disband, releasing its atoms back into interstellar space. Mud flung from a rapidly spinning bicycle wheel is a good example: outward forces dominate any surface tension tending to keep the mud on the wheel. The only way an interstellar cloud can preserve itself against the threat of dissipation via rotation is to gather more and more atoms, thereby increasing its collective strength of gravity. The upshot is this: rapidly rotating interstellar clouds need more mass to guarantee continued contraction toward starlike objects than do clouds having no rotation at all.

Magnetism, turbulence, and several other physical effects can also hinder the contraction of a gas cloud. Magnetic forces permeate interstellar clouds, much as they do more strongly the Sun and Earth; in all these cases, the magnetism probably arises from the motions of charged particles. Gas turbulence, or disordered bulk motion, is also present within each cloud, yet nearly intractable mathematically; turbulence is possibly the result of collisions among clouds over eons of time or shock waves pummeling the clouds as cosmic rays from exploded stars plow into them. Observations made during the past few decades show that most interstellar clouds are very cold, spin slowly, and are only slightly magnetized and turbulent, so individually these factors shouldn't amount to much. But theory suggests that, taken together, even small quantities of each of these agents sometimes unite to compete effectively with gravity.

So it's not a simple case of gravity sweeping up matter to build a star. Many additional factors serve to complicate the process, making star formation challenging to understand in detail. The upshot is that even those clouds that do manage to contract often do so in highly distorted ways, greatly altering the dynamical behavior and subsequent evolution of a typical gas cloud.

We now return to our original question: How many (hydrogen and helium) atoms need to accumulate for the collective pull of gravity to prohibit a pocket of gas, once formed, from dispersing back into the surrounding interstellar environment? The answer, even for a cool cloud having no rotation or magnetism, is a very large number. In fact, nearly a thousand billion billion billion billion billion billion (i.e., 10^{57}) atoms are needed for gravity to bind a gaseous condensation. There's no doubt about the truly huge magnitude of this number. It's much larger than the number of grains of sand on all the beaches of the world

($\sim 10^{22}$), even larger than the million billion billion billion billion billion (i.e., 10^{51}) elementary particles in all the atomic nuclei throughout the entire Earth. It's large compared to anything with which we are familiar because there's simply nothing on Earth comparable to a star.

This number, 10^{57} atoms, just about equals the mass or our Sun—which is no coincidence. Our Sun is an ordinary, average star (if a little on the small side), implying that most stars form from galactic fragments having approximately this number of atoms. In all, stars originate from slightly larger and smaller clumps, for the range of known stars varies from about one-tenth to one hundred times the mass of our Sun—a rather small variation in astronomical terms, but a variation nonetheless among populations of stars.

Spanning ten to a hundred light-years, typical interstellar clouds usually harbor thousands of times more matter than in normal stars. Observations prove this, especially as regards the so-called giant molecular clouds, which dwarf all other "objects" in the Galaxy (well, save the possible black hole and its accretion disk at the core of the Milky Way). If these clouds are to become the birthplaces of stars, they cannot remain homogeneous blobs. The clouds must gradually break up into smaller parcels, often less than a light-year across. Theory suggests that fragmentation into subunits occurs naturally because of inherent gravitational instabilities; the result is localized inhomogeneities in the gas. A single such cloud can therefore divide into tens, even hundreds, of fragments or clumps, each imitating the shrinking behavior of the larger, parent cloud as a whole, albeit contracting even faster than the bigger cloud in which it's embedded.

In fact, astronomers have no evidence that stars are born in isolation, one star from one cloud. Some interstellar clouds end their long evolutionary trek either by forming several stars each much larger than our Sun or clusters of hundreds of stars each comparable to or smaller than the Sun. In reality, most clouds give rise to a whole family, or population, of stars, smaller ones outnumbering larger ones (much as at the seashore small pebbles far outnumber larger boulders). Perhaps all stars originate as members of groups. Those now appearing alone and isolated in space, such as our Sun, probably wandered away from the rest of their litter, though only after all were fully formed.

Once a fragment assumes its own identity within an interstellar cloud, it passes through a series of inevitable stages. It first begins to contract as gravity grows with the ever-accumulating group of atoms; the fragment literally shrinks under the stress of its own weight. Simultaneously, as the density rises, the atoms collide more frequently, in turn causing the increasingly compact fragment to warm steadily.

By the time a typical self-heating fragment has shrunk to about a tenth of a light-year—which is yet hundreds of times the size of our Solar System—its temperature has risen to nearly zero degrees Celsius. That's still colder than our twenty-degree-Celsius room-temperature standard on Earth, but it's a lot warmer than the original interstellar cloud prior to its clumping. This individual, gaseous blob has begun the long, formative trek that will ultimately produce a star. However, it must change still further, reorganizing itself into a smaller, denser, hotter object, before it can be rightfully called a genuine star.

Our description is more than a theoretical scenario. Its rough outline has now been clearly, though not visually confirmed, using specialized equipment developed during the past quarter-century. Radio and infrared observations have produced solid evidence that huge interstellar clouds are, in fact, fragmenting into smaller clumps of gas. Pockets of slightly hotter and denser matter within otherwise tenuous, cold, and enormous clouds are now known to be the rule rather than the exception.

Fragmentation might be expected to continue indefinitely, dividing again and again and ultimately yielding ever-smaller clumps impossible to form stars. Fortunately, the process halts before it's too late. Rising gas density stops the process of fragmentation from reducing all parts of the cloud without limit, lest the cloud become homogeneous again. As individual fragments compress their gas, they eventually become compact enough to prohibit radiation from easily escaping. With the cloud's natural vent partially blocked, the trapped radiation causes the temperature to rise, pressure to increase, and fragmentation to cease.

As each gas clump continues to evolve, computer models predict much the same story: the fragment's size diminishes, its density grows, and its temperature rises at both its core and periphery. Several tens of thousands of years after it first began contracting, a typical fragment's dimensions will have become comparable to those of our Solar System, a size still ten thousand times larger than our Sun. Core temperatures at this stage will have reached many thousands of degrees Celsius, values greater than those inside the hottest steel furnace built by our civilization on Earth.

Roughly a million years later, the full expanse of an interstellar fragment could fit within Earth's orbit around the Sun. This increasingly structured object is now looking less like an irregular fragment and more like a round blob. Its core temperature has steadily mounted to nearly a million degrees Celsius, and that manifests as a warming surface that now begins to glow. Although charged elementary particles, ripped from disintegrating atoms, are whizzing around inside, they are still too sluggish to overcome their natural electromagnetic repulsion in order to penetrate the smaller realm of the more powerful nuclear force. In other words, the blob's matter is still far from the ten million degrees Celsius needed to initiate the nuclear fusion that will one day transform this gaseous heap into a bona fide star. Nonetheless, the hot, dense object at this stage resembles a star closely enough to merit the special name protostar—an embryonic, glowing blob perched at the dawn of star birth.

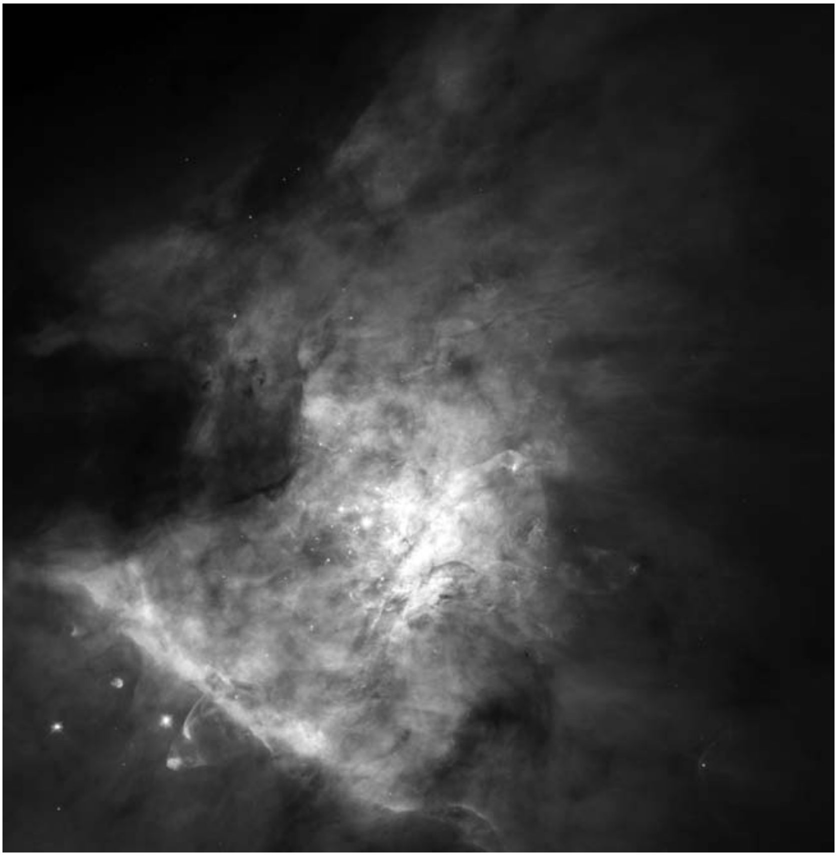
Theoretical modeling aside, is there any observational evidence that hot, dense fragments have Solar System dimensions? Indeed there is. Within the past couple of decades, radio and infrared telescopes have captured radiation emitted by small clumps at or near the cores of many cloud fragments. The diameter of each clump is hardly more than a thousandth of a light-year, or just about the size of our Solar System. Their total gas densities, inferred best from the radio observations, reach nearly a billion particles per cubic centimeter. And their temperatures have been measured by infrared techniques to be many hundreds of degrees Celsius. Most experts agree that these dense, warm blobs are real protostars—large gassy balls about the size of Mercury's orbit and poised on the verge of stardom.

In a few such cases, especially in some of the better-known star-forming regions several thousand light-years away, such as the Eagle and Trifid nebulae, astronomers have been able to detect and monitor the dynamical effects of infalling clumps. Not that the act of infall is directly seen, for the timescale typical of cloud contraction is millions of years—well longer than the age of humankind. But, by studying Doppler-shifted radio radiation (which is a velocity diagnostic) arising from such suspect regions, we can clearly perceive cloud fragments well on their way to becoming stars—or, more likely, clusters of stars.

The hunt for protostars themselves shifts to the infrared part of the spectrum, since these objects should have more heat than their ancestral clouds. Some do indeed stand out as fuzzy little telltale sources of heat, often deep inside cocoons of dust that hide protostars from direct optical view. What's more, protostars often exhibit strong "winds" and "jets" of fast-moving gas expanding outward at high velocities, typically tens of kilometers per second; imagery often reveals the bipolar jets billowing from adolescent stars and churning up the surrounding interstellar medium. The jets, which sometimes extend for a few light-years, emanate perpendicular to much smaller disks where planets are probably beginning to form. Such outflows resemble the vastly larger lobes of hot plasma seen near active galaxies, such as the quasars as noted earlier in the Galactic Epoch, and are yet another way that young stars "manage" their energy budgets, all the while ridding themselves of excess energy. Eventually, such a star breaks through its placental envelope, its winds blow away much of the disk, and henceforth its energy is emitted less in the form of twin jets and more as a normal, uniformly bright, visual star, such as our Sun today.

Notably, but only in a few of the closest stellar nurseries such as the Orion Nebula some fifteen hundred light-years distant, circumstellar disks have recently been spotted. Big telescopes are needed to resolve the details, yet even a naked-eye amateur can spot this fuzzy nebula at the business end of the "sword" hanging below the "hunter's belt" of three blue-white stars strikingly aligned in this winter-sky constellation. There, relatively new stars by the dozens, including four notable ones called the Trapezium group and viewable with good binoculars, glow intensely, many of them as young as a hundred thousand years. At higher magnification—though well beyond that discernible with even the best binoculars—thin little oblong smudges are seen on digital images all through the Orion region, each one apparently a dirty disk. Again, their size and scale resemble those of our Solar System, much as expected for protostars and possibly planets emerging from the turbulent mishmash of interstellar gas and dust.

Some protostellar objects emit intense, highly focused radiation, much of it coming from small molecules containing two or more atoms linked together. The radiation is especially intriguing because of its terrific strength and localized source. The first observations, a few decades ago, of radio radiation from one such blob were so mysterious that puzzled astronomers began calling the emitting molecules "mysterium." They were later identified properly as hydroxyl (hydrogen plus oxygen, OH) molecules, and their enormously powerful signals are now known



Evidence of stellar birth.

Glowing regions of hot, thin plasma, gaseous nebulae are signposts of star formation. Here, the Orion Nebula, some fifteen hundred light-years away, is the closest of the rich stellar nurseries near the disk of the Milky Way. Itself several light-years across, this nebula is embedded within a much larger, invisible molecular cloud extending hundreds of light-years. Observations have shown that the nebula houses scores of young stars in addition to hundreds of new protostars. *Source: European Southern Observatory/Space Telescope Science Institute.*

to be enhanced or amplified by a special "maser" process a little like that found in "lasers" on Earth.

The word *laser*, the name of common everyday devices in supermarket checkout counters and compact-disk players, is actually an acronym for *l*ight *a*mplification by *s*timulated *e*mission of *r*adiation. Lasers are artificial gadgets that emit concentrated streams of light radiation in very narrow beams. Only within the past few decades has our civilization become smart enough to build such tools, relying as they do on both advanced technology and a good understanding of atomic and molecular physics. Lasers operate by exciting atoms or molecules in a gas and then stimulating them to emit radiation simultaneously. In this way, a tremendous burst of radiation can result, much more powerful than that from an ordinary light bulb.

Masers are similar to lasers, except that they produce *m*icrowave (a special type of radio) radiation rather than optical light. Physicists know how to build them in terrestrial laboratories, though masers are very delicate machines, requiring special conditions and much patience to operate. When working properly, they are the best amplifiers known, much more effective than the ordinary transistors that do yeoman service in our personal computers and household appliances.

Interestingly enough, certain regions of interstellar space are naturally suited to produce amplified microwave radiation. Protostellar blobs apparently enjoy the special conditions required, first, to excite some molecules and, second, to stimulate them to emit intensely. The blobs' warm temperatures and moderate densities seem ideal for this unique emission mechanism. Accordingly, the intense maser radiation observed from certain molecules—not just hydroxyl, but also water vapor and a few others—can be analyzed for additional clues about protostellar regions. Such studies frame one of the most exciting areas of contemporary astrophysics.

The very fact that molecules populate the near void of interstellar space is remarkable. Harsh radiation and alien environments would clearly compromise the molecules' existence unless they were protected, which is probably why they are invariably found in and around the dark, dense, and dusty parts of space. These are the giant molecular clouds again, our Galaxy's largest entities that overwhelm in both size and mass, indeed often fully engulf, even the biggest nebulae, such as one known as Sagittarius B2, not far from the Galaxy's center. About a thousand such molecular clouds are currently known in the Milky Way, some of them millions of times more massive than our Sun. Ironically, the minute dust grains within those huge clouds not only serve to shield the fragile molecules but also likely act as catalysts to help form them. The grains provide both a place where atoms can stick and react as well as a means of dissipating any heat generated by the reaction, which might otherwise destroy the newly formed molecules. Even so, the details of this new frontier subject—astrochemistry—are still subject to debate and testing. After all, textbooks entitled *General Chemistry* are not general at all but are really books on terrestrial chemistry familiar to us on Earth. The truly "general" or universal chemistry texts are only now being written based on astronomers' findings in the wider extraterrestrial domain where extraordinarily low temperatures and densities prevail quite unlike anything naturally on Earth.

During the past few decades, well more than a hundred different types of molecules have been detected in spectra (which display radiation's "fingerprints"), many of them in the particularly rich Sagittarius B2 area. Given that they often radiate long radio waves, which can penetrate dust, the molecules act as important tracers of a cloud's structure and physical properties as well as its chemistry. Carbon monoxide (CO), ammonia (NH_3), and water vapor (H_2O) are especially ubiquitous. Most intriguing, a pharmaceutical array of rather complex organic (carbon-rich) molecules has also been discovered in the darkest and densest of the molecular clouds, such as formaldehyde (H_2CO, a popular cleaning fluid and preservative), formic acid (H_2CO_2, prominent in ants and other insects), ethyl alcohol (C_2H_5OH, or galactic booze), and cyanodecapentayne ($HC_{11}N$, a thirteen-atom molecule not naturally found on Earth). Their presence has fueled speculation about life having originated in interstellar space, especially since a report by radio astronomers in the mid-1990s (still unconfirmed) that deep space harbors glycine (NH_2CH_2COOH), which is one of the key ingredients of protein molecules in living cells. The likelihood that such precursors of life, indeed perhaps life itself, could have formed outside Earth makes for another exciting interdisciplinary subject—astrobiology—to which we shall return in the Chemical Epoch. Organic molecules in space are harbingers of objects of greater complexity—to be sure, much greater complexity—yet to come in this book.

The very existence of interstellar molecules has forced astronomers to rethink and reobserve the vast realms well beyond Earth. In doing so, we have begun to realize that this active, fertile domain is far from the void suspected by theoreticians not so long ago. Regions of space recently thought to contain nothing more than galactic "garbage"—the empty-looking darkness among the nighttime stars—now play a critical role in our understanding of the interstellar medium in which stars, and at least the building blocks of life, are born.

We are not done with protostars, which in this recounting have not yet become real stars. Theory suggests that protostars should be a bit unstable, as their inward pull of gravity doesn't quite balance their outward push of hot gas pressure. The temperature is still too low to establish that "gravity-in, pressure-out" equality that guarantees stability—fortunately for us, since if the heated gas managed to counter gravity before reaching the point of nuclear burning, there would be no stars. The nighttime sky would be fully abundant in dim protostars, though completely lacking in actual stars. And it's likely that neither we nor any other intelligent life-forms would exist to appreciate this distinctly duller Universe.

Computer models predict that as protostars continue to follow the dictates of gravity, the gas has no choice but to contract more, alas ever so slightly. The result is renewed heating. But even after a thousand centuries of infall, and with a core temperature of several million degrees Celsius, not enough heat has yet built up to initiate nuclear fusion. Only when the temperature deep down in the core reaches fully ten million degrees Celsius do the nuclear reactions commence. Atomic nuclei then have enough thermal energy to slam violently into each other and to overwhelm their own mutual repulsion by means of the very same process described earlier for the transformation of hydrogen into helium during the Particle Epoch. The upshot is more energy released and a halt to the contraction. A genuine star has finally formed, its principal function thereafter being the consumption of hydrogen, thereby producing helium and especially energy.

So, even as the average cosmic density and temperature continue to decline with the expansion of the Universe on the largest scales, small, localized "islands" called stars arise wherein their densities and temperatures increase. Stars buck the cosmic trend of decreasing temperature and density; they also go against the tendency of the Universe to become more disordered, for stars are clearly sites of rising complexity and greater order, especially as their thermal and elemental gradients steepen with time from core to surface. These, in turn, accompany increased energy flows, but again, only locally where stars reside. Such energy flows are likely key to the emergence of order and structure in the Universe.

Hearts of stars, then, are sites where atomic nuclei viciously collide, interpenetrating the realm of the nuclear force, thus releasing copious amounts of energy. Contrary to popular opinion, it's not the nuclear reactions that create the high temperatures in stellar cores. Rather, it's the high temperatures that allow the nuclear reactions to proceed there.

Only with a sufficiently high temperature—that's the ten-million-degree Celsius threshold just noted—can the positively charged proton nuclei of hydrogen get up sufficient speed to ram into one another fiercely enough to allow the nuclear force of attraction to overcome the electromagnetic force of repulsion. As with many large-scale phenomena in the cosmos, it's gravity that triggers this change—a change that causes fusion to literally light up the star—for the needed temperatures rise only because the infalling cloud manages to convert some of its gravitational potential energy into frictional heat.

Once fully formed, a star becomes a prodigious emitter of radiation. Every second, our Sun fuses six hundred million tons of hydrogen into helium, converting the equivalent of more than four tons of matter into pure energy according to that same simple yet profound equation, $E = mc^2$. In gee-whiz terms, the Sun releases, again each second, an amount of energy equivalent to the detonation of about trillion atomic bombs. That's more energy than humans have generated in all of history and the Sun does it *each second.* In fact, that's enough solar energy, if suitably focused, to evaporate all of Earth's oceans in about six seconds, or melt our planet's crust in a mere three minutes. Fortunately, the energy from the nuclear inferno moves up through the interior of a star and is radiated isotropically, unfocusedly, and equally from all parts of its surface in the guise of ordinary starlight.

All the starry points of light seen in the nighttime sky owe their existence to nuclear fires churning deep in the cores of each and every one of them. Ponder all that astronomical activity, that sheer cosmic power, while looking upward at the stars some clear, moonless evening. Even the nocturnal quiescence is home to continual change, the black heavens above pierced by the myriad, brilliant signposts of stellar birth.

The remarkable change from galactic cloud to contracting fragment to protostellar blob to nascent star takes a few tens of millions of years. Obviously a long time by human standards—in fact, tens of thousands of millennia—this is still less than one percent of a typical star's lifetime. The entire process amounts to a steady metamorphosis—an evolution of sorts—a gradual transformation of a cold, tenuous, flimsy pocket of gas into a hot, dense, round star. The prime instigator in all this stellar evolutionary change is, once again, gravity. And the net effect of it all is increased energy flow and rising complexity.

∞

Once heat and gravity are balanced, a star like our Sun is stable. It experiences "storms" at its surface in the form of flares, spots, and prominences, but these are minor irritations for an object as large as the Sun (though perhaps not so minor for any nearby planets). The star's main agenda is to produce energy steadily for some ten billion years. A combination of theory and experiment implies that the Sun has already done so for about half this duration. So "our star" can be regarded as middle-aged, a celestial body expected to burn literally morning, noon, and night for another five billion years into the future. (Its total lifetime as a star, all the way through the red-giant and white-dwarf evolutionary stages explained below, is projected to be more like twelve billion years.)

Stars smaller than our Sun take more time to form from interstellar matter. They also last longer while fusing more slowly; they resemble efficient compact cars in that they carry less fuel but burn it more effectively. For example, stars having one-tenth the Sun's mass require nearly a billion years for birth and should endure for as long as a trillion years. Since the latter value is much longer than the current age of the Universe, all small stars that have ever formed must still be fusing hydrogen into helium, producing a constant flux of energy for the benefit of any attendant planets.

By contrast, stars larger than our Sun tend to form faster from interstellar clouds, some in as little as a million years. The more massive stars in fact seem to do everything at a quickened pace. They burn their hydrogen fuel more rapidly, and they pass through all their evolutionary paces more quickly. The reason is that their greater masses gravitationally compact the big stars more strongly, causing matter within them to collide more frequently and violently, which, in turn, hastens their nuclear reactions. As a result, and somewhat surprisingly despite their huge masses, the biggest stars endure for much less than the ten-billion-year lifetime of our Sun. The most massive ones, for example, are nearly a hundred times the mass of the Sun, yet last for only about ten million years. They expend their stability with great flurry, a mere wink of an eye on the normal scale of cosmic lifetimes; they resemble gas-guzzling cars that carry more fuel but burn it less efficiently. Alas, a quickened pace is not always a desirable one, for the big stars live fast and die young. Just as it's unhealthy for humans to rush through life, the largest stars hardly seem to settle down at all. In the end, while small stars shrivel up and fade away, stars more massive than our Sun perish by catastrophically collapsing and then exploding. Apparently some clichés have universal applicability: the bigger they are, indeed the harder they fall.

Other objects worth noting are the "failed stars," for some cloud fragments never do achieve legitimate stardom. The planet Jupiter is one such case, having contracted under the influence of gravity and heated up somewhat, yet not having enough mass for gravity to crush its matter to the point of nuclear ignition. With only about a thousandth the Sun's mass, Jupiter never evolved beyond the protostellar stage. Space might well be heavily strewn with such compact, dark "clinkers" of unburned matter frozen in time.

Theory holds that some objects having at least a dozen Jupiter masses might be able to ignite a special form of hydrogen, namely the isotope deuterium, which has a proton and a neutron, but only for short periods of time. Deuterium is generally present only in trace amounts in any celestial object, and this minimal fusion process stops as soon as it's gone. These objects, called brown dwarfs, are distinctly more massive than Jupiter but a good deal less massive than the Sun. In fact, nearly a hundred Jupiter masses (or a tenth of a solar mass) are needed to generate core temperatures high enough to sustain the normal fusion process of hydrogen → helium burning that is a hallmark of a true star. Astronomers know of no brown dwarfs in or near our Solar System, nor anywhere in the extended solar neighborhood of thousands of cubic light-years, but we are beginning to find some in the Milky Way beyond.

Even our best telescopes have difficulty spotting such brown dwarfs in deep space. They are intrinsically very faint, glowing mostly with the heat left over from their formation, and even a smattering of interstellar dust further dims our view. Recent advances in detector technology, especially in the infrared part of the spectrum where warmth can be sensed against the cold background of space, have enabled astronomers to begin to catalog what is perhaps a whole new population of these objects. It's not inconceivable that hundreds of billions of brown dwarfs populate our Galaxy, comparable in numbers to all the genuine stars in the Milky Way. However, only a few dozen of these elusive objects have been found to date—many in binary systems whose bright stellar member often betrays the presence of a small, dark companion—yet that's enough to prove the reality of this intermediate stage of abortive stars.

Brown dwarfs' inherent darkness makes them potentially relevant to one of the great unsolved problems in science: Could they be part of the missing dark matter plaguing astronomy today? Until the past few years, cosmic inventories had failed to account for these small, dim objects, now thought to be scattered liberally about the Galaxy. Clearly, brown dwarfs must contribute something to the dark matter and some astronomers are inclined to think that they might hide much of it. However, given that they are made of normal, baryonic matter, brown dwarfs cannot be the entire solution. Current censuses imply that their small masses accumulate to no more than a few percent of our Galaxy's dark matter.

A whole array of smallish, compact bodies—from dwarf stars to asteroidal rocks and planet-sized objects, as well as myriad old and dead stars encountered in the next section—could all be roaming the Milky Way in prodigious numbers undetected thus far. Any object having a size midway between, on the one hand, stars large enough to illuminate themselves to become visible from afar and, on the other, atoms and molecules small enough to reveal themselves spectroscopically even at great distances, would be virtually undetectable by any observational means currently available. Ironically, galactic space could be chock full of "interstellar basketballs," yet we have no way of knowing about them.

∞

Hardly anything notable befalls a star during most of its lifetime. Provided the nuclear events at its core continue to offset the relentless onslaught of gravity, nothing spectacular happens to the star as a whole. Predictably, its core fuses hydrogen into helium, its surface erupts in flares and storms, and its atmosphere releases vast amounts of radiation. But, by and large, stars experience no sudden changes while in equilibrium. They simply "burn" hydrogen during this, the longest phase in the history of all stars, lasting about ninety-nine percent of their total lifetimes.

Actually, stars enjoy hydrostatic equilibrium, not thermodynamic equilibrium. The former pertains to the structural integrity of a normal star, noting again its delicate balance in the tug-of-war between gravity pulling in and heat pushing out. Technically, it's not heat that pushes out as much as gas pressure; heat is a form of energy while pressure—the product of temperature and density—is more akin to a force. Hydrostatic equilibrium—as in a "compressible fluid," which is the way stars are modeled—tends to stabilize a star at every point within the star, to keep it from collapsing or exploding, in either case catastrophically. By contrast, thermodynamic equilibrium occurs when temperatures are uniform throughout, a state most definitely not achieved by any star. In fact, stars have a clear and obvious temperature gradient, from their fiery cores to their cooler (but still hot) surfaces. What's more, such gradients grow as stars age, driving them further from thermal equilibrium.

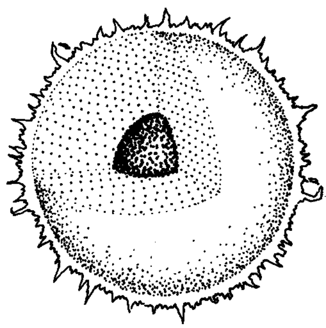
Note also, to make another clarification, that even in hydrostatic equilibrium, a star like the Sun continues to change its luminosity—that is, its rate of energy flow—ever so slightly over the course of its lifetime. Specifically, for the Sun's case, that amounts to an increase in brightness of about one percent every hundred million years. Although that seems minute, extrapolating back some three or four billion years means that the early Sun was probably only a third as luminous as it is today—and that might pose a problem understanding the origin and maintenance of life if planet Earth were at the time too cold for water to be liquefied. We shall return to discuss this "faint-Sun paradox" later in the Cultural Epoch.

Stars, then, in their normal, balanced state continue to produce energy indefinitely, pending some drastic change. The great struggle between heat and gravity remains stabilized, typically for billions of years. Eventually, however, something drastic does occur: all stars eventually exhaust their fuel.

Computer simulations are again our foremost guide to the specific changes experienced by any star near death. Identifying numerous physical and chemical factors and adjusting their values repeatedly, theoreticians have built models to describe the wide variety of stars seen in the real Universe. Let's first detail the finale of a star like our Sun, after which we can extrapolate to all stars, large and small. Keep in mind, though, that all these fatal events occur within the last one percent of a star's lifetime.

As the Sun ages, its hydrogen steadily depletes, at least in a small, central core about a hundredth of the star's full size. After nearly ten billion years of slow and steady burning, little hydrogen will remain within the innermost fusion zone. The star literally runs out of gas. Much like an automobile cruising along a highway at a constant speed for many hours without a care in the world, its engine starts to cough and sputter as the gas gauge approaches empty. Unlike automobiles, though, stars are not easy to refuel.

Widespread exhaustion of hydrogen in the stellar core causes the nuclear fires there to cease. Hydrogen combustion continues unabated in the intermediate layers, above the core though well below the surface. But the core itself normally provides the bulk of the support in a star, acting as a foundation and guaranteeing stability. By contrast, the lack of core burning assures instability because, although the outward gas pressure weakens in the cooling core, the inward pull of gravity most assuredly does not. Gravity never lets up; it's relentless. Once the outward push against gravity is relaxed—even a little—structural changes in the star become inevitable.



As the Sun ages, its hydrogen steadily depletes, at least in a small, central core . . .

Generation of more heat could bring the aged star back into hydrostatic balance. If, for example, the helium at the core began fusing into some heavier element such as carbon, then all would be well once again, for energy would be re-created as a by-product to help reestablish the outward gas pressure. But the helium there cannot burn—not yet, anyway. Despite the phenomenal temperature of millions of degrees, the core is just too "cold" for helium to fuse into any heavier elements.

Recall that a temperature of at least ten million degrees Celsius is needed to initiate the hydrogen → helium fusion cycle. That's what it takes for two colliding hydrogen nuclei (protons) to get up enough steam to overwhelm the repulsive electromagnetic force between two like charges. Otherwise, the nuclei cannot penetrate the realm of the nuclear binding force and the fusion process simply doesn't work. Even ten million degrees Celsius, however, is insufficient for helium fusion, since each helium nucleus (two protons and two neutrons) has a net charge twice that of the hydrogen nucleus, making the repulsive electromagnetic force greater. To ensure successful fusion by means of a violent collision between helium nuclei, even higher temperatures are needed. How high? About a hundred million degrees Celsius.

Lacking that degree of heat, the star's core of helium "ash" does not remain idle for long. Its hydrogen fuel spent, the core begins contracting. It has to; there's not enough pressure to hold back gravity. However, this very shrinkage allows the gas density to increase, thereby creating more heat as gas particle collisions become ever more frequent. Once again, it's gravity, in the guise of gravitational potential energy converting to frictional heat energy, that drives this process—indeed drives up the temperature.

The increasingly hot core continues to roil the overlying layers of this stellar furnace. It's very much like a domestic thermostat that calls for more heat in our homes, thereby keeping the air temperature comfortably stable. In an aged star, Nature seeks more energy to restabilize events, and when the star generates enough of it, negative feedback terminates the contraction—at least for a while. ("Feedback" because, as in a central heating system, a change in the effect is fed back to modify its cause, and "negative" because the feedback loop controlling the process ensures that the effect doesn't increase or decrease without limit.) But first, higher temperatures—at this stage, well over ten million degrees Celsius—cause hydrogen nuclei in the star's intermediate layers to fuse even more furiously than in the core before. All the while, helium ash continues to pile up around the core.

The aged star is really in a predicament now. Its days are numbered. The core is unbalanced and shrinking, on its way toward generating enough heat for helium fusion. The intermediate layers are also scrambling to maintain some semblance of poise, fusing hydrogen into helium at faster-than-normal rates. Alas, the gas pressure exerted by this enhanced hydrogen burning does build up, forcing the star's outermost layers to expand; not even gravity can stop them. So, although the core is shrinking, the overlying layers are expanding! Clearly, the star's structural stability is completely ruined.

Two observable aspects of such a perverse star are interesting. To an astronomer far away, this celestial object would seem gigantic, nearly a hundred times larger than usual. Captured radiation would also imply that the star's surface was a little cooler than normal. This is not to say that the act of ballooning and chilling of an aged star could be observed directly during any one human lifetime. The transition from a normal star to an elderly giant still takes about a hundred million years.

The second change—surface cooling—is a direct result of the first change—increased size. As the star expands, the sum total of its heat spreads throughout a much larger stellar volume. Hence, visible radiation emitted from such a cooling, yet still hot, surface shifts in color. Like a white-hot piece of metal that turns red while cooling, the whole extended star displays a reddish tint. Over the course of time, again long by human though short by stellar standards, a star of normal size and yellow color slowly changes into one of giant size and red color. The bright normal star has evolved into a dim red giant.

To recapitulate these momentous events and give them some local color, once the Sun exhausts its hydrogen fuel supply at its core, instability is sure to set in. Its core will shrink, its overlying layers swell, and its equilibrium becomes shot. As such, the Sun is destined to become a bloated sphere hundreds of times its normal size, perhaps large enough to engulf many of the planets, including Mercury and Venus, and maybe even Earth and Mars as well.

Humans need not panic, not yet at any rate. Provided the theory of stellar evolution is reasonably correct as described here, we can be sure that our Sun will not swell to this red-giant stage for another five billion years. Whether life can remain viable on Earth that long is debatable; there are two competing arguments: First, owing to the faint-Sun paradox noted a few pages earlier, the future Sun seems likely to increase its luminosity again by ten percent in a "mere" billion years, possibly rendering our planet unsuitable for life well before the Sun itself expires. Planet Earth will eventually get quite steamy regardless of any global pollution caused by humankind. Second, countering that long-term heating is a natural cooling forecast by the expected outward migration of the planets' orbits as the Sun loses mass and lessens its gravitational grip. Hard to believe, the Sun is shedding its own matter (in a "solar wind") at the prodigious rate of about a million tons *each second*, yet even in a billion years will have lost less than a tenth of one percent of its total mass, which might not be enough for the planets to drift away much. Whether the resulting cooling trend caused by the receding planets can offset the heating trend caused by increased sunlight is an unsolved problem—a rare astronomical problem with life-and-death terrestrial implications. Whichever, life's days on Earth are surely numbered, its oceans destined to evaporate, its atmosphere dissipate, our planet eventually resembling a ceramic-encrusted Mercury. Not to worry, such a hell on Earth will not commence for nearly another trillion or so days.

Red-giant stars are not the fiction of some theoretician's mind. They really do exist, scattered here and there about the sky. Even the naked eye can perceive the most famous of all red giants—the bright star Betelgeuse, that swollen, elderly, distinctly reddish member of the constellation Orion—a prominent signpost in the northern hemisphere's winter sky. This star is so luminous, it can be seen even through the smog and light pollution of our biggest cities. Look up!

Should the inherent imbalance of a red-giant star be maintained unabated, the core would eventually implode while the rest of the star drifted into space. Various forces and pressures at work inside such a decrepit star would literally, though slowly, pull it apart. Fortunately for the stellar veteran, this tortuous shrinkage and expansion is not expected to continue indefinitely. Within a hundred million years after the star first begins to panic for lack of hydrogen fuel, something else happens—helium begins to burn. This is when the natural thermostat shuts off the flow of additional heat as the core stabilizes once more. Though this seems like a whole new lease on life, it amounts to only a brief reprieve.

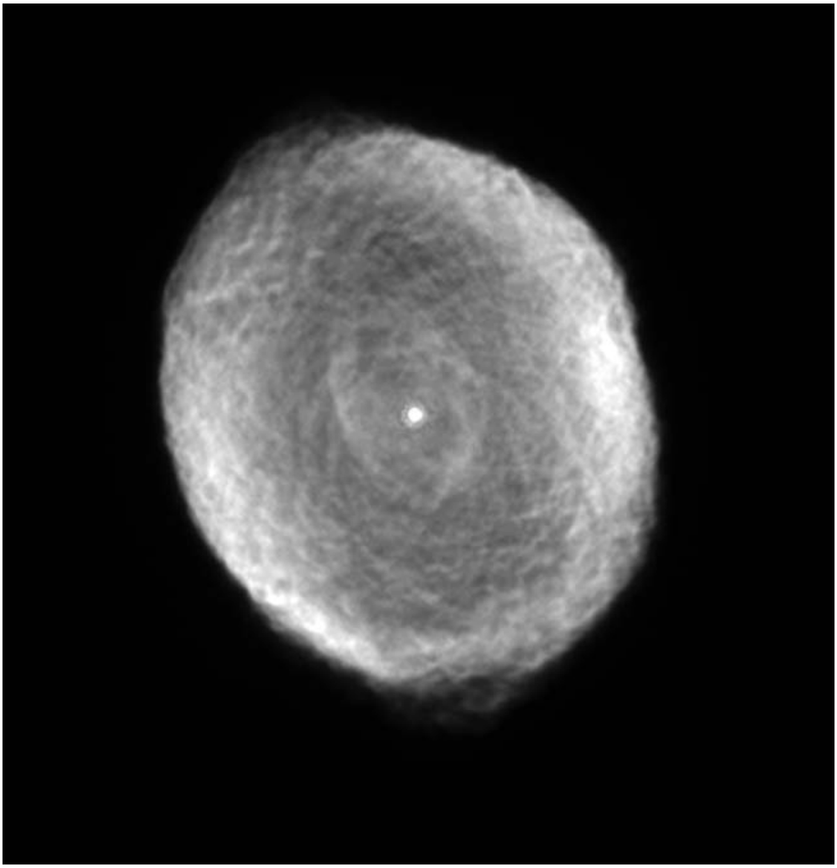
Deep down inside a red-giant star, the density increases as the interior pressure builds. Once the matter in the star's core becomes a thousand times denser than that of a normal star, collisions among the gas particles are violent and frequent enough to generate sufficient heat, via friction, to reach the hundred-million-degree temperature needed for helium fusion. Helium nuclei henceforth collide, ignite the central fires once again, and begin transforming into carbon. Thereafter for a period of a few hours, the helium burns ferociously, like an uncontrolled bomb. It's remarkable that the star doesn't explode.

Despite their brevity, these renewed nuclear events release an enormous flood of new energy. The energy is potent enough to etherealize the core matter somewhat, thereby lowering its density and relieving some of the pent-up pressure among the charged nuclei. This small expansive adjustment of the core halts the gravitational contraction of the star, reestablishing an equilibrium of sorts—in this case, a balance occurring at the quantum level among the densely packed electrons whose tiny, pointlike spheres are essentially touching one another, thereby physically holding up the aged star against gravity.

To make yet another clarifying technical comment, note that, in actuality, the nuclear reaction that changes helium into carbon occurs in two steps known as the "triple-alpha process." First, two helium nuclei (which are also known as alpha particles) combine to form beryllium, which is a very unstable nucleus that would normally break right back down (in less than a microsecond) into two helium nuclei—causing the process to be stuck in an endless cycle. However (and this is the second step), the huge densities in the helium ash guarantee that a third helium nucleus sometimes collides with newly made beryllium before it has a chance to decay. This is not a miracle, or some sort of "anthropic" event, implying that some supernatural being designed it that way to permit heavy elements and therefore life. Rather, given the very high densities in a red giant's core, the timescale for collision among three helium nuclei is naturally shorter than for the breakdown of beryllium. The result is carbon, the nucleus of a vitally important element later in the Chemical Epoch of our cosmic-evolutionary story.

Once the helium → carbon fusion reactions commence, thus stabilizing the core, the hydrogen → helium fusion reactions churning in the layers above subside. In that way, the star expands its outer layers a bit too rapidly, overshooting the distance at which it reestablishes a relaxed structural balance. The entire star is then able to shrink a little, losing somewhat its swollen appearance. Like all the other evolutionary changes in the early or late phases of a star, this slight size adjustment is made quickly—at least by cosmic standards—that is, in about a hundred thousand years.

Though the timescales for marked stellar change are deemed rapid for a star's emergence from dust as well as its thrust toward doom, all these notable durations are still long compared to human lifespans. Observers have little hope of watching a given star move through all, or even some, of the evolutionary paces of the Stellar Epoch. Instead, much as before, astronomers search the Galaxy for evidence of diverse cosmic objects at different stages of their evolutionary cycles, trying to position them like puzzle pieces into a self-consistent picture. Or, to use another metaphor, like social behaviorists charged with the task of unraveling the population dynamics of animals, astronomers are finding that the deeper they peer into galactic lairs, the more instructive the menagerie



Evidence of stellar evolution.

The sky is replete with hundreds of bloated old stars, especially in the halo of the Milky Way. This one, called the Spirograph Nebula, has expanded into its red-giant phase only during the past few thousand years, revealing a bright core, or white-dwarf remnant, at the center. This so-called planetary nebula, now more than ten times the size of our Solar System, was once a middle-aged object like our Sun. It's a good example of the likely fate of ole Sol some five billion years from now. *Source: Space Telescope Science Institute.*

of stellar inhabitants becomes. In the end, we rely on mathematical calculations to match the theoretical models with the observational evidence of the many varied stages in the birth and death of stars.

This reliance on computer modeling is exactly what made the results of an important experiment so disturbing—until recently. The one experiment that bears directly on the physical events inside stars did not jibe well with the predictions for a star like our Sun. For decades, scientists were puzzled by the number of neutrino elementary particles found in the solar radiation reaching Earth. Derived from an Italian word meaning "little neutral one," neutrinos are known from experiments on Earth to be virtually massless and chargeless and to travel at (or very close to) the velocity of light. Interacting with almost nothing, neutrinos are ghostlike particles endowed with an ability to pass freely through several light-years of lead! Hence, they should be able to escape unhesitatingly from the solar core, where they are created in copious amounts as by-products of nuclear reactions. Ordinary radiation scatters around (or "random walks") in the solar interior for about a million years before being emitted from the Sun's surface into space, but neutrinos should pierce the solar surface in two seconds and arrive at Earth a mere eight minutes after being made at the core. They thus embody the only direct test of the nuclear events responsible for powering the Sun.

Solar neutrinos nonchalantly penetrate Earth all the time. Some five million neutrinos pepper every square centimeter of our bodies each second, though we are neither aware of nor harmed by them. Despite their elusiveness, however, the effects of neutrinos can be studied with carefully built instruments made of rare materials. One of those materials is a chemical with the tongue-twisting name of tetrachloroethylene. As toxic as it sounds, C_2Cl_4 is a safe fluid often used in the dry-cleaning industry. Such a "neutrino telescope" was originally built at the bottom of a South Dakota gold mine by filling a large tank with four hundred thousand liters (about a hundred thousand gallons) of this stuff. In that way, some of the solar neutrinos arriving at Earth can be counted and analyzed, though actually only one is detected for every million billion of them streaming through the tank. Depth is essential to shield the experiment from interference due to cosmic rays and other elementary particles hailing from nonsolar sources such as ancient supernovae. Although the equipment seems to have worked properly for decades, the rate of neutrino detection has been consistently less than theory predicts; they are seen about twice per week, rather than once per day—about a threefold discrepancy.

Astrophysicists have wrestled with these puzzling results for many years. Both theorists and experimentalists are reluctant to blame any underabundance of solar neutrinos on conceptual errors in the theory of stellar evolution. No one wants to discard what seems like a good understanding of solar fusion, all other aspects of which agree so well with observations. Some researchers (mostly theorists) suspect the experimental gear: perhaps it was not quite tuned properly, and in any case a factor of three or so is not a large issue in astronomy. Others (mostly experimenters) are leery about the computer models: if the Sun's core were only ten percent cooler than theory maintains, the predicted number of solar neutrinos would be less. Still others argue that we don't yet know enough about the odd neutrino particle itself; the physical properties of the neutrinos might make them the culprit, especially if they turn out to have even minute amounts of mass.

More recently, in the year 2000, this factor-of-three discrepancy seems to have been resolved during experiments in new underground laboratories located in Japan and Canada, the latter using a thousand-ton sphere of ultrapure water suspended more than a kilometer beneath the surface and surrounded by ten thousand sensors. The new results do indicate that neutrinos have minute amounts of mass—roughly a millionth the mass of an electron, which is itself nearly two thousand times lighter than a proton. However, even this ultratiny mass is enough to cause the apparently schizophrenic neutrinos to change their properties, even to transform them into other particles, during their eight-minute journey from the Sun to Earth. And that is what most astronomers now think is happening: neutrinos are produced in the Sun at the rate predicted by theory, but some change into something else—actually, they morph into other types of neutrinos—en route to Earth. The original experiments were insensitive to the changes, but the newer experiments are seeing evidence of them. At issue now is the need to fix up the standard model of particle physics, in which neutrinos are expected to have precisely zero mass—or to begin a whole new search to solve a new contradiction between quantum theory and delicate experiment.

Assuming these latest results are correct—namely, that neutrinos have both intrinsic mass and mutable properties—we once again wonder if the neutrinos could be the solution to the elusive dark-matter quandary. Given the tremendous number of neutrinos likely flooding our Galaxy—both leftovers from particle interactions in the early Universe as well as new ones created in all the stars of the Milky Way—it still seems doubtful. Although neutrinos are surely part of the cosmic mix, their total accumulation likely amounts to less than one percent of the overall mass of the Galaxy.

In any event, few researchers regard the surprising solution to the solar-neutrino problem as a threat to our understanding of the way stars shine. This decades-old dilemma now seems to have been more of a problem with the physics of the particle than with the astronomy of the Sun. By checking and double-checking both theory and experiment, all the while continuing to address the issue with reason and skepticism—which is exactly the way science progresses—what once loomed as a serious misunderstanding of stellar fusion has apparently now been resolved.

Uncertainties limit our understanding in every epoch of cosmic evolution. Here in the Stellar Epoch, as elsewhere, we seem able to identify the broad outlines of many possible events, but the fine details are not always in hand. What causes flaring on our Sun, resulting in huge prominences of matter and radiation that escape our star and impact our planet? How does the eleven-year solar cycle work, turning surface sunspots off and on at near-decade intervals? Can we explain satisfactorily the million-degree-Celsius corona, or outer atmosphere of the Sun, when its surface is only six thousand degrees Celsius? What is the role of magnetic fields in the origin, maintenance, and demise of all stars?

Even the brightest star in the nighttime sky seems a little puzzling, at least as regards the historical record. Sirius A, only nine light-years away, appears twice as luminous as any other visible star (excluding the Sun) and has been prominently observed by many ancient civilizations. Cuneiform texts of the Babylonians refer to this star as far back as 1000 B.C., and historians know that the star strongly influenced the agriculture and religion of the Egyptians of about 3000 B.C. So, given the lengthy record of observations of Sirius, here is an object for which we might have a chance to study slight evolutionary changes, despite the long time scales usually needed to produce such changes. Yet herein lies the puzzle.

Sirius A does seem to have changed its appearance over the ages; the historical records clearly imply it. But the naked-eye observations of the ancients are confusing. Every piece of information about Sirius recorded between the years 100 B.C. and A.D. 200 claims that this star was red. In contrast, modern observations now show it to be white or bluish white, but definitely not red. According to the theory of stellar evolution, no star should be able to change its color from red to blue-white so dramatically in such a short time—even over thousands of years. Any change of this sort should take roughly a hundred thousand years, perhaps a lot longer, and in any case would more likely change from blue to red.

Astronomers have offered many explanations for the rather sudden change in Sirius A. These include the possibility that some ancient observers were wrong and other scribes copied them. Or perhaps a galactic dust cloud passed between Sirius and Earth some two thousand years ago, reddening the star much as Earth's dusty atmosphere often does for our Sun at dusk. Or maybe a companion to Sirius A, namely Sirius B, was a red giant and dominant star of this double-star system two thousand years ago and has since expelled its outer envelope to reveal the small (white-dwarf) star that we now observe as Sirius B.

None of these explanations seems plausible, however. How could the color of the sky's brightest star be incorrectly recorded for hundreds of years? Where is the intervening galactic cloud now? Where is the shell of the former red giant? We are left with the uneasy feeling that the night's brightest star doesn't seem to fit well into the currently accepted scenario of stellar evolution.

As if that were not troubling enough, our resolute navigational beacon, Polaris, the North Star, is also a bit of a conundrum. Despite Shakespeare's closing line for Julius Ceasar, "But I am constant as the Northern Star," the light from Polaris is not so steady, yet the sky is replete with variable stars so that is alright. Alas, the extent of its variability is also changing, and quickly too, and that's what's puzzling. Greek astronomers of two thousand years ago claimed that Polaris's average brightness was three times dimmer than now, a rate of change, if real, much greater than that allowed by current models of stellar evolution. That not all the loose ends are yet tied up is not meant to imply major cracks in our understanding of stars; rather, plenty of work remains regarding those picky little details that often serve to fine-tune that understanding.

These subtle yet bothersome issues aside, our understanding of stellar evolution is judged one of the great success stories of modern astrophysics. Theory and observation have advanced hand in hand over the last many decades, refining our knowledge of stars as they proceed from cradle to grave. Today, the subject of stellar evolution is a cornerstone of the cosmic-evolutionary narrative, a key part of that broadest view of the biggest picture that we've come to know rather well.

Nuclear reactions in an old star's helium core churn on, but not for long. Whatever helium exists in the core is rapidly consumed. The helium $\rightarrow$ carbon fusion cycle, like the hydrogen $\rightarrow$ helium cycle before it, runs at a rate proportional to the temperature; the greater the core heat, the faster the reactions progress. Under these very high temperatures, helium fuel simply won't last long—probably less than a few million years.

Buildup of carbon ash in the inner core causes a series of physical events similar to those in the earlier helium core. Helium first becomes depleted at the star's very center, after which fusion there ceases, the temperature being too low for carbon detonation. The carbon core then shrinks and heats a little, as Nature's thermostat kicks in again while searching for more energy from renewed gravitational infall. This, in turn, causes the hydrogen and helium burning cycles to ramp up in the intermediate and outermost layers of the star. Such an aged star begins to resemble a huge onion, with different shells of progressively heavier elements toward its center. All this additional heating causes its outer envelope ultimately to expand, much as it did earlier, making the star once again a swollen red giant.

Provided the core temperature does become high enough for the fusion of two carbon nuclei, or more likely a union of carbon and helium nuclei, still heavier products can be synthesized. Newly generated energy supports the star at each stage in the nuclear chain, returning the star to its accustomed hydrostatic equilibrium. Again, this is not a thermodynamic equilibrium, for such decrepit old stars develop strong thermal and elemental gradients from core to surface. For this reason, such aged stars are decidedly more complex than their younger counterparts. Ironically, as the fusion process advances, old stars continue getting brighter; all the while, they are dying.

This contracting-heating-fusing-cycle is generally the way that many of the heavy elements are fashioned within the last gasps of stellar cores. All elements heavier than carbon are created within the final one percent of some stars' lifetimes. Our Sun, however, is not one of them.

∞

How do stars die? What do they become at the end of their long and brilliant existence? The answers rely partly on computer modeling and partly on what is seen in the sky—the usual marriage of theory and observation. The details of this problem, quite frankly, are once again tricky, in that no one has witnessed a nearby star die since the invention of the telescope nearly four centuries ago. Guided by theoretical predictions of how stars ought to behave near (and after) death, astronomers search the Universe, seeking evidence of objects resembling the predicted hulks.

The best models hold that the final stages of stellar evolution depend critically on the mass of the star. As a rule of thumb, low-mass stars die gently, whereas high-mass stars die violently. The dividing line between these two very different outcomes lies around eight times the mass of the Sun. Since a Milky Way census confirms that hardly one percent of all stars have more than this mass, our Sun and the great majority of stars are members of the low-mass category. Only rare stars much larger than our Sun are grouped in the high-mass category.

The demise of our Sun is destined to be straightforward and unspectacular. The Sun's core will become extremely hot and compact as it heads toward its end state. A single cubic centimeter of stellar core matter would weigh a ton on Earth. That's a thousand kilograms of matter compressed into a volume the size of a pea. Yet, even at these very high densities, collisions among nuclei are insufficiently frequent and violent to raise the temperature to the extraordinarily high six hundred million degrees needed to ignite a new round of nuclear reactions and thereby change carbon into any of the heavier elements. There is simply not enough matter in the overlying layers of the smaller stars to bear down any harder. The density reaches maximum compression, the temperature stops rising, and oxygen, iron, gold, uranium, and many other elements cannot be created in low-mass stars.

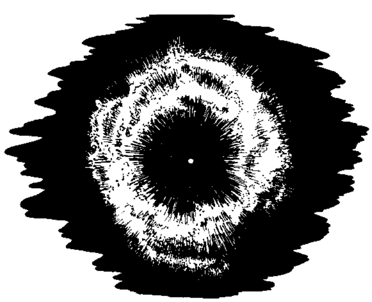
Small stars like our Sun manage to work themselves into quite a predicament in their old age. Their carbon core is, for all intents and purposes, dead. Helium just outside the region of carbon ash continues to transform into more carbon, while hydrogen in the intermediate layers above converts into more helium. This onslaught of heating slowly pushes away the outermost layers to even greater distances. The expected upshot is an object of distinctly odd posture having two separate parts. Called a planetary nebula, it's predicted to have a halo of warm, rarefied matter veiling a hot, dense core.

Nebula is Latin for "mist" or a "cloud" of great extension and extreme tenuity, but these planetary nebulae should not be confused with the even larger galactic nebulae noted earlier in this epoch. Galactic nebulae are signposts of recent stellar birth; planetary nebulae are indicators of impending stellar death. The adjective "planetary" is also misleading, for these celestial objects are not related to planets in any way. Their designation dates back to the eighteenth century, when astronomers could barely distinguish among the myriad faint, fuzzy patches of light in the nighttime sky, and some observers mistook them for planets. Later studies clearly demonstrated that the nebula's fuzziness results from shells and rings of warm gas surrounding a small luminous object. Modern telescopes well resolve planetary nebulae, enabling us to recognize their true nature.

Odd or not, nearly a thousand examples of planetary nebulae have been discovered in our Galaxy alone. Though they sometimes appear to have a ring surrounding a bright core, their halo-shaped appearance is often an illusion owing to accumulations of their emitted gas along our line of sight. Direct observations confirm the theoretical predictions that their expelled matter consists of a mostly spherical envelope in the act of gently escaping from the core of an aged red-giant star. Exceptions abound, and weird patterns are common, as the star's rotation and mottled environment often distort the receding gaseous shell, making some planetary nebular shapes, well . . . nebulous. A few planetaries even have illuminated jets, streams, and spirals of gas emanating from near their aged cores, sporting peculiar geometries thus far unexplained.

The evolution of a nebula's expanding envelope is not very interesting thereafter. It simply continues spreading out as time passes, becoming evermore diffuse and cool, and gradually merging imperceptibly with the interstellar medium. Its most important role is to enrich space with additional helium atoms and possibly some carbon atoms as well.

Continued evolution of the core remnant at the center of a planetary nebula is not much more exciting. Formerly concealed by the atmospheres of red-giant stars, cores appear once their flimsy envelopes re-



Formerly concealed by the atmospheres of red-giant stars, white dwarfs appear once their flimsy envelopes recede.

cede. These cores are relatively small, glowing objects, highly abundant with very hot carbon but not nuclear burning. They shine only by stored energy, though their small size and intense heat guarantee a white-hot appearance. Not much bigger than planet Earth, these shrunken carbon cores—balls of nuclear wastes, really—are called white-dwarf stars. They, too, are seen in the heavens.

Analysis of radiation emitted by white-dwarf stars shows their properties to agree well with the computer models of elderly, low-mass stars. Scores of dwarfs are found at the very centers of planetary nebulae, but only one per nebula. Several hundred additional ones have been discovered "naked" in our Galaxy, their envelopes having been expelled to invisibility long ago. The most famous of all white dwarfs is Sirius B, the dim binary companion to the bright Sirius A, about which the aforementioned nagging problems remain.

So astronomers readily identify red-giant stars, planetary nebulae, and white-dwarf stars in the nearby cosmos. At different stages in their old age, each of these objects seems to match the overall disposition predicted by the theoretical calculations for ancient low-mass stars. Once again, though, we should not expect to witness the act of envelope expulsion during the course of a single human lifetime. Several tens of thousands of years are typical for a red giant's atmosphere to recede sufficiently for a white dwarf to become visible.

Nothing exciting befalls the dwarf stars thereafter. For all practical purposes, these "stars" are dead. They continue to cool, becoming dimmer with time while slowly transforming from white dwarfs to yellow dwarfs and then red dwarfs, eventually crystallizing into trillion-trillion-carat diamonds in the sky. Their final state is that of a black dwarf—a cold, dense, burned-out ember in space. Such stellar corpses have reached the graveyard of stars.

No one knows how many black dwarfs populate our Galaxy—which is not surprising since they're unlit. They're also in that hard-to-probe size range between normal bright stars, on the one hand, and atoms and molecules, on the other. Even if these dark clinkers could somehow be detected, we would probably find few of them. The total duration of a low-mass star is very long, typically comparable to or longer than the age of the Galaxy. So don't look for them to solve the dark-matter problem. Our Milky Way has not likely yet endured long enough for many low-mass stars to have completed the whole stellar-evolutionary trek from birth to death. Perhaps none has.

Different fates await stars having more than eight times the mass of our Sun. By and large, they evolve much like their low-mass counterparts up through the red-giant stage, with only one difference. All the evolutionary paces occur more quickly for the high-mass stars because their greater mass enables them to generate more heat. As before, it's gravity that creates that heat and the subsequent energy flows speed all evolutionary events. That's why the biggest stars, consuming fuel at prodigious rates, endure for shorter periods, in fact well less than a billion years. Some of the biggest ones, containing tens of solar masses, last for as little as ten million years, or a thousand times less than the Sun—a mere cosmological wink of an eye, yet still longer than a hundred thousand human lifetimes.

At the red-giant stage, the core of a high-mass star is able to attain the six hundred million degrees Celsius to begin fusing carbon into even heavier elements. Again, mass is the key. Truly massive stars generate stronger gravitational forces than solar-type stars, and the added gravity can crush matter in the core to a high enough density to ensure frequent and violent collisions among the gas particles.

Theoretical models call for a highly evolved star of large mass to have several internal layers where various nuclei burn simultaneously. The insides of this star even more strongly resemble an onion. At the relatively cool periphery just below the surface, hydrogen fuses into helium. In the middle layers, helium and carbon fuse into heavier nuclei. Just above the core, magnesium, silicon, sulfur, and many other heavy nuclei are present, and some of these in turn fuse into even heavier nuclei. As the temperature rises with depth, the ash of each burning stage becomes the fuel for the next stage. The core itself is full of iron nuclei, rather complex pieces of matter each containing several dozen protons and neutrons, midway between the lightest and heaviest of all known nuclei.

Each of these fusion cycles, during which nuclei for new elements are created at various depths in a star's interior, is induced by periods of stellar instability. The core cools somewhat, contracts a little, heats some more, fuses heavy nuclei, and depletes its fuel, after which the cycle starts over by contracting again, heating again, fusing again, and so on. At each stellar-burning stage, energy is released as a by-product of the fusion process, effectively supporting the star (at least for a while) against gravity. And at each stage, as the star's interior evolves, its burning rate accelerates. For example, for a star some twenty times more massive than the Sun, hydrogen burns for ten million years, helium for a half-million years, carbon for a thousand years, oxygen for one year, and silicon for a week.

With iron accumulating in the core, complications develop for this sick and dying star in less than a day. Nuclear physicists say that iron has the "highest nuclear binding energy," meaning that iron is the most stable of all the elements. Consequently, nuclear events involving iron don't produce energy; iron nuclei are so compact that they tend to consume energy. That's because further contraction actually breaks down the iron back into helium, absorbing energy instead of emitting it. In lay terms, iron nuclei play the role of fire extinguisher, suddenly damping the stellar inferno, at least at the core. With the buildup of iron at the core, the central fires quench and the nuclear events cease for the last time.

Potential for disaster now clearly exists. No longer is this very massive star upheld by nuclear fusion at its core. The star's foundation is gone, its structural stability destroyed. Although the temperature in the iron core has by this point reached several billion degrees Celsius, the strong and now unopposed gravitational pull of the great mass of overlying matter ensures catastrophe in the very near future. Unless nuclear events continue unabated, trouble is a certainty for any such defunct star.

Once gravity overwhelms the pressure of the hot gas, the star implodes, falling in on itself. The implosion doesn't take long, perhaps only minutes after cessation of core kindling; this is not a gentle contraction as much as a dramatic collapse. Internal temperatures and densities then rise phenomenally, causing the star to rebound instantaneously like a coiled spring, detonating parts of the core while jettisoning all the surrounding layers. The details of how such a massive star physically rebounds like this are not well known, but the outcome surely is: much of its mass—including a variety of heavy elements cooked within—is expelled into neighboring regions of space at speeds initially reaching tens of thousands of kilometers per second (or nearly a hundred million miles per hour). The expulsion is much, much more violent than that for a planetary nebula; this is a titanic event, since much of an entire star has literally exploded. All stars much larger than our Sun are slated to perish in this way. Such a spectacular death rattle is known as a supernova.

Supernovae are the most tumultuous events in any galaxy, indeed among the most energetic in all the Universe. The exploded stellar debris is intensely hot and altogether can radiate a flash equal to more than a billion times the brightness of our Sun. This amounts to a single star suddenly rivaling the brightness of nearly our entire Milky Way Galaxy within a few hours after its outburst. Eventually, as in any explosion, the surge subsides and the debris cools, but not before the galactic neighborhood has been irradiated with plenty of potent energy and heavy elements.

Astrophysicists are unsure of precisely when and how supernovae explode because a nearby star hasn't erupted in this way since early in the seventeenth century. Nor are the theoretical models entirely clear on the intricate details of the explosion. Glibly stated: How does most of an *entire star* not just implode catastrophically, but completely reverse that sudden implosion by exploding dramatically?

Supernova models imply that while heavy elements such as carbon, nitrogen, oxygen, sodium, magnesium, silicon, and much of everything up to iron are produced in stellar interiors, the explosion itself is responsible for elements heavier than iron. At the moment of detonation and for about fifteen minutes thereafter, intermediate-weight nuclei are fiercely jammed together, thus creating some of the heaviest of all nuclei; neutron capture creates the rest. Many of the rare elements are synthesized at this time, including silver, gold, uranium, and plutonium. Matter most valued by society on Earth therefore originated in the very last gasps of shattered stars once big and bright, though now dead and disintegrated. Ironically, the heaviest of all elements are made only after their parent stars have perished. But because the time available for making the heavies is so brief, elements heavier than iron are billions of times less abundant than most light nuclei—which is precisely what makes many of them so valuable.

The sprinkled debris of erstwhile stars then mingles with fresh interstellar hydrogen and helium made during the earliest Particle Epoch of the Universe. This messy mixture of all the elements can then undergo contraction, heating, and fusion yet again, thus fabricating second, third, and nth-generation stars in a seemingly endless cycle of birth, death, and rebirth. Our own Sun is at least a second-generation star, for it already contains heavy elements, including lots of iron. Since these heavies could not have been made in a low-mass, relatively cool star like the Sun, they must be the products of formerly massive stars that exploded long ago.

How do we know that stars really do create heavy elements in this way? Can we be sure that the theory of stellar nucleosynthesis is correct? One piece of circumstantial evidence is in hand, and one item of direct support is telling, in addition to the obvious wreck of the exploded debris itself as seen in the sky. First, the rate at which various nuclei are captured and the rate at which they decay are known from laboratory experiments of the past few decades; some of this work was done to support America's nuclear weapons program. When all these rates are incorporated into sophisticated computer models, which also take account of the temperatures, densities, and compositions at many layers within a massive star, the relative amounts of each type of synthesized nucleus match remarkably well the known abundances of intermediate-weight elements up to and including iron. Thus, despite the fact that no one has ever directly observed atomic nuclei in the act of production—other than the trace debris collected after nuclear bomb tests on Earth—the agreement between theory and observation is striking. We can thus be reasonably sure that Nature's way of making elements is well understood, given our knowledge of nuclear physics and stellar evolution.

Second, close study of one type of nucleus—a rare and unstable one named technetium—provides direct evidence that heavy-element formation really does occur in massive stars. This nucleus, much heavier than iron, is known from laboratory measurements to have a radioactive half-life of about two hundred thousand years. This is a very short time astronomically speaking, hence the reason why no one has ever found even traces of naturally occurring technetium on Earth; all of it decayed long ago. (It can, however, be studied as a newly created byproduct of nuclear laboratory experiments.) By contrast, the identification of technetium in the spectrum of many red-giant stars implies that it must have been created within the past million years or so. Elemental production is indeed underway in stars today.

And finally, the blast signatures of a detonated supernova and that of a nuclear bomb are identical. The flash of a supernova displays a rapid rise in luminosity at the moment of explosion, followed by a notably slower decrease in brightness over the next few years. The peculiar dimming of the light owes mostly to the radioactive decline of unstable nuclei produced in the fireball itself. Studies of the flash and decay of light in the aftermath of thermonuclear weapons tests on Earth imply that the process is one and the same—though, thankfully, far less intense. As destructive as bombs are, they help us understand Nature. Destructive as supernovae are, they help enrich Nature.

Supernovae are not just idle predictions of theoreticians. Plenty of evidence has been amassed that cosmic explosions have occurred throughout the ages. One of the most heavily studied remnants of a supernova is the Crab Nebula, so aptly named largely because its appearance resembles that type of marine animal. About six thousand light-years from Earth and in the constellation Taurus, its glowing debris is strewn over a ten-light-year extent. Now greatly dimmed, the Crab, as it's known for short, can be seen only through a telescope. But the measured motions of its expelled matter—still now racing outward at some thousand kilometers per second, or roughly two million miles per hour—imply a brilliant explosion whose light must have been easily seen with the naked eye more than nine hundred years ago.

Although *nova* is Latin for "new," modern astronomers realize that the sudden brightening of supernovae, briefly changing from near invisibility to great prominence, only made them seem new to the ancients. The original explosion of the precursor to the Crab Nebula was so spectacular that old Asian and Arab manuscripts claim its brightness surpassed that of Venus and even rivaled that of the Moon in the year A.D. 1054. Native Americans also saw it, having left engravings of the event in the rocks of what is now the Midwestern United States.

Numerous other massive stars no doubt self-destructed in earlier times. Written documents refer to relics of at least a dozen supernovae in our Galaxy within the past several thousand years. The nighttime sky harbors additional evidence of many wispy remnants of former stars that must have blown up well before the advent of recorded history. The closest of these was probably the Veil supernova, whose remains lay a mere fifteen hundred light-years from Earth. Astronomers can still see that remnant well enough to measure its expansion velocity and thus to infer that its progenitor star must have detonated around 18,000 B.C. Based on the amount of matter strewn throughout its debris field, this supernova likely exceeded the brightness of the full Moon. We can only speculate about the impact that such a suddenly luminous orb might have had on the myths, religions, and cultures of Stone Age humans.

Nowadays, hundreds of supernovae have been sighted in other galaxies—many of them momentarily as bright as their parent galaxies. Astronomers patrolling the skies on any given night often notice a sud-

Evidence of stellar death.

The Crab Nebula is the remnant of a star that nearly blew itself to smithereens. This is the way we see this supernova today—the glowing debris of a spectacular explosion that was actually witnessed in A.D. 1054 by Chinese, Arab, and Native American observers. The nebula now covers an area only one-fifth the diameter of the full Moon, but at its distance of six thousand light-years that debris is scattered over several light-years and is rich in heavy elements. *Source: European Southern Observatory.*

den brightening of a portion of some faraway galaxy, enabling them not only to verify that high-mass stars are common to all galaxies, but also to refine the predictions of the stellar-evolutionary models. If a month-long videotape could be made of a distant galaxy cluster, we would see supernovae popping off like minute and silent flashbulbs in a darkened stadium. Disconcertingly, however, humankind has been unable to inspect a supernova closely since a massive star in our own Milky Way hasn't viewably detonated in this way during all of the previous four centuries when telescopes have been in use.

The most recent supernovae observed in our Galaxy caused a sensation during the Renaissance and helped overthrow the leading philosophy of the time. Oddly, the earlier Crab Nebula had apparently gone unnoticed, or at least unrecorded, by Europeans several centuries before. Perhaps the influence of the Church was so strong and its dogma of immutability so rigid that faithful (or fearful) citizens simply put it out of their minds. But the sudden appearance and subsequent fading of very bright stellar objects in the years A.D. 1572 and 1604 were unavoidably noticed—the latter visible during the day for weeks—and together they shattered the Aristotelian idea of an unchanging Universe beyond Earth. Little did anyone then realize that these brilliant flashes in the heavens—now known as Tycho's and Kepler's stars, respectively—provided the mental seedlings for the eventual emergence of the scenario of cosmic evolution, wherein the concept of change is central and ubiquitous in the Universe.

Much more recently, astronomers were treated to a spectacular supernova in the Large Magellanic Cloud, that small galaxy less than two hundred thousand light-years away yet orbiting our own Milky Way. Amateur astronomers in Chile first spotted in it the winter of 1987 and within a few hours most of the world's telescopes, on the ground and in orbit, had focused on the suspect object. SN1987A, as it's called, was one of the most dramatic cosmic changes observed in the past four hundred years. Apparently, a giant star at least fifteen times the mass of our Sun had become unstable late in life, exploding with such vengeance that its glowing debris can still be clearly seen, despite its distance, more than a decade later. By and large, the observed properties of this celestial explosion agree well with our computer models of stellar evolution, including the detection of a short burst of neutrinos that reached Earth nearly coincident with the initial flash of light. As noted below, these neutrinos—the same kind of poorly understood particles implicated in the solar mystery discussed earlier—might well have triggered the rebound of the infalling star, causing the stupendous explosion whose remnants astronomers are still studying today.

Not everything about supernovae is well understood. Certainly their infrequence in our part of the Milky Way is a bit disturbing. Knowing the rate at which stellar evolution is thought to occur and estimating the number of massive stars in the Galaxy, we expect a supernova to occur in an observable location (away from the dusty parts of the Milky Way's plane) every century or so. Yet only six such galactic supernovae have been recorded in the past thousand years, and none at all in the past few hundred. Since it's hardly likely that any such garish explosions could have been missed since the last one several centuries ago, the Milky Way seems long overdue for a blast from the past. Unless massive stars explode much less frequently than predicted by theory, we should be treated to one of Nature's most spectacular exhibitions any day now. Let's just hope it's not too spectacular—or too close!

Supernovae may be more than splendid light shows. Should a massive star detonate in the galactic suburbs where our Sun resides, it could well inundate Earth for months with radiation and matter harmful to life. An initial pulse of high-energy X and gamma rays, followed quickly by neutrinos as well, would impact our planet suddenly, certainly destroying Earth's ozone layer, probably making our atmosphere radioactive, and possibly allowing sunlight to fatally fry most life on the surface. The topmost layers of massive stars are also physically ripped off and sent flying into space as extremely fast-moving elementary particles, known as cosmic rays, which would arrive somewhat later and for an extended period (perhaps decades) of bombardment. Such violent events might have triggered episodes of mass extinction as revealed by Earth's fossil record over the past few hundred million years—yet another fertile area of research in which astronomy and biology meet in the interdisicipline of astrobiology. Knowledge of the nearby stars—especially their masses—is thus of more than just passing interest. An ability to predict the manner in which nearby stars die is downright critical. Of particular concern is the possibility that one of our stellar neighbors might explode as a supernova, although we probably couldn't do much about it even if one did.

Statistics of the stars in our galactic neighborhood imply that one supernova can be expected within three hundred light-years of the Sun once every half-million years, or within thirty light-years once every half-billion years. Too close for comfort? Fortunately, none of the stars currently close to Earth is massive enough to die by self-detonation. Luckily for us, they all seem destined to perish, as will our Sun, via the more placid red-giant → white-dwarf route.

Fascinatingly, a viewable massive star has almost surely *already* exploded, but the light from this stupendous event has yet to reach us. Owing to the finite speed of light, some of the brightest stars now seen above could have actually blown up centuries ago and we wouldn't yet know it. The ten-solar-mass star Rigel, for instance and for all we know, looking proud and mighty some eight hundred light-years away, could have detonated during Earth's Middle Ages, and that "message" would still be winging its way toward us. Its "companion" in the Orion constellation, the fifteen-solar-mass red-giant Betelgeuse, only four hundred light-years distant, might have already done so more recently. These are certainly candidates to explode someday, as are the north star Polaris, the red supergiant Antares, and mighty Deneb, the brightest star in the constellation Cygnus.

Should such a supernova abruptly appear in the sky, we can be certain every major piece of astronomical instrumentation will immediately focus in the direction of this, the grandest of fireworks. Some major observatories, such as the Harvard-Smithsonian Center for Astrophysics, have established "supernova alert teams," where several astronomers stand ready to commandeer, within an hour's notice, all ground-based telescopes and orbiting spacecraft operated by that institution. A memorable false alarm on a Labor Day weekend twenty years ago even helped to smooth out some communications hazards, should a supernova inconsiderately time its earthly arrival on a human holiday! The team's prime objective is the study of the early phases of a supernova outburst, especially the various types of emitted radiation stretching from relatively harmless radio waves to potentially lethal gamma rays.

∞

Supernovae are extraordinarily energetic, easily and efficiently piling up nearby matter into dense clumps, much as a plow effortlessly pushes around snow. A single supernova can sweep up far more mass than it formerly contained as a star. For example, the explosion of a ten-solar-mass star would launch a shock wave extending out to sixty light-years and gathering up about eight thousand solar masses of interstellar matter. Newly created shock waves, either the result of expanding nebular gas produced by stellar birth or that of more violently expelled gas produced by stellar death, create "second-generation" stars, some of which in turn explode and give rise to still more shock waves, and so on. Like a chain reaction, old stars trigger the origin of new ones ever deeper into an interstellar cloud—a sequential wave of star formation, aided and abetted by those most-massive stars living fast and dangerously. Observations of young stars in the vicinity of supernova remnants do imply that the gentle births of new stars is often initiated by the explosive deaths of others.

If we think broadly enough and over long enough timescales, stars begin to seem as replicative as bugs in a petri dish. Though it is perhaps a bit of a stretch, stars nonetheless provide good examples of physical evolution. As stars naturally change over time, their interiors develop steeper gradients in temperature and elemental composition; their cores heat up and the heavies are created. Stellar size, color, brightness, and makeup all change, while progressing from protostars at "birth," to mature stars at middle "life," and on to red giants, white dwarfs, and pre-supernovae near "death." At least as regards energy flow, matter circulation, internal gradients, and nonequilibrium while undergoing change, stars have much in common with life.

None of this is to say that stars are alive, a common misinterpretation of such an eclectic stance. Nor do stars evolve in the strict and limited biological sense—a subject best examined more closely later, in the Biological Epoch. Yet close parallels are apparent, including selection, adaptation, and perhaps even generational offspring among the stars. It's all reminiscent of a Malthusian-inspired scenario, hereby liberally stated:

Galactic clouds spawn clusters of stars, only a few of which (the more massive ones unlike the Sun) cause (via supernovae) other, subsequent populations of stars to emerge in turn, with each generation's offspring showing slight variations, especially among the heavy elements contained within. Waves of star formation propagate through many such clouds like slow-motion chain reactions over eons of time—shocks from the death of old stars triggering the birth of new ones—no single kind of star displaying a dramatic increase in number nor the process of regeneration ever being perfect. Those massive stars selected by Nature to endure the fires needed to produce heavy elements are in fact the very same stars that often foster new populations of stars, thereby both gradually and episodically enriching interstellar space with greater elemental complexity on timescales measured in millions of millennia. As always, the necessary though perhaps not sufficient conditions for the growth of complexity depend on the environmental circumstances and on the availability of energy flows in such (here, galactic) environments.

On and on, the cycle churns: buildup, breakdown, change—a kind of stellar "reproduction" minus any genes, inheritance, or overt function, for these are the value-added qualities of biological evolution that admittedly go well beyond stellar evolution.

∞

What remains in the aftermath of a supernova explosion? Is the entire star just blown to bits and ejected into the surrounding interstellar medium? Astronomers aren't quite sure, though most computer models predict that some portion (perhaps twenty percent) of the star survives. Like planetary nebulae that expel matter less violently and bequeath a white-dwarf core remnant, most supernovae are also expected to leave behind tiny, severely compressed cores. The matter within this centralized cinder is unlike anything found on Earth; indeed it constitutes one of the strangest states of matter in all the Universe.

During the moment of implosion of a massive star, just prior to its explosion, all the electrons in the core violently smash into the protons. The electrons were there all along, given the plasma state extant throughout, but the protons are freed when some of the heavy nuclei disintegrate under the phenomenal onslaught. The result is an elementary-particle reaction that races through the core of the massive star, converting within seconds all the electrons and protons into neutrons and neutrinos. Though the neutrinos rapidly leave the scene at nearly the speed of light, they are suspected by many theorists of playing a major role in triggering supernovae. The reason is that neutrinos transport much of the energy of the collapsed core to the overlying layers of the star, deposit it there, and cause the rest of the star to detonate like a colossal bomb. In contrast to our Sun, where the neutrinos escape in seconds without hesitating, the collapsed cores of more massive stars are so much denser that the neutrinos are actually stopped in their tracks before escaping—but only momentarily, as their accompanying energy is sufficient to blow the rest of the star to high heaven.

Made of nuclei much heavier than the neutrinos, the shattered debris from everything above the star's core departs at fast speeds, though much less than the velocity of light. These are the wisps and filaments seen today still racing outward from old supernova remnants. Only the core remains intact as an ultradense ball of pure neutrons hardly more than a few kilometers across—about the size of an asteroid. Astronomers breezily call this core remnant a neutron star, but it's not really a "star" in the true sense of the word since all its nuclear reactions have ceased forever.

. . . a neutron star is not much larger than a typical city.

The theory of stellar evolution predicts that neutron stars are very small, though very massive. Composed simply and solely of neutron particles crammed into a tight-knit sphere, a neutron star is not much larger than a typical city. Despite its diminutive size, each unexploded core remnant usually contains a few times the mass of our Sun, making neutron stars extraordinarily compact. Their average density is estimated to reach at least a quadrillion times that of Earth's rocks. Not merely a huge density, this is an incredible density, nearly a billion times denser than the already supercompact white-dwarf stars; a single thimbleful of neutron-star stuff would weigh a hundred million tons on Earth. In fact, the density of a normal atomic nucleus is not much greater; neutron stars are just about as compressed as the matter within the nuclei of normal atoms. Such an extraordinary density was already encountered in the tale of cosmic evolution during the first, Particle Epoch of the Universe. The weird neutron stars might therefore act as "primeval laboratories," enabling scientists to study the physical conditions prevalent just after the start of the Universe.

Once stars explode as supernovae, all nuclear events end. Calculations predict that the remnant neutron stars are probably solid objects, more like planets than stars. We might imagine standing on one, provided it's cooled enough, though it wouldn't be easy; a neutron star's gravity is unbelievably powerful. A person weighing an Earthly seventy kilograms (hundred-fifty pounds) would weigh the equivalent of about a billion kilograms (a million tons) while on a neutron star. Actually, standing on one wouldn't even be possible, for the severe pull of gravity would instantly squash a person to the thickness of a postage stamp. Gravity is so strong on a neutron star that the entire population of the world, if shipped there, would be crushed to a volume the size of a pea!

Can we be sure that objects as strange as ultradense neutron stars actually exist? The answer is again yes. In the past few decades, radio and X-ray astronomers have made some remarkable discoveries, proving that neutron stars are real. These observers regularly monitor rapidly blinking stars, or pulsars for short, that emit short radiative pulses lasting about a hundredth of a second apiece. Each pulse contains a burst of radiation, after which there is nothing. Then another pulse arrives, again and again ad infinitum. We humans are insensitive to such rapid flashes (even when they occur optically), making it impossible to observe a pulsar's flickering by eye, but suitably designed instruments can record the quick machine-gun-like signals. The time intervals between pulses are so astonishingly uniform—good to one part in a hundred million (thus the equivalent of losing one second in ten years)—that the repeated emissions can be used as a highly accurate clock. Yet even these cosmic timepieces are not perfect, for they too change. Over billions of years, pulsars gradually slow down and die.

More than a thousand pulsars have been charted among the cindral remains of supernova remnants, though not all such debris fields have pulsars buried within them. Perhaps the latter are examples of those stars that did blow themselves totally to smithereens. The most prominently studied pulsar resides close to the center of the Crab Nebula, the site of that tumultuous explosion observed nearly a thousand years ago and now known to have occurred about six thousand light-years away—therefore whose explosion actually took place some seven thousand years ago. By determining the speed and direction of travel of the ejected matter observed for that supernova remnant, researchers have been able to trace the debris backward, pinpointing the location in space at which the progenitor star exploded. There the supernova core remnant is expected to be located. And that's exactly the region in the nebula from which the pulsing signals arise. Apparently, pulsars are indeed the dregs of the once-massive stars—in the case of the Crab, one that launches a signal nearly thirty times each second.

Astrophysicists reason that the only physical process consistent with such precisely timed signals is a small, spinning source of radiation. Only rapid rotation—not pulsation—can cause the high degree of regularity of the observed pulses. Typically, once around on its axis in a second makes for a pretty amazing cosmic object, even more so for the Crab at thirty times faster. And only an object less than ten kilometers in diameter can account for the sharp, snappy quality of each pulse; radiation emitted by a larger object would arrive at Earth at slightly different times, blurring the pulse to a droning hum. Normal stars, even white dwarfs, would be completely torn apart by such rapid rotation; they're too big in size and too weak in gravity to hold themselves together. It's hardly surprising then that the best theoretical model of a pulsar envisions a small, compact, spinning neutron star that periodically flashes radiation toward Earth. The experimentalist's "pulsar" and the theoretician's "neutron star" are synonymous.

According to leading theoretical ideas, a "hot spot" on or near the surface of a neutron star, or perhaps in the atmosphere above it, continuously emits radiation in a sort of narrow "searchlight" beam. This spot could be akin to a violent surface quake or atmospheric storm much like flares on the Sun or volcanoes on Earth. Most likely, it's a localized region near the neutron star's poles, where charged particles, possibly ripped from a companion star and accelerated by intense magnetism to extremely high energies, emit radiation along the neutron star's axis—not entirely unlike the active galaxies whose accretion disk around a central black hole spews forth fast, magnetically guided electrons into huge lobes along polar axes orthogonal to the disk. Magnetism comes into play since, during collapse of the progenitor star, any magnetic field would be amplified to as much as trillions of times that on the Sun. Although the energized spot sprays radiation steadily into space, the star's spin rate of typically once per second guarantees that the emitted radiation in any direction behaves like a hail of discrete bullets. The radiation sweeps through space like a revolving lighthouse beacon or a spinning lawn sprinkler. Arriving at Earth, perhaps thousands of years later, the rapid pulses are captured by our telescopes—provided that the star is oriented so that its beam sweeps past our planet. The details of the theoretical model are sketchy and controversial, for researchers have little hard information about the behavior of matter with densities so great that an entire kilometer-sized object begins to resemble a cyclopean atomic nucleus.

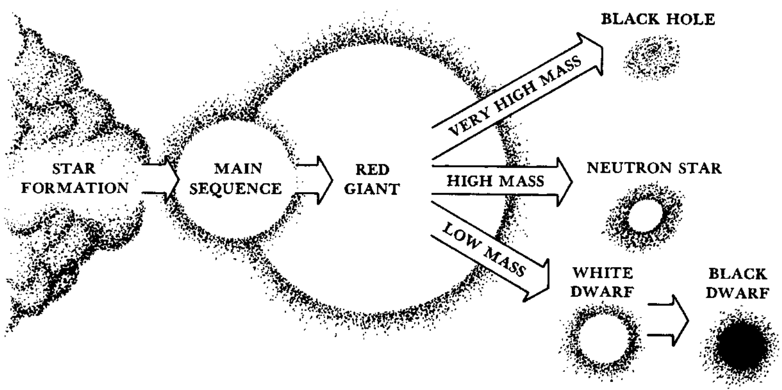
Neutron stars are indeed outlandish objects. Even so, modeling stipulates them to be more or less stable, much like most other stars. In this case, however, neutron stars are not balanced by the inward pull of gravity battling the outward pressure of hot gas. As best we can tell, neutron stars have no hot gas. Instead, the outward force arises from the crystalline nature of the tightly packed neutrons themselves: quantum theory restricts the number of neutrons occupying any one space, and that's sufficient to buoy the star against its own gravity. Side by side and virtually touching one another, the neutrons form a solidified ball of matter resembling, but much denser than, the compacted electrons in a white-dwarf star. Compositional details differ, with some models implying a crusty surface of hard iron ash, others claiming whole new superconducting materials, and still others proposing a core of unspecified "strange matter" that levitates the iron crust above it, but twentieth-century science could not reach a consensus on these true oddballs. Any object having strong magnetism and rapid rotation will have an unsure status since modern physics is still unable to solve this problem, even with high-speed computers and assumptions of spherical symmetry. The essential feature of neutron stars is that their hypercompressed neutrons generate enormous pressure that not even gravity can counter—with one notable exception.

Hypotheses have been advanced that galaxies ought to naturally house stellar core remnants with masses so large that the inward pull of gravity can in fact crush even the seemingly incomprehensible sphere of pure neutrons. According to some theories, should enough matter (at least three solar masses) be packed into a small volume (no bigger than a kilometer or so), the collective efforts of gravity can overwhelm *any* countervailing force. In this case, gravity becomes powerful enough to compress an entire star into an object the size of a planet, a city, a pinhead, or even smaller!

As the core of a formerly massive star shrinks, the gravitational pull in its vicinity eventually becomes so great that light itself is unable to escape beyond its event horizon; the light would return much as baseballs fall back to Earth when thrown into the air. Such freakish objects are expected to emit no light, no radiation, no information whatsoever—providing grist for both natural scientists' models and science-fiction writers' dreams. Incommunicado, such a massive star effectively collapses onto itself and utterly vanishes—into a "hole" typically less than a kilometer across, but a hole nonetheless into which all nearby matter falls, trapped by its own gravity perhaps forever. These are the most bizarre end points of stellar evolution, encountered earlier on much larger scales in the Galactic Epoch as candidates for the vibrant cores of active galaxies. These are the celebrated black holes.

∞

Here again is the sequence of events expected if these late stages of stellar evolution are valid: A star having at least several times the Sun's mass ends its burning cycle by exploding as a supernova. Much of the star's original content is ejected as fast-moving debris. Provided a few solar masses of material remain behind in the core, the unexploded remnant collapses catastrophically, the whole core diving below the event horizon in less than a second. The core simply winks out—not merely becoming invisible but literally disappearing—leaving a small dark region from which no radiation or matter, indeed nothing at all, can escape. This is the way black holes are born as blackened domains in space. These bizarre end points of stellar evolution are not really objects as much as holes—black holes in the fabric of spacetime.



. . . the sequence of events expected if these late stages of stellar evolution are valid.

So what's the story? Do black holes really exist, or are they nothing more than figments of theorists' fertile imaginations? Maybe all massive stars are blown to bits when they explode as supernovae, never leaving much of a remnant core at all. Or perhaps another, as yet undiscovered force is capable of competing with gravity despite these extreme conditions of ultracondensed matter. Each of these possibilities would preclude the existence of black holes—though what remained in their place would still be awfully bizarre. Just how much observational evidence is there for black holes?

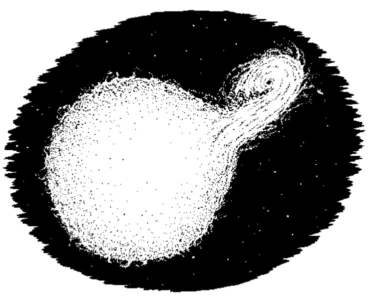
In the Galactic Epoch, black holes were invoked as the most likely answer to the vast energy pouring forth from the centers of galaxies and quasars. Despite the term *black* holes, the environments around such regions are expected to be intense sources of emission, the result of matter falling into the clutches of the holes themselves, heating and radiating all the while. In earlier times, when the Universe was more violent and not much matter was yet wrapped up into organized structures, the conditions were ripening for the formation of supermassive black holes at the hubs of the emerging galaxies. Astronomers now reason that what we perceive today at most galaxy centers—though little of it is seen directly—are the galactic remnants of erstwhile holes, millions to billions of times more massive than for mere stellar remnants, some apparently still gulping matter. How sure are we of all this theoretical modeling? In truth, despite a growing consensus in the astronomical community that black holes are probably real, we are unsure how sure.

Black holes may be invisible, but that doesn't mean they lack gravity; they still house mass and thus still exert gravity. Accordingly, astronomers can test for holes' existence by probing the gravitational force in and around the space near them. For example, the motion of a spacecraft or of a nearby celestial body could conceivably be used to study the nature of a black hole. Any object outside an event horizon should behave just as though a massive, visible object resided at the site of the hole. In other words, all conventional means of assessing a black hole disappears, yet its gravity persists. If astronomers could find some small, neighboring black holes, perhaps we could examine, if not them directly, then at least their environments. The closer they are to us, preferably in our own Milky Way, the better we should be able to resolve their features, infer their mass, and test our ideas about them.

Surely, our civilization doesn't have the capability to maneuver spacecraft into the neighborhood of suspected black holes, even if we knew their exact locations. However, the Galaxy is populated with many double-star systems whose members orbit about one another. Yet for many of these only one star is visible, the other betrayed only by indirect means such as periodic spectral shifts or starlight dimming. Of course, each unseen companion could be just a dwarf star hidden in the glare of a bright stellar partner. Or the unseen object could be shrouded by interstellar dust, making it invisible to equipment on Earth, but not necessarily indicative of a black hole. And these alternatives are probably true for the majority of binaries having an unrevealed member, in which case the invisible candidates are not black holes.

A few of these atypical binary systems, however, display peculiar properties strongly suggestive of black holes. Pioneering observations, made by Earth-orbiting satellites as long ago as the 1970s, revealed binaries that emit copious amounts of X-ray radiation. This high-frequency emission cannot easily penetrate dust, making it unlikely that galactic debris has camouflaged one of the systems' partners. More recently, advanced satellites equipped to pinpoint X rays from suspected targets have been monitoring just this type of radiation from several binary systems, some of which show only one star visually. For each, the X-ray radiation arises from million-degree-Celsius gas flowing from a large visible star toward a small, invisible companion. Furthermore, each invisible member harbors several solar masses and spans no more than a hundred kilometers. That's several times the mass of our Sun all compacted within the size of a large city. These properties have all the earmarks of black holes.

A tentative model for the nearest and best case for a "classical" black hole—Cygnus X-1, some seven thousand light-years from Earth—stipulates that much of the gas drawn from the visible star flows onto a nearby, tire-shaped oval of matter. This is probably an accretion disk where the superheating occurs and the X-ray radiation is launched, often variably and implying that the unseen companion is very compact—a neutron star or black hole. If the latter, some of this gas inevitably streams toward the black hole, sucked ever deeper into its whirlpool and eventually trapped forever by it.



. . . gas drawn from the visible star flows onto a nearby, tire-shaped oval of matter.

A handful of other stellar black-hole candidates are frequently monitored in or near our Galaxy, each of them displaying observational traits similar to those of Cygnus X-1. And far beyond, extremely intense gamma-ray bursts have been detected nearly daily by civilian and military satellites during the past few decades, possibly the result of two neutron stars first colliding and then triggering a supernova (or "hypernova") that instantaneously releases monstrous flashes of narrowly collimated energy just before the whole mess core-collapses into a black hole. But nagging problems plague interpretations of most of these as black holes, not the least being that all the suspected holes in binary systems have masses close to the neutron star–black hole dividing line of a few solar masses. When the effects of rotation and magnetism are fully incorporated into the physics of dead stars—tasks untenable at present—there's a chance that the dark objects in question may turn out to be more mundane neutronlike stars and not foreboding black holes at all. Or perhaps something else entirely.

"Quark stars" come to mind as cutting-edge, theoretical alternatives to black holes—and that may be where they will stay as none has yet been unambiguously found anywhere in Nature. Such strange objects, purportedly at least a thousand times denser than neutron stars, or a billion billion times more compact than in the core of our Sun, would nonetheless be a substitute for black holes. A quark star would form when neutronic matter in a neutron star cracked under unbearable strain, splitting many of the neutrons back down into their constituent quarks while compressing the star to the absolute limits of known physics, yet without, however, plunging it indefinitely into a black hole that seems beyond physics. Strange quarks, strange stars; this matter, if real, has reverted to the form in which it was born a microsecond after the big bang.

That said, the payoff would be great if astronomers could demonstrably prove the existence of a nearby black hole. It would become a virtual laboratory in space for the study of bizarre states of matter ruled by quantum gravity. Such peculiar phenomena so foreign to our everyday senses are thought to mimic, more than any other astronomical entity, the decidedly unearthly conditions prevalent near the very start of the Universe. Ultimately and ironically, these burned-out corpses of massive stars may someday help us comprehend one of the foremost quandaries in all of science—the emergence of that singular, superhot, superdense state from which the Universe itself originated.

∞

Stellar evolution well explains how myriad stars in the celestial sky form, mature, and die. It enables us to date numerous types of stellar objects scattered across the Galaxy, ordering each of them into a consistent temporal sequence along the arrow of time. And it shows how all those many varied objects—from galactic clouds and protostars to red giants and white dwarfs, and including nebulae, pulsars, and supernovae—fit into an overall framework of understanding based on the unifying concept of change. These varying interrelationships among the many diverse components of stellar and interstellar matter in our Milky Way compose nothing less than a "galactic ecosystem," an evolutionary posture nearly as intricate and delicate as life in a tidal pool or a tropical forest.

Without a theory of stellar evolution, we would witness but a huge unrelated zoo of objects strewn throughout space. Astronomers would resemble stamp collectors, with many data but not much insight into how one cosmic object relates to any other. With stellar evolution, we enjoy a more powerful position, intellectually. We can place each type of cosmic object into a comprehensive evolutionary perspective, thereby deciphering both the bigger picture of all the stars in toto as well as their detailed relationships among odd and varied kinds of stars. Indeed, we can indirectly follow their evolutionary tracks, as one type of celestial object changes into another.

In contrast to our poor grasp of galactic evolution, astronomers know much about stellar evolution. The way stars change, especially elemental evolution—whose theory matches the observed cosmic abundances remarkably well—is one of the best developed and well understood aspects of cosmic evolution. The subject of stellar evolution can naturally account for the observed differences in elemental abundance between the old, globular-cluster stars, which are perhaps as ancient as the Milky Way itself, and the young, open-cluster stars that more recently formed in our Galaxy. The oldest stars have few heavy elements, the youngest stars the most heavies. All things considered, our knowledge of stellar aging and elemental creation is surprisingly robust, given that the stars are so far and foreign.

Though some of the evolutionary events described here in the Stellar Epoch are cyclical, they actually lead slowly and surely toward increased energy flows and ordered material structures. Operating at countless localized sites within the vastly larger galaxies, the ongoing process of star birth, maturity, and death constantly enrich and fertilize interstellar space with heavy elements that sow the seeds for later-generation stars—as well as planets. And much as elsewhere in our story, with time comes change, and with change comes rising complexity.

On and on, that cycle does churn. Stellar buildup, breakdown, change. Dust to dust, and to dust some more—in this case stardust engaged in a kind of cosmic reincarnation. From the ashes of dead stars come the origins of new ones. Indeed, without the heavy elements made inside stars, both life on Earth and Earth itself would not exist.

4. PLANETARY EPOCH

Habitats for Life

PLANETS ARE GLOBES OF SOLIDS, liquids, and gases, smaller than stars and made partly of heavy elements. These worlds could not have formed early in the Universe. There simply were no appreciable heavy elements in the first few billion years. Planets had to await the birth and death of countless high-mass stars, the only known locales suitable for molding the heavies. Planets, then, are quite literally collections of the cinders of burned-out stars—balls of matter hardly relevant in the cosmological scheme of things, yet comfortable abodes for us sentient beings seeking to decipher (or is it create?) that larger worldview.

The origin of the planets and their mottled moons is a challenging, as-yet-unsolved problem. Most of our knowledge of the Solar System's formative stages comes from studies of galactic clouds, fallen meteorites, and the Sun and Moon, as well as from various planets and moons examined from a distance with peering telescopes and sometimes closer with robotic probes. Added understanding now also comes from observations of planets circling nearby stars, a new and exciting area of research that has led to an avalanche of discoveries in recent years. Ironically, studies of Earth itself do not help much since our planet's youngest stages eroded away long ago.

All signs point to an Earth that formed nearly five billion years ago. Initially cold, our planet heated enough to melt completely, partly from without because of violent, macroscopic asteroid bombardment but mostly from within owing to serene, microscopic radioactive decay. During its first half-billion years of existence, Earth's interior evolved, its crust solidified, and much of its atmosphere escaped. Change was initially rampant, energy flows surging. And although that change slowed thereafter, it produced nonetheless mountain ranges, oceanic trenches, and atmospheric rejuvenation. As environmental change continued, the stage was progressively set for the origin and evolution of life—including human life, an especially interesting, if seemingly anthropic, chapter in the cosmic-evolutionary story.

∞

The planetary group in which we live is a varied lot. The Solar System includes one star, nine planets, more than a hundred moons, thousands of asteroids ranging in diameter from meters to hundreds of kilometers, countless comets of kilometer dimensions, and myriad meteoroids less than a meter across. With the Earth-Sun distance of about a hundred-fifty million kilometers termed an "astronomical unit," the whole Solar System extends end to end for nearly eighty such units. That may sound large, but it's only about a thousandth of a light-year, hardly more than a billionth the size of our Milky Way. Planetary systems occupy spaces much smaller than the distances separating stars—at least out in the galactic suburbs where we live—making each such system a celestial island unto itself.

The four innermost planets—Mercury, Venus, Earth, and Mars—are often termed the Terrestrial Planets because of their physical and chemical similarity to rocky Earth. In contrast, the larger, outer bodies—Jupiter, Saturn, Uranus, and Neptune—are called the Jovian Planets because of their resemblance to gassy Jupiter. Between these two groups, in a broad band, or thick "belt," some two to three astronomical units from the Sun (roughly between Mars and Jupiter), roam the stony asteroids, sometimes labeled "minor planets" or even "planetoids," for they are actually not starlike at all. Pluto, usually the outermost planet (though it sometimes ventures inside Neptune's orbit, as it did between 1979 and 1999), doesn't fit well into any of these categories. Much smaller than Earth's Moon, the literal oddball Pluto might have once been a moon of Neptune or a distant asteroid and not originally a planet at all. Despite recent attempts to reclassify it as merely a large icy stone (yet the most prominent member of the newly discovered Kuiper belt of trans-Neptunian rocks), official retreat from its historical designation as a planet seems unlikely.

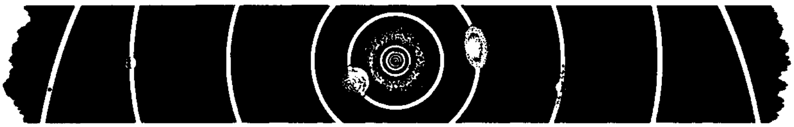
From a remote vantage point far beyond our home planetary system, the Sun overwhelmingly dominates our cosmic neighborhood, with Jupiter an inferior second. Our star has a mass more than a thousand times Jupiter's and about seven hundred times that of the whole rest of the Solar System, including Jupiter. The Sun, then, houses more than ninety-nine percent of all the matter in the Solar System. Everything else, especially the small Terrestrial Planets and notably Earth, resembles a collection of nearly insignificant debris.

Draw a distinction in Jupiter's case, however, for this is no ordinary heavenly body. Jupiter in fact just missed becoming a star. The composition and structure of this giant planet—and possibly all the big Jovian Planets—is largely stellar. They are rich in hydrogen and helium, light gases that have long ago escaped from the smaller Terrestrial Planets. But none of the Jovians is quite big enough to ignite—to start a thermonuclear reaction at its core by virtue of its own overlying mass. Had Jupiter gathered several tens of times more matter, its central temperature would equal that needed to commence nuclear fusion, converting it into a dwarf star. Thus, our Solar System almost formed as a binary-star system, an astronomical posture that would have rendered Earth life improbable, perhaps impossible.

We owe a debt of gratitude to the Sun for lighting up, and to Jupiter for not.

Erstwhile, growing complications of a perceived clockwork Solar System—especially the retrograde motions of some of the planets, such as Mars—were greatly simplified in Renaissance times. Looking (observation) and thinking (theory) combined to build more objective mental models than those deduced by the ancients; testing became a vital part of the process of inquiry. The sixteenth-century Polish cleric Nicholas Copernicus recognized that a heliocentric (Sun-centered) model improved the harmony of the tangled geocentric (Earth-centered) models proposed by the Greeks and Romans of old.

Despite the support of empirical data and a mathematical underpinning by two seventeenth-century scholars, the German Johannes Kepler and the Englander Isaac Newton, the Copernican model was not easy to accept even as recently as a few hundred years ago. Heliocentricity



. . . a heliocentric model improved the harmony of the tangled geocentric models . . .

rubbed against the grain of all previous logic, and it violated many religious teachings of the time. Above all, it relegated Earth to a noncentral and undistinguished location within the Solar System and the Universe. Earth became just one of many planets.

Although we now realize that these Renaissance workers were correct, none of them was able to prove to his contemporaries that our system is centered on the Sun, or even that Earth moves. Unambiguous proof of the latter came only in the mid-nineteenth century when the German astronomer Friedrich Bessel first observed stellar parallax—the yearly to-and-fro artificial motion of a nearby star caused by Earth's real motion around the Sun. Heliocentricity of the Solar System has been verified repeatedly over the years with an ever-increasing number of experimental tests, culminating with the recent expeditions of our robot space probes that have toured through an obviously Sun-centered planetary system.

Initial motivation for the heliocentric model was simplicity, at least in the mind's eye. Heliocentricity provides a more natural explanation of the observed facts than can any geocentric model. Even today, scientists are often guided by simplicity, symmetry, and beauty in modeling all aspects of the Universe. Those models in science having a measure of elegance are often closer to reality; those that are complicated are usually wrong.

Development and eventual acceptance of the heliocentric model is an awesome milestone in our thinking as human beings. Discovering the framework of our planetary system freed us from an Earth-centered view of the Universe and enabled us to realize that ours is only one of many planets orbiting the Sun. Surprisingly, it was less than a century ago that the American astronomer Harlow Shapley took the next bold step, in turn proving that as a resident of the suburbs of the Milky Way, neither was our Sun centralized, unique, or special in any way. The more we look and the more we test, the more mediocre our niche in the Universe seems to be.

∞

Any model capable of explaining the origin and architecture of our planetary system must adhere to the known facts. Generally, these facts derive from studies of interstellar clouds, landed meteorites, and Earth's Moon, as well as from observations of numerous planets both within and beyond our Solar System. The meteorites provide especially useful information, for they contain entrapped traces of solid and gaseous matter uneroded from the early Solar System. Radioactively determined dates of all meteorites uniformly imply that our system formed, with the Sun and Earth as part of it, some four-and-a-half billion years ago. Laboratory analyses of the oldest lunar rocks generally confirm this date, as does theoretical modeling of the Sun itself.

Among the many observed properties of our Solar System, seven stand out most boldly:

Each planet is relatively isolated in space, none of them being bunched together; each planet (out to Saturn anyway and skipping the asteroid belt) resides roughly twice as far from the Sun as its next inward neighbor, implying a certain geometric harmony—the kind of order and elegance alluded to earlier.

The orbits of the planets describe nearly perfect circles, with only two exceptions; Mercury's noticeable elliptical orbit is surely caused by this innermost planet's proximity to the neighboring Sun, while Pluto's more pronounced eccentricity likely results from this outermost planet being more an asteroidal rock than a genuine planet.

The orbits of the planets all lie in nearly the same plane, Earth's such plane being called the "ecliptic"; each of the planes swept out by the planets' orbits aligns with the others to within a few arc degrees (excepting again Mercury and Pluto), the whole system of planets having the shape of a rather flat disk.

The direction in which the planets orbit the Sun is the same in which the Sun rotates on its axis (counterclockwise from terrestrial north); virtually all the angular momentum in the Solar System—the planets' orbits and the Sun's spin—seems systematized, again implying a high degree of unison.

The direction in which most planets rotate on their axes also mimics that of the Sun's spin (again counterclockwise); the two exceptions are Venus, which spins oppositely (retrograde), and Uranus, whose poles are tipped over so as to lie in the plane of its own orbit. (Pluto might also be out of bounds—again.)

Most of the known moons revolve about their parent planets in the same direction as the planets rotate on their axes; some moons, like those associated with Jupiter, resemble miniature Solar Systems, revolving about their parent planet in roughly the same plane as the planet's equator, once more evincing unison throughout our planetary system.

Finally, the Solar System is highly differentiated; the inner, Terrestrial Planets are characterized by small sizes, rocky makeup, high densities, moderate atmospheres, slow rotations, and few or no moons and rings, whereas the outer, Jovian Planets have large sizes, gaseous makeup, low densities, thick atmospheres, rapid rotations, and many moons and rings.

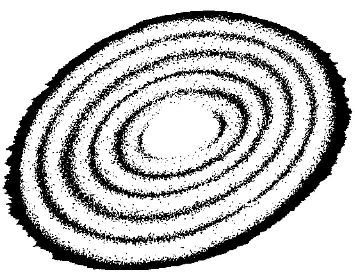
All these observed properties, when taken together, clearly denote a high degree of order within our Solar System. Although much diversity prevails among individual planets and moons, the whole ensemble is apparently not a random assortment of objects spinning and orbiting this way or that. It hardly seems possible that the Solar System is a pickup team, amassed by the slow accumulation of already-fashioned interstellar bodies casually captured by our Sun over the course of billions of years. The overall architecture of our Solar System is too neat and tidy, and the ages of its members too uniform, to be the result of chaotic events or haphazard circumstances. All signs point toward a single formation, the product of an ancient but one-time event not quite five billion years ago.

A comprehensive account of all these properties has been a principal goal of astronomers for well more than a century. The Solar System is, after all, our extended home in space, and it would be nice to know, specifically and in detail, how it all came to be.

Though not all these planetary properties were known hundreds of years ago, the crux of the modern theory of our Solar System's origin dates back at least that far. Called the nebular model, the original idea is often attributed to the German philosopher Immanuel Kant, but he merely elaborated upon an earlier proposal made in the seventeenth century by the French philosopher René Descartes. In this conceptual model, a giant, swirling gas cloud gradually contracts to form a central Sun, and the planets and their moons are assumed to be natural by-products of the star formation process. But these philosophers failed to work out the mathematical details of their models; their proposals amounted to little more than qualitative words and untested ideas.

Later in the eighteenth century, the French mathematician-astronomer Pierre-Simon de Laplace tried to give this type of model a quantitative basis. Using angular-momentum arguments, he showed mathematically that gaseous bodies spin faster as they contract. A decrease in the size of a rotating mass must be balanced by an increase in its rotational speed, much like a pirouetting figure skater who spins faster while closely retracting her arms, or a high diver who somersaults quickly by tightly curling his body. An interstellar cloud would eventually flatten into a pancake-shaped disk, for the simple reason that gravity can pull matter toward the center of the region more easily along its rotation axis than perpendicular to it—which is why a spinning body tends to develop a bulge around its middle. This model provides a plausible origin of some of the ordered architecture observed in our Solar System today—the planets' near-circular orbits, their residence in a well-defined disk, and many of the other properties just listed. These properties are among the natural results of simple changes expected in any galactic cloud, a straightforward obedience of a parcel of gas to the known laws of physics.

Continued contraction of such a primitive Solar System forces the entire cloud to spin more rapidly as time proceeds. Near the fringe, the outward centrifugal push eventually exceeds the inward gravitational pull. This push creates a thin ring of gaseous matter that breaks away from the rest of the system, which in turn contracts a little more until such time as another ring of matter is deposited inward of the first. Progressing in this way, an entire series of rings were imagined by Laplace to form around the central protosun. Each ring is furthermore theorized to condense, over long intervals of time, into a planet. Several outer planets might develop quickly while the interior of the early Solar System continued to shape the inner planets and the Sun.



Each ring is theorized to condense, over long intervals of time, into a planet.

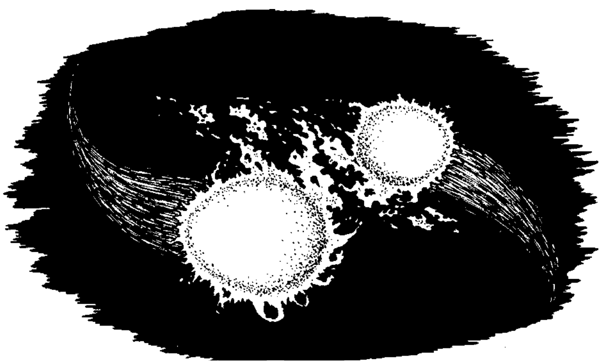
As sensible as this nebular model seems, it's not without difficulties. Detailed analyses show that material in a ring of this sort would not likely assemble into a planet. In fact, computer simulations predict just the opposite. The rings would tend to disperse, owing to both a wealth of heat and a lack of mass within any one ring. Gravitational clumping of interstellar matter is one thing—it works reasonably well when making stars because vast amounts of mass are often housed within a typical, cold galactic cloud. But the coagulation of a warm, protoplanetary ring is another thing—not nearly enough matter is present for gravity to best the heat and thereby gather the gas into a planet-sized ball. Instead of coalescing to form a planet, computations predict the ring will break up and fade away.

Don't be too hard on Laplace. He didn't have a computer, and it's both tricky and tedious to account for all the statistical subtleties of this problem without one. Even today, experts disagree on some of the details in the best computer models, as noted below.

A second problem further complicates the nebular model for the origin of the Solar System. It's well known that the Sun spins on its axis once in about thirty days, a good deal more slowly than Earth. This solar sluggishness baffles astronomers for one simple reason: although the Sun contains more than a thousand times the mass of all the planets combined, it boasts less than two percent of the system's angular momentum. Jupiter, for instance, has a lot more momentum than our Sun. Not merely does it spin on its axis so fast (less than ten hours to go once around), but for an object with its sizable mass so distant from the Sun, Jupiter carries a great deal of orbital momentum. In fact, Jupiter presently harbors more than half of the Solar System's total momentum. All told, the four big Jovian Planets account for some ninety-eight percent of the momentum of our Solar System. By comparison, the lighter Terrestrial Planets have negligible momentum.

The puzzle here is that the nebular model predicts the Sun should command most of the Solar System's angular momentum. It should be spinning much faster. After all, since the Sun has most of the system's mass, why shouldn't it also have most of its momentum? This is especially true since contracted objects are expected to increase their spin rate, again in the manner of the figure skater's pirouette. Expressed another way: if all the planets, with their large amounts of orbital momentum, were hypothetically deposited inside the Sun, it would spin roughly a hundred times faster than at present. Instead of rotating about once a month, the Sun would spin around once every several hours.

These and other problems with the nebular model forced researchers, for a while at least, to consider alternative ideas—ones that are less evolutionary and more catastrophic. One such idea is embodied in the so-called collision model, which does indeed invoke near catastrophe. Here, the planets are imagined as end products of hot, streaming debris torn from the Sun during a close encounter with another star. The flaming streamers induced by such a near collision are surmised to remain gravitationally bound to the Sun, to be captured into orbits about it, and eventually to assemble into planets. Despite the phenomenal tides surely accompanying the near collision of two stars, the predicted aftermath agrees with the common orientation of the planets' orbits and the Sun's spin, as well as perhaps the close planar alignment of all the planets in a disk.



. . . planets are imagined as end products of hot, streaming debris torn from our Sun . . .

Although first proposed during the eighteenth century, the collision model enjoyed its greatest popularity about a hundred years ago when astronomers not only began realizing that the nebular model was impractical but also began discovering a few minor exceptions to the overall harmony of our planetary system—a retrograde moon of Neptune and a sizable tilt to Uranus, among other irregularities. The absolute beauty and ordered architecture of the Solar System diminished somewhat, giving a boost to models of violence that invoke accidental or unlikely celestial events.

That said, few astronomers take the collision model seriously today. Though it has some points in its favor, models that depend on stellar collisions also have their pitfalls—some of them seemingly fatal. The high improbability of a near collision between two stars is the foremost problem. Stars are large by terrestrial standards but minute compared to the distances separating them, as noted in the Galactic Epoch. For example, the Sun is about a million kilometers in diameter, whereas the distance to Alpha Centauri, the nearest star system, is well more than a trillion kilometers. Probability studies predict that, given the number of stars, their sizes, and their typical separations, not more than a handful of such near collisions ought to have occurred throughout the entire expanse and history of the Milky Way Galaxy (at least outside of the heavily congested galactic central regions). Although galactic collisions are frequent and clearly seen at many locations on the sky, stellar collisions must be extremely rare; in fact, none has ever been observed in the history of astronomy.

The improbability of stellar collisions does not, of course, disprove the idea. Our Solar System could conceivably be the foremost—even the only—example of such an extraordinarily uncommon phenomenon. Should this idea be correct, we can justifiably conclude that our planetary system is an extremely rare type of astronomical system. Very few stars would be expected to have planets and the chances for extraterrestrial life would greatly decrease. However, as noted later in this epoch, recent discoveries of numerous planets orbiting many nearby stars virtually rule out the collision model. Too many extrasolar planets—those beyond our Solar System—are now being found for all of them to be the result of close encounters among stars.

As if the small chance of collision were not enough to badly wound this idea, several other problems plague the notion of planetary origins via encounters of any kind. First, the momentum puzzle besetting the nebular model is again troubling. Second and more formidable, it's hard to understand how hot matter torn from the Sun could contract; hot gases usually disperse. Consequently, although such a near collision between two stars might happen occasionally, it's unlikely that the resulting hot fragments would form planets. Some of the hot streamers would surely fall back into the Sun. Others would tend to dissipate even more quickly than the merely warm matter in the purported rings of the nebular model. A third quandary concerns the nearly circular orbits traced by each of the observed planets. If matter were tidally ripped from the Sun to form the planets, why should each of the clumps of debris end up orbiting the Sun in a near-perfect circle? The collision model cannot explain this observed fact even qualitatively.

The model of Solar System formation most embraced by astronomers today is termed the condensation model. Really a sophisticated version of the nebular concept explained earlier, this model mixes all the attractive features of the old nebular model with our recently revised assessment of interstellar chemistry—or "astrochemistry," that rich and vibrant interdisciplinary area of frontier research previously noted in the Stellar Epoch. Theorists can now concoct a modern condensation model that alleviates several of the aforementioned theoretical problems. And what's more, the new models generally agree with the wealth of observational data now being acquired with today's telescopes and spacecraft.

Recall that the first problem with the nebular model is its inability to assemble ringed material into a tight-knit ball of protoplanetary matter. Each ring would have likely had too little mass and too much heat to begin gravitational contraction. However, a new twist has been added only within the past decade or two. We have come to realize the ubiquity and importance of dust in interstellar space. Dust grains—solid microscopic bodies of rock and ice—are liberally strewn throughout the Galaxy, doubtless the ejected debris from long-dead stars.

Much of our knowledge of interstellar dust comes from meteorite fragments and captured radiation. Ironically, the dust grains within fallen meteorites provides only indirect information even though we can touch those rocks, whereas that information is more direct when we analyze infrared radiation emitted by the dust particles themselves that are far out of reach. The reason is that in rocks the dust is embedded and contaminated, yet in space it's more pristine. Some exceptions are submillimeter-sized, fluffy dust particles collected by high-flying (U2) aircraft in the stratosphere, but these are probably biased samples of chemically altered dust near Earth and not representative of native "stardust" formed in the outer atmospheres of ancient red-giant stars or in the debris fields propelled into space by supernovae. Much of the dust comprises rock (rich in silicon and iron) and ice (mostly dirty water), but a good deal of it also includes carbon, especially a class of organic compounds known by the tongue-twisting name of polycyclic aromatic hydrocarbons (or PAHs, for short), similar to the large benzene-ringed molecules found in cigarette smoke and automobile exhaust. Not surprisingly, then, some grains are made of the widespread interstellar molecules noted in the previous Stellar Epoch. Ubiquitous in space, the dust particles are sized midway between atoms and planets—indeed they reside on the evolutionary path whereby atoms make planets. What's more, given its organic nature, the dust might also have been the source material for the origin of life on planet Earth, a topic best addressed next in the Chemical Epoch.

Such miniature dust grains play an important role in the evolution of any gas. At issue here is the way that thermodynamics works in the presence of gravity. Dust helps to cool warm matter by efficiently radiating away its heat in the form of infrared radiation, thereby reducing the outward pressure of the heat and allowing the inward contraction caused by gravity to proceed more easily. This is the radiation detectable with infrared telescopes, granting information about both the emitting dust and the infalling cloud.

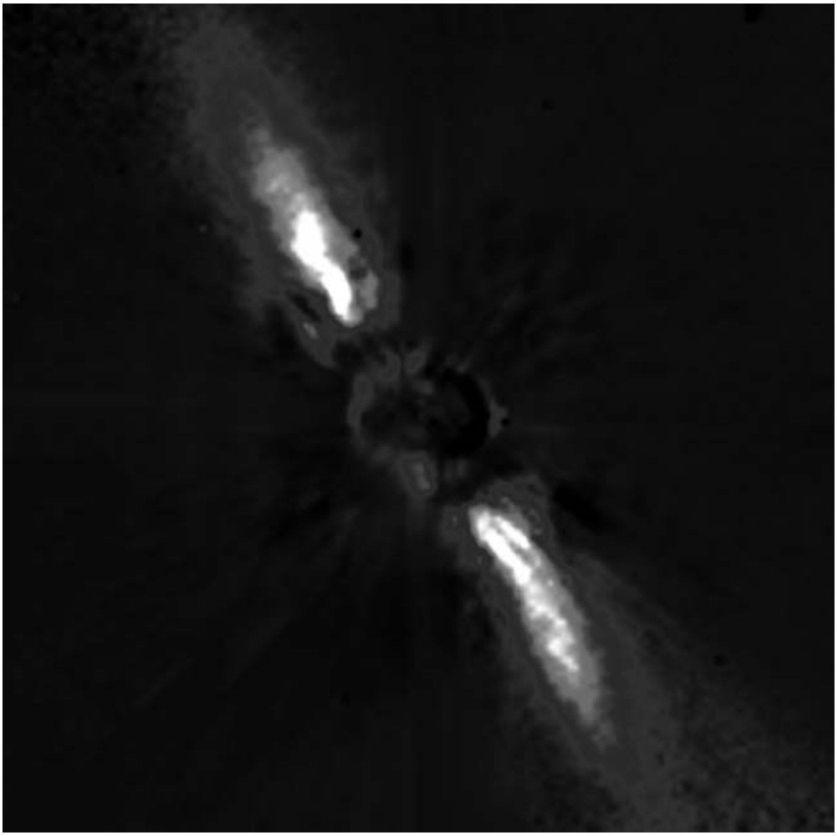
The condensation model, then, assumes that dust was peppered throughout the warm gas of the primitive Solar System, helping to cool it by releasing heat from the protoplanetary blobs. Furthermore, dust grains accelerate the clustering of atoms within the gas, acting as miniature condensation kernels (hence the name of this model) around which other atoms can aggregate, in turn forming larger and larger balls of matter. (This resembles the way that raindrops form in Earth's atmosphere, when dust and soot in the air act as condensation foci around which water molecules cluster.) In short, the presence of dust often guarantees that gravity wins—at least usually, and at least gradually—in the ceaseless tug-of-war between the pressure of heat pushing out and the onslaught of gravity pulling in.

For the particular case of our home in space, by postulating the existence of a dusty interstellar cloud some five billion years ago, theorists reason that dust-grain cooling must have occurred before the gas had a chance to drift away. Accordingly, modern observations of sooty interstellar matter suggest, but do not prove, the likelihood of assembly rather than dispersal of protoplanetary matter. Alas, nagging problems still do remain, yet ones that are actively being tackled observationally, not just theoretically.

Astronomers are fairly confident that the solar nebula formed such a dusty disk long ago because similar disks of loose gas and dust have been observed around young stars not too far away. Foremost among them is the naked-eye star Beta Pictoris, a very young object some sixty light-years distant. Here, copious quantities of dust orbiting the star absorb the starlight, heat up, and, like city pavement on a summer's evening, reradiate the energy. When the light from this star itself is suppressed (with a suitable instrument that blocks receipt of most of its light), a faint disk of warm matter is apparent—especially in the infrared part of the spectrum where dust radiates most strongly. Although this particular disk is roughly ten times the diameter of Pluto's orbit, modeling implies that a star like Beta Pictoris, perhaps as young as twenty million years, is only now undergoing its earliest, somewhat turgid, evolutionary phase, akin to that probably experienced by our own Sun nearly five billion years ago.

The archetypical star-forming region, namely, the Orion Nebula noted in the Stellar Epoch, also provides dramatic confirmation of the above ideas. Direct imagery shows more than a hundred newborn stars throughout the area, each barely a million years old and enveloped by disks of gas and dust seen in silhouette against the nebula's bright background. No planets are seen in the disks, nor are any expected in such juvenile regions. At least several million more years, and perhaps tens of millions more, are needed for genuine planets to emerge from the hodgepodge of gas and dust. This formative sequence has already occurred for the four well-known Trapezium stars now illuminating the nebula, although at a distance of fifteen hundred light-years any planets orbiting them are impossible to discern by any current observational technique—and in any case such planets would be bathed in ultraviolet radiation lethal by human standards.

Not all protoplanetary disks give rise to planets. Many of these disks get blown away by young stars that develop fierce winds shortly after nuclear ignition. This is especially likely for congested clusters such as Orion, where many massive stars energize the nebula, which probably houses thousands of low-mass, Sun-like stars. The process might well resemble a kind of "survival of the fittest" among disks: Those able to withstand the onslaught of ionizing radiation and the battering by loose gas and dust will at least have a chance to form planets from leftover matter. For many "wannabe" planets, it may be a race against time and a battle against blistering radiation. Those disks that manage to coalesce rocky ice balls quickly in potentially hostile surroundings might give birth to planets, and those that don't surely won't. Nature selects some protoplanets to become planets and chokes off the others by destroying



Evidence of planetary disks.

Beta Pictoris is a premier example of several nearby young stars having flattened disks of warm matter surrounding them. Such disks are especially apparent when (as here) the light of the central star is artificially blocked and the disk is sensed by its infrared heat. Beta Pic is some sixty light-years away, and the full extent of its disk, seen here edge-on to our line of sight, measures about a thousand times the distance from Earth to the Sun, or nearly ten times the diameter of Pluto's orbit. *Source: European Southern Observatory.*

their raw materials. As with the star formation noted in the Stellar Epoch, planet formation is a hazardous process with an unpredictable outcome—a little like life in a changing, challenging environment.

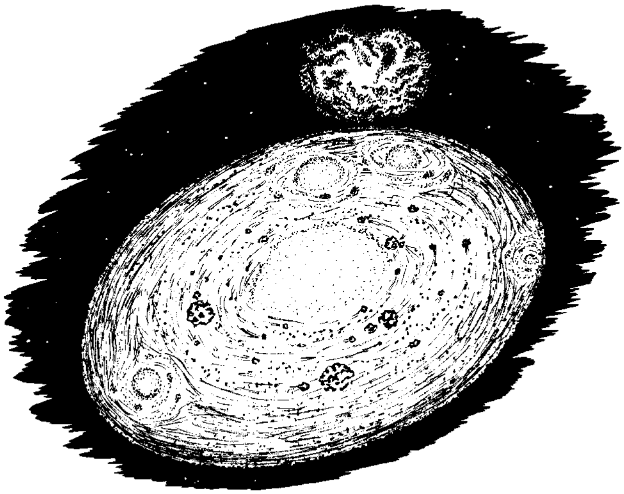
Closer to home, circumstellar disks are now evident at many places in our sector of the Milky Way, wherever the radiation of young stars warms surrounding dust not yet swept away. They include the famous bright stars Vega and Fomalhaut, only twenty and twenty-five light-years distant, respectively, and Epsilon Eridani, even closer at only ten light-years. These and other young stars all emit telltale signatures of infrared emission from surrounding clouds, some of which are warped perhaps owing to perturbations by unseen giant planets. Such planetary nurseries (namely, the disks) are easier to spot than the toddlers (the planets themselves) for the sole reason that the dusty disks are a few trillion times larger than the suspected planets. During the 1990s, astronomers discovered—and mapped in some detail—dozens of disk-shaped regions of emitting dust around adolescent stars. It is in these regions, many of which span solar-system dimensions, that the growth of planets is thought to be now underway. In fact, some of the disks seem to show hollow centers, comparable in size to our own planetary system and perhaps carved out by newborn (yet unseen) planets. These new and exciting findings lend support to the specifics of the condensation model now being fine-tuned by planetologists—those pioneering researchers who are themselves carving out a whole new cottage industry while striving to understand the origin and evolution of our Solar System.

To trace the formative stages of a planetary system such as ours, the modern condensation model stipulates the following broad scenario, starting with a large, dusty interstellar cloud. The original cloud itself might have extended for tens or even hundreds of light-years, but the smaller fragment that would become our home probably spanned no more than about a light-year across. Intermingled with the usual plenitude of hydrogen and helium atoms, the cloud was surely sprinkled with some heavy-element gas and dust ejected from many prior supernovae. Gravitational instabilities started the parent fragment contracting, routinely in fact down to a size of several hundred astronomical units—roughly the size of Beta Pictoris's disk today. All the while, it rotated faster and flattened yet more, after which dense protoplanetary eddies emerged of their own accord.

The initial instability that triggered the infall of our ancestral cloud could have been caused by many events. Perhaps a collision with another interstellar cloud did it, or maybe the passage of a galactic spiral arm; either type of event is suspected to happen relatively frequently—by cosmic standards—roughly once every ten million years. However, the view now favored by the astronomical community is that a nearby supernova was probably the culprit. Old, uneroded meteorites found on the Antarctic ice sheets contain overabundances of certain elements (especially the residues of some mildly radioactive metals, such as iron and aluminum), implying that the genesis of our Solar System might have begun with the concussion of a nearby supernova some five billion years ago. Dating of the meteoritic grains support this idea, implying that the supernova blazed forth less than a few million years before the meteorites condensed into solid rocks. Apparently the ejected debris from the supernova didn't have time to become completely mixed with the primordial matter of our parent galactic cloud before our planetary system formed, the result being microscopic inclusions embedded within today's captured meteorites.

Such supernova explosions inevitably create shock, or blast, waves that compress matter, as noted in the Stellar Epoch. In the case of our parent interstellar cloud, those shocks would have piled up matter into dense sheets, much like snow swept by the blade of a plow. Calculations show that the shocks race around the thinner exterior of the cloud more rapidly than they penetrate its thicker interior. Such sudden pressures would not just compress the cloud from only one direction; they would squeeze it all around. Nuclear weapons tests conducted at the Pacific's Bikini Atoll during the Cold War experimentally demonstrated this squeezing. Shock waves created in the bomb blast literally surrounded buildings, causing them to be blown together (imploded), rather than apart (exploded). In like manner, shock waves can cause the initial compression of an interstellar cloud, after which natural gravitational instabilities divide it into fragments that eventually form stars and planets. Ironically, the demise of old stars may be the trigger needed to conceive not only new stars but whole new worlds as well.

Once the shock wave passed, turbulent, whirling eddies arose naturally in the loose gas and dust at many locations throughout the primitive, rotating Solar System, the bulk of which by this time would have flattened into a Frisbee-shaped disk. As in the earlier cases of galaxy and star formation, these eddies were nothing more than density fluctuations that came and went at random. It's *partly* an issue of statistics, probability, and chance, but it's also the laws of physics at work—a mixture again of randomness and determinism. The phenomenon is akin to the eddies forming in the wake of a spoon stirring a coffee cup or of a hurricane gathering strength while cruising the Atlantic. Provided an eddy could sweep up enough matter while orbiting the protosun, including a rich enough mixture of dust to cool it, then gravity alone would virtually ensure the formation of a planet. The process can be



. . . the demise of old stars may be the trigger needed to conceive . . . whole new worlds.

likened to a snowball thrown through a fierce winter storm; the ball grows fatter while sweeping up more snowflakes in its path. In this way, individual planetesimals the size of moons grew by accretion—the gradual accumulation of small objects by ongoing collision and sticking—and they in turn fashioned protoplanets at various distances from the protosun. The smaller planets eventually and preferentially formed in the inner disk where the amount of matter was less. The larger planets, in the outer disk where more matter naturally settled, likely formed more quickly, as their greater gravity aided and abetted their growth—the rich got richer in the early Solar System. As in any disk, seventy-five percent of its area resides in its outer half, and even if the disk density gradually decreased with distance from the protosun, the ancient disk housed most of its matter well away from the young Sun.

The natural satellites, or moons, of the planets presumably formed in similar fashion but on smaller scales, as mini-eddies of gas and dust condensed in the vicinity of their parent planets. Fragmentation, collisions, and accretion would have aided the growth of miniature solar systems in the gravitational fields of at least the big Jovian planets. Surely, the larger moons formed in this way; the smaller ones may have been chipped off their parent planets during collisions with asteroids; still others may be captured asteroids themselves. Admittedly, the details are lost to times long past, forever irreproducible in our computer simulations, if only owing to the (limited) role played by chance.

Assuming the "sweeping" process of accretion was reasonably efficient throughout the primitive disk, we can appreciate how our present Solar System came to exist as a collection of rather tiny, well-separated planets wheeling around a huge sunny sphere in an otherwise empty region of space. Mathematical modeling and meteorite analyses imply that the bulk of the formation process probably took no longer than ten million years to evolve nine protoplanetary eddies, scores of protomoons, as well as the big protosolar eddy in their midst. Within a few tens of millions of years, the whole region had come to resemble a dirty version of our present Solar System. Nearly a billion more years would have been needed to sweep the system reasonably clear of interplanetary trash.

Those bodies that did not eventually collide with a planet or moon ended up as rocky asteroids in inner belts around the Sun or as icy comets normally resident far from the Sun. Whether a rare and spectacular Halley's comet or the minute debris that showers Earth like the Perseids or Leonids each year, these are all vestiges of an antiquated formative stage. Comets and meteors, then, ought to serve as reminders of birth and construction, not (as in historical lore) as omens of death and destruction.

The weakest link in the condensation model is, once again, the anomalously small momentum of our present Sun. Every quantitative analysis of the young Solar System stipulates the Sun to have spun very fast. Somehow, it must have slowed its rotation dramatically—and friction alone in the early, congested protosolar disk would surely have been a factor—but no consensus has yet been reached on how it actually managed to do so.

Some researchers speculate that the solar "wind," discovered only in the 1960s by some of the first robotic satellites, could have helped slow the Sun's spin ever so gradually over the course of five billion years. High-velocity elementary particles constantly escaping the Sun through flares and other surface storms could conceivably have acted as accruing microscopic brakes to diminish its early rotation. That's because the particles are charged and tied to the Sun's extended magnetism, all of which acts as a drag on any spinning star. The Sun's terrific mass-loss rate of roughly a million tons of matter per second could have indeed robbed it of much of its initial spin, as each and every solar particle must carry with it a minute amount of the Sun's momentum over the course of billions of years. Manned and unmanned space vehicles are now trying to measure the intensity of current solar activity, though it's controversial to estimate the level of that activity billions of years ago.

Other researchers prefer to solve the Sun's momentum problem by postulating a primitive Solar System much more massive than the present-day system. They argue that the accretion process was not overly successful during the system's formative stages, especially in its inner parts where the smaller planets never became massive enough to capture lightweight gases. Matter not captured by the Sun or the planets may well have transported some momentum while escaping back toward interstellar space. The matter that was then lost, or nearly so, might now be in the so-called Oort Cloud, a vast reservoir of comets theorized decades ago by a Dutch astronomer but never observed to date. This proposal is tough to test because the escaped matter would be currently far beyond the range of today's spaceprobes, only a handful of which—the *Pioneer* and *Voyager* missions—have now passed the orbit of Pluto. What specifically lies at those outer realms of the Solar System—or beyond even that, some thousand times the distance to Pluto where interstellar space begins—is frankly unknown.

In building a viable model of our planetary origins, it's imperative to touch base periodically with reality—to use the traditional, and sometimes sobering, actuality of genuine data. In recent years, astronomers have managed to acquire increasing evidence that all young stars apparently do experience a highly active evolutionary stage known as the T-Tauri phase. (The twentieth, or "Tth," star in the constellation Taurus is the premier example.) It is at this stage, when the stars have only recently fired up their nuclear fusion, that their brightness is especially great, their winds extremely intense. Nebular particles in the form of opposing jets travel outward along rotational axes, as noted previously in the Stellar Epoch, carrying with them much mass and momentum from their spinning source. Such bipolar jets are now observed for many T-Tauri stars, none of which is more than a few million years old.

Although we have no direct information about our embryonic Sun itself, observations of strong stellar winds emanating from young stars elsewhere suggest that much of the nebular gas left over among the planetesimals of our system could have been blown away into interstellar space—and with it some of the system's "missing" angular momentum. The youthful Sun must once have had narrow, fast-moving jets that reached light-years across. What turned them on is a contentious issue, but what turned them off is simply unknown, other than suppositions that perhaps the infalling matter simply ran out. In any case, the adolescent Sun's solar wind, energetic flaring, and radiation pressure would have aggressively, and not so gently, blown away much of the loose nebular disk within a few million years of its formation, even before hydrogen fusion commenced.

Despite some lingering controversy as to how to solve this momentum quandary, nearly all astrophysicists agree that some version of the condensation model is correct. The details, however, not yet worked out and still under debate, form the essence of a most challenging problem now being addressed at the frontiers of science at several leading observatories around the world. To repeat: The big picture of our Solar System's origin is in place. At issue are the specifics; we want to know, as best we can, exactly how our home in space did materialize.

∞

Diversity of physical conditions in the earliest years of the Solar System is probably responsible for the large contrast in content and structure between the Terrestrial and the Jovian Planets. This is where the adjective "condensation"—as in the condensation model—takes on its true meaning. Again we return to consider thermodynamics operating in a gravitational field.

As the primitive Solar System contracted under the influence of gravity, it heated, spun up, and flattened. Well before even the initial protoplanetary eddies began taking shape, the rising warmth broke apart the dust gains into simple molecules, and they in turn split into simpler atoms. Since the density, and hence the collision rate, were surely greater close to the protosun, matter there would have become hotter than in the outlying portions of the youthful system. A temperature gradient naturally developed—one that surely caused the original dust in the inner regions to incinerate, but not necessarily so in the outer regions where the grains would have probably remained mostly intact. While the gas temperature was several thousand degrees Celsius near the core of the contracting system, it would have been well below the freezing point of water some ten astronomical units away, out where Saturn now resides.

Such a gas cannot continue to heat indefinitely, lest the region blow up. Like any hot gas, the primitive Solar System must have released some of its newly gained energy. So, even as the protosun continued heating upon contraction, the outer regions of the primordial system cooled. As a result, heavy elements several astronomical units from the protosun began reversing their fate by crystallizing from their hotter gas phase to their cooler solid phase. (Again, the same process occurs today on Earth, though on a smaller scale, as raindrops, snowflakes, and hailstones condense from moist, cooling air.)

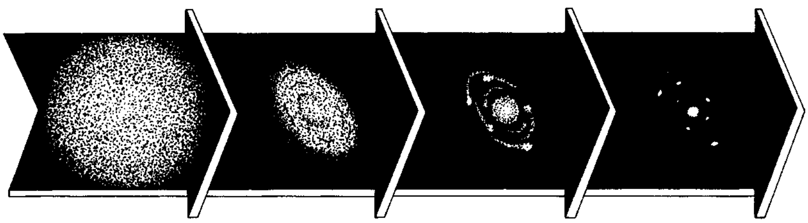
With the passage of time, the temperature decreased at all locations, except at the very core where the Sun, not yet a genuine star, was still forming. Everywhere beyond the protosun, atoms slowed while returning to their low-energy states, after which some of them collided and stuck to form molecules, which in turn clustered to form dust grains once more. This is the accretion process described above, but now one that would have operated more selectively, creating a compositional gradient in the early Solar System.

We might think it amusing that although plenty of interstellar dust grains uniformly peppered the area early on, Nature saw fit to destroy them only to rebuild them again later. However, a critical change had occurred in the meantime. Initially, the interstellar gas was evenly sprinkled with an array of all sorts of dust grains. When the dust later reformed, the mixture was much different, for the condensation of solid dust from hot gas depends on the temperature. The act of contractive heating had served to sterilize much of the region, thus setting the stage for a Solar System highly diversified in planetary composition.

In the outer, colder regions of the nascent planetary system, beyond several astronomical units from the Sun, where temperatures would have been a few hundred degrees Celsius or less, reasonably abundant heavy elements such as carbon, nitrogen, and oxygen combined with the most abundant element, hydrogen, to form some well-known simple chemicals, including ice crystals of water, ammonia, and methane—to be sure, the primary constituents of the Jovian atmospheres seen today. (Helium is an inert element and does not combine chemically with other atoms.) The ancestral fragments destined to become the Jovian Planets were fashioned under rather cold conditions by gravitational instabilities much like those noted earlier for the formation of galaxies and stars.

Accretion was no doubt at work way out there as well. Microscopic icy grains orbiting throughout the outer nebular disk gradually collided and stuck together, fabricating increasingly larger aggregates of ice in much the same way that fluffy snowflakes can be compressed into snowballs. Together with leftover hydrogen and helium atoms trapped by the strong gravitational pull of these huge protoplanets, gassy and icy compounds are now known to constitute the bulk of the Jovian Planets. Had we been there to see it, the emergence of these massive planets probably resembled the formation of our Sun. None of them is quite massive enough, though, to kindle nuclear fusion, the hallmark of any star.

By contrast, in the inner, warmer regions of the young Solar System, the average temperature would have been about a thousand degrees Celsius at the time when condensation from gas to solid began. The environment there was simply too hot for ices to survive. Instead, many of the abundant heavier elements such as silicon, iron, magnesium, and aluminum would have combined with oxygen in order to make iron oxides, crusty silicates, and a variety of other rocky minerals. Planetesimals in the inner system were therefore rocky in nature, as were the protoplanets and planets they ultimately formed.



. . . the act of contractive heating served to sterilize much of the region, thus setting the stage for a Solar System highly diversified in planetary composition.

These orbiting rocky grains gradually coalesced into objects of pebble size, boulder size, kilometer size, and larger—another bottom-up scenario. The bigger they grew, the quicker gravity helped them coalesce, sweeping more and more matter from the surrounding regions of the flattened nebular disk and eventually fabricating planet-sized objects. That the Terrestrial Planets are smaller than their Jovian counterparts owes mostly to the relative lack of material in the inner disk; their difference in chemical composition owes mostly to the formative system's temperature gradient. The very abundant light elements of hydrogen and helium, as well as many other gases that failed to condense into solids, would have surely escaped from these small protoplanetary objects. Their temperature was too high, and their gravity too low, to prevent light gases from escaping the inner planets. What little hydrogen and helium did manage to stick around was probably blown away by the wind and radiation of the newly ignited Sun. What remained, so say the theoretical models, were a few rocky planets, each cool, hostile, and largely devoid of an atmosphere.

Why the myriad rocks of the asteroid belt between Mars and Jupiter failed to coalesce into a planet remains a mystery. Perhaps one did exist, after which it blew up, the puzzle then being for what reason. If a planet did once exist there, it must have been a small one; the total asteroid belt now contains only a tenth the mass of the Moon or a thousandth the mass of Earth. More likely, these old, uneroded rocks (most less than a few meters across) never did manage to clump together to form a planet, given the incessant tug of Jupiter's gravitational tides that caused the asteroids to collide destructively rather than assemble constructively. That destructive process is probably still underway today, preventing the development of any protoplanet in the belt. If so, then the asteroids are the sole surviving witnesses and must hold primal clues to the grand event that did occur here nearly five billion years ago.

∞

Current consensus envisions the genesis of a planetary system as a natural, perhaps frequent, outgrowth of the birth of a star. The condensation model can generally account for each of the seven properties noted earlier that characterize our Solar System today. But exactly how those atoms of gas and grains of dust managed to coalesce into the present planets and moons remains one of the great riddles of modern science. Frankly, we may never know the precise details, given the role played by chance. Not that chance dominated events in the early Solar System, for scientific determinism was also functioning. But chance is an essential factor in all evolutionary events, and the birth and development of our planetary system were not exceptions.

Chance, contingency, catastrophe, and the like are real factors in modern science, especially complexity science. They are increasingly acknowledged these days, if not well understood. All such stochastic events function as agents of imperfection, helping us appreciate, at least in general terms, the many deviations from Nature's otherwise well-ordered scheme of things. Variations from "perfect" categories abound among both the living and the nonliving. But, aside from the quantum realm of elementary particles, chance does not usually rule our world or our lives—probably not ever. Contrary to popular opinion, chance is not in control of Nature's multifarious phenomena. Not everything we see around us is an accident; far from it.

Astrophysicists build grand theoretical models for the origins of systems such as the planets described here, guided by the known principles of deterministic science. Yet, and admittedly, embedded in that determinism, which derives from the laws and formulae of predictable, mechanistic physics, is an ingredient of chance. The two go hand in hand like a dance: chance flirts with necessity, randomness with determinism. To be sure, it is from this interchange that novelty and creativity arise in Nature, thereby yielding unique forms and novel structures. Much as for the galaxies and stars of earlier epochs along the arrow of time, chance did indeed play some role in the origin of our planetary system—and continues now with the ongoing evolution of Earth itself. Chance and necessity are twin features that weave in and out of the cosmic-evolutionary tapestry. We shall return to them repeatedly.

Regarding the early Solar System, chance would have manifested itself not only as planetesimals collided randomly to build protoplanets in the formative stages but also as the established planets and moons were later bombarded with incoming, leftover debris well after those formative stages. The effects of such collisions can still be seen today in many parts of the Solar System, not least on the marred surface of our own Moon. And it's those chancy, catastrophic effects of collision that likely explain many of the planetary anomalies noted earlier, among them: Two large bodies probably collided and merged to form Venus, yielding its abnormally slow, in fact retrograde, spin. Uranus was likely affected by a grazing encounter with a massive body (possibly as large as Earth), tilting that planet almost completely over on its side. Also, many of the moons would have been nearly destroyed in random collisional events, accounting for some of their truly bizarre surface terrains. While it is impossible to test these assertions directly—for these events are long over and done with—we can reasonably suppose that some of the decidedly odd aspects of our Solar System, especially those deviating from its well-ordered architecture, can be explained by means of untraceable, chancy events during and after its formative stages. Sadly, we shall never know with any certainty the specific incidents triggered by chance, for most of this collisional activity must have occurred in the first billion years or so of our system's history. These are not only times long gone but also events almost hopelessly confused with more recent evolutionary changes that have continued to sculpt our home in space.

Despite the catastrophic implications they carry, comets and asteroids are not entirely "vermin of the skies," as these interplanetary wanderers were often termed only a half-century ago. Ironically, for they caused much lasting damage, these celestial vagabonds also provide perhaps the most useful information available about our home's origins. Much of the debris that has survived to this day has preserved within it specks of solid and gaseous matter from eons past. Comets, in particular, may harbor material of a pristine, unevolved nature and thus have much to teach us about our local beginnings. Traveling in highly elliptical orbits and sometimes closely encountering the Sun, comets are often seen as faint and fuzzy patches of light, their tails beautifully sweeping across the nighttime sky. Yet, these "dirty snowballs" are more than spectacular sights of sublimating ice and rock, indeed more than mere inspiration for poets. Each time a comet appears in the heavens, think of it as a harbinger of news from an outer, ancestral reservoir—even if it is the merely postulated and yet unsighted Oort Cloud about a light-year away. Each comet that graces our skies brings us a little more of the story of our Solar System's origins.

Meteorites that land on Earth's surface are especially telling about the original state of matter in the solar neighborhood. The smaller ones, less than a few centimeters across, are mostly liberated swarms of cometary debris; the larger ones are more likely stray asteroids from the famous belt between Mars and Jupiter. Either type of meteorite extends our understanding by bringing us novel information about extraterrestrial matter close to home yet well beyond Earth. The blackest, most primitive ones, known as carbonaceous chondrites, which are rich in organic molecules, are of special interest to the cosmic evolutionist since their clues may harbor data not only about planets and moons but also perhaps about life itself. We are left wondering if rocks on museum shelves, or even some of those in our backyards, might contain vital signs about the origins of home and selves.

We shall return in the Chemical Epoch to debate the relevance of these alien intruders in our Solar System, especially the meteoritic chunks that survived the plunge to Earth's surface. These are the extraterrestrial bodies of rock and ice that have altered the geology of our planet throughout the course of history and have impacted—literally—the biology of planet Earth. And like mutations among life-forms, not all impacts have been negative. During periods of bombardment, the motor of evolution often accelerated, granting the potential for diversity, death, and rebirth among Nature's complex systems. The effects of impacting bodies on the evolution of life have rich implications for the interface between astronomy and biology—the crux of the newly emerging interdisciplinary subject of astrobiology.

Earth's Moon may well be the foremost example of a cosmic catastrophe in the local realm, the intrusion of chance into an otherwise straightforward condensation of gas and dust underway five billion years ago. The origin of our Moon is surprisingly uncertain, although its age implies an event contemporaneous with the formation of Earth itself. The four-and-a-half-billion-year-old age of the oldest Moon rocks, collected in the lunar highlands by manned American and robotic Russian missions of the 1970s, agrees closely with the age of all meteorites, as well as with the implied ages of both the Sun and Earth. Apparently, the Moon partook of a grand event in our cosmic neighborhood that spawned our Solar System eons ago. However, none of the several theories advanced in the twentieth century to account for the Moon has proved entirely satisfactory.

One theory holds that the Moon condensed as a separate object near Earth, in much the same way as did our planet. The two objects would have then essentially formed a binary-planet system, each revolving about the other. Though favored by many planetologists, this idea suffers from a major flaw: the Moon has only about half the density and a different composition than Earth's, making it hard to understand how both could have originated from the same protoplanetary blob.

A second possibility has it that the Moon originated far from Earth and was later captured by it. In this way, the density and composition of the two objects need not be similar, for the Moon and Earth would have presumably formed in unrelated regions of the early Solar System. However, the concern here is that the Moon's capture could not have been easy; it might well have been impossible. Why? Because the mass of our Moon relative to that of Earth (about one percent) is larger than for any other moon of any other planet. It's not that our Moon is the largest natural satellite in the Solar System, but it is unusually big compared to its parent planet. Mathematical modeling implies that it would have been highly unlikely for Earth's gravity to have captured the Moon in just the right way to avoid either its crashing into Earth or skipping off into deep space.

A third idea maintains that the Moon materialized out of Earth itself. The Pacific Basin has often been mentioned as the place from which protolunar matter might have been torn, the result of centrifugal forces on a young, molten, and rapidly rotating Earth. As absurd as this idea may seem, the early findings of the *Apollo* program seemed to favor it. Both the lunar composition and density were found to mimic those of Earth's mantle, that region just below the crust. However, recent, more exacting studies of our Moon's makeup show significant dissimilarities to Earth's underbelly. What's more, there remains the fundamental mystery of how Earth could possibly have ejected into a stable orbit an object as large as the Moon.

Clearly, none of these theories is compelling. Each suffers from a major flaw or two. Yet, it would seem that one of them, or some version of them, must be correct. In fact, astronomers now favor a hybrid model combining the best features of each of the above ideas. The most popular model today postulates a vast, ancient collision between a young, molten Earth and a large, Mars-sized object. Impacts were undoubtedly common in the early Solar System, although one of this magnitude would have been nearly catastrophic; perhaps it was more of a glancing blow than a head-on collision. Matter that dislodged from our planet, as well as parts of the impacting object itself, presumably then aggregated to form the Moon. Computer simulations do show that, for a collision at an oblique angle, debris having largely the composition of Earth's mantle could have been ejected into a stable orbit nearly halfway to where the Moon resides today. It probably would have happened quickly, with the far-flung material reassembling into a single clump within a few weeks and forming a spherical rock resembling today's Moon within a year. As Earth aged by billions of years and its spin slowed largely due to lunar tides, the Moon then naturally receded to its current distance, thereby preserving the total angular momentum of the Earth-Moon duo, as required by physical law. This dynamical scenario is supported by tidal layers deposited in rare rocks formed along prehistoric shorelines that show Earth's day to have lasted eighteen hours and its year to have equaled nearly five hundred days roughly a billion years ago. Furthermore, laser measurements (that time light's travel back and forth) of the Moon's distance have proved that it continues to move away from Earth at a rate of several centimeters per year. Such a huge wallop might also explain not only how Earth spins so rapidly on its axis but also how the tilt of its axis (which causes our seasons) is so large, at twenty-three arc degrees relative to the plane of its orbit. Not everyone agrees with this hypothesis, however, as other modeling suggests that for typical collisions, the whole of Earth most likely would have shattered into pieces if hit with such a large object.

One of humankind's most ancient questions seems still up for grabs. The origin of the closest celestial body to us, indeed one that hovers above us while reflecting brilliant splendor most evenings, is not well understood. Perhaps the formation of our Moon was the product of circumstances so rare that we shall never be able to unravel the details of its birth. Some researchers have been so perplexed by the origin of Earth's Moon that they have felt forced, in desperation, to argue that the Moon cannot possibly exist!

Absolute and relative ages of all structured objects found in Nature provide vital clues to help calibrate the cosmic-evolutionary story. Not only do we seek to know the specific ages of many ordered systems and key events in universal history; we especially want to be sure that those ages fare well in a temporal sequence along the arrow of time. Recall that in the prologue, we were concerned that stars better be younger than the Universe, for the obvious reason that nothing can be older than its parent. Likewise, here in the Planetary Epoch, we want to be sure that Earth is younger than the oldest stars, indeed contemporaneous with the Sun, lest something be amiss in the planetary-system models sketched here. In turn, in the next Chemical Epoch, we shall need to be alert that fossilized life is in all cases younger than the rocky Earth (unless life arrived intact from beyond our planet). These kinds of "sanity checks" are useful periodically while developing the scenario of cosmic evolution. A reasonable and consistent sequence of objects' relative ages is just as important as their absolute ages.

Take Earth for example. Estimates of Earth's age have increased dramatically over the past couple of centuries. The most widely quoted of the early estimates is attributed to Anglican Archbishop James Ussher, who in the mid-eighteenth century used the Bible to reason that Earth had been created in 4004 B.C.—October 23, by most accounts! Other researchers at the time, however, preferring to ponder Earth itself and not just beliefs about it, were convinced that our planet must be a good deal older than this six-thousand-year-old value implied by Scripture.

Although few scholars claimed precision, by two hundred years ago they mostly agreed with the pioneering French naturalist Georges Buffon, who argued that Earth is at least a hundred thousand years old. Reportedly, even heretically for the time, he maintained in his unpublished diary that Earth likely formed several million years ago. Gradually during the early nineteenth century, and led especially by the Englander Charles Lyell (who heavily influenced Darwin's thinking), most geologists came to accept Earth's age as spanning millions of years, yet still fully a thousand times younger than what we know today. As for a specific value for the duration of our planet, many were then content to see, in the words of the father of modern geology, the Scottish farmer James Hutton, "no vestige of a beginning, no prospect of an end."

Lord Kelvin was an exception. By the mid-nineteenth century, this British physicist had become familiar with the new subject of thermodynamics (the science of changing heat, or energy), and he used it to try to calculate Earth's age. Arguing that any gravitationally contracting object cools at a certain rate, he reasoned that our planet would have been molten hot sometime between tens of millions and hundreds of millions of years in the past. However, as noted in the prologue, even these longer durations fell short of our planet's true age. While Kelvin did the calculation correctly, he was unaware of a most important phenomenon—radioactivity.

Not until the early twentieth century did French scientists, mainly the Curies, isolate radium from pitchblend, thereby learning how that heavy element decays into several lightweight elements. Such decays, naturally occurring throughout Earth, provide an additional source of energy on our planet and thus extend Kelvin's inferred value for Earth's age. Soon thereafter, a pioneering British atomic physicist, Ernest Rutherford, championed the idea of using radioactive elements to directly date Earth materials. Finally, the true age of Earth could be found—well, almost.

Technically, many heavy nuclei, such as those of uranium, thorium, and plutonium, are inherently unstable. If left alone, they gradually break up into lighter nuclei, in the process emitting some elementary particles and releasing energy. This change happens spontaneously, without any external influence. (The energy released by the disintegration of radioactive elements drives the process of nuclear fission, in either controlled nuclear reactors or uncontrolled atomic bombs.) The decay from "parent" nuclei to more stable, "daughter" offspring is not immediate, however; rather, it happens at a characteristic pace—measured by the nuclei's half-life. Half-lives vary greatly, as measured in the laboratory; for example, half a sample of plutonium will decay in a few million years, while half a sample of uranium needs seven hundred million years. Thus, if we can measure the amount of unstable parent nuclei of a given element remaining today in, say, a rock, and also measure the amount of its stable decay product, then we can specify the time during which the decay occurred. This method is widely used by geologists and gives age estimates with an accuracy of a few percent, but there is a caveat.

The radioactive-dating technique rests on the assumption that the rock has remained solid while the radioactive matter decayed. If the rock melts, there is no particular reason to expect the daughter nuclei to remain in the same locations their parents had occupied, and the whole method fails. Therefore, radioactive dating measures the time elapsed since the rock in question last solidified. In many cases, this will be a lower limit, given that most rocks underwent some heating in their past.

Throughout the first half of the twentieth century, radioactive methods gave Earth ages variously in the range of one to three billion years. As our understanding of nuclear physics advanced during the second half of the century, so was progressively older rock found on our planet. Today, the oldest rocks are found in Greenland and Labrador, dated to nearly four billion years and proving that our planet is at least that old. Furthermore, since Earth is highly differentiated—its heaviest elements are mainly at the planet's core, in contrast to its lightweight elements being at or near the crust—it must have been molten at some earlier time, lest the heavies not have been able to sink into the core. A combination of thermodynamic tests of rock cooling rates, radioactive dates of stray meteorites and the lunar highlands, and theoretical studies of the Sun's evolutionary state all converge on an age for planet Earth of close to four-and-a-half billion years.

This episode in the changing estimates of Earth's age is a good example of how the scientific method, though sometimes affected by the subjective whims and biases of individual researchers, does eventually yield a definite sense of objectivity. Over the course of time, groups of scientists checking, confirming, and refining experimental tests will neutralize the subjectivism of single workers. Often a few years of intensely focused research is enough to bring much objectivity to bear on any given problem, although some particularly tricky issues—such as Earth's advanced age in this example, as much as Earth's orbiting the Sun in Galileo's day—were swamped for generations by cultural and institutional biases fostered by tradition, religion, and even politics.

Today, with an open mind and a readiness to revise our models to reflect new theoretical ideas and better experimental tests, scientists maintain that Nature yields a certain measure of objectivity through revealed facts, thus granting us a progressively better "approximation of reality." It is in this sense that science claims to make progress, both in quantitative terms of a fuller, more accurate knowledge and in qualitative terms of a richer understanding of what this knowledge means.

∞

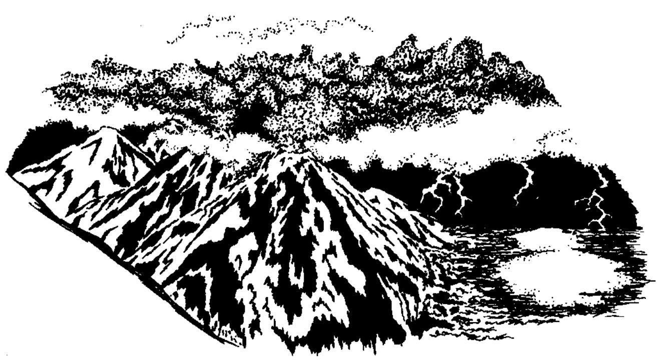
The most troubling aspect of Earth's origin is our inability to probe the geological record for the first half-billion years of our planet's history. Studies of Earth itself are surprisingly useless. Evidence from this critical time period, which would ordinarily provide clues to the youthful environment in which our planet was born, is missing, having been literally melted, eroded, and chipped away long ago. What we do know is that, in nearly every respect, primordial Earth and its global environment of several billion years ago must have differed substantially from the world we now inhabit.

Drawing a mental picture, we can surmise that shortly after Earth formed, it was hot, oceanless, free of oxygen, and pelted with all sorts of energy from within and without. Solar ultraviolet radiation, fierce thunder and lightning, radioactive rocks, and violent volcanoes all energized our young planet. Intense meteorite barrages at this time, known as the period of heavy bombardment, must have caused our early planet to resemble a hell on Earth for its first half-billion years or so. We need look no further than the heavily scarred and anciently cratered Moon for ample proof that Earth was, in fact, belted frequently by comets and asteroids. The whole globe must have melted right down to its center since, when Earth rings like a bell during earthquakes, geologists infer a complete differentiation of our planet's interior, from core to surface. Not long after Earth's formative stage, the dense, iron-nickel metals must have sunk to the center while the lightweight, granite-silicate rock rose toward the surface. Gradually, the restless Earth cooled, cracked, exhaled steam, and secreted an ocean and atmosphere.

Earth's original atmosphere almost certainly contained all of the most abundant elements—hydrogen, helium, nitrogen, oxygen, neon, carbon—as well as a long list of trace elements. These gases mimicked those of the interstellar cloud from which our Solar System formed. This primary atmosphere did not likely stick around very long, however. Earth's surface was much hotter during its first billion years than it is today, and many of the atmospheric gases present then must have escaped to outer space. Gravity just could not hold back the early hot gases.

The relative scarcity of several noble gases—those that are inert and unable to react with other chemicals—such as neon, argon, krypton, and xenon, is the best evidence that Earth failed to retain its original atmosphere. If our primordial atmosphere were still here, even if modified by later evolutionary events, those inert gases should be present in quantities comparable to those in the Sun where they do in fact exist. Apparently, the heavy bombardment, high surface heat, and fierce solar winds were too much for the small young planets to bear. None of the Terrestrial Planets likely retained their original gaseous atmospheres left over from the primitive solar nebula.

Despite the depletion of Earth's initial atmosphere, one obviously surrounds our planet today. We wouldn't be here if it didn't. Hence, the air we breathe must be a secondary atmosphere acquired by our planet at a later date. What's more, these secondary gases in turn evolved, owing to the presence of plants as explained later in the Biological Epoch, to become the air we do now breathe—so perhaps we should correctly call the current gases in which we are bathed a "tertiary atmosphere."



. . . primordial Earth . . . differed substantially from the world we now inhabit.

For the same reason that ice cubes congeal from the outside in, the surface of the gradually cooling primordial Earth would have been the first part of the molten planet to solidify into rock. Intense heat trapped below the crust had to get out somehow. The result was surely volcanoes, geysers, quakes, and a variety of other geological events that literally blew off steam and pent-up heat through cracks in the surface. "Outgassing" of this sort happens even today, though at only a few locations on Earth and rather infrequently at that. But several billion years ago, this type of geological activity was surely more widespread and frequent. Scrutiny of modern volcanoes shows that lots of steamy water vapor, carbon dioxide, and nitrogen would have then undoubtedly emerged, along with vast quantities of ash and dust. Smaller amounts of hydrogen, oxygen, carbon, and other gases doubtless accompanied these early planetary eruptions. Calculations imply that over the course of Earth's history, enough gas was exhaled through fissures from Earth's interior to create much of our current atmosphere. The rest presumably came from comets and meteorites that could have salted the young Earth with large quantities of matter, including prebiotic molecules. Even today, some forty thousand tons of extraterrestrial matter fall to Earth each year, almost all of it burning in the air or splashing in the ocean.

The origin of our present atmosphere is therefore a combination of terrestrial outgassing and interplanetary assault (further changed by later biological events). In truth, Earth's atmosphere is perhaps still adjusting, as present-day volcanoes occasionally sputter gas and heat amid incoming debris arriving from space. Today's atmosphere was not, however, derived directly from a mixture of interstellar gases. The composition of Earth's secondary atmosphere thus differs considerably from the average cosmic abundance of the elements. By contrast, Jupiter and the other Jovian Planets have atmospheres rich in hydrogen, helium, and many light gases. These planets are large enough to have retained their primitive atmospheres, which *were* formed directly from interstellar matter.

Since the atmosphere and ocean of planet Earth are so closely linked, they almost certainly originated, at least partly, from the same sources—our planet's interior and interplanetary bolides. As regards the ocean, geologists argue that as the surface cooled sufficiently, the first store of liquid water pooled as water vapor condensed. After all, steam is the main component of volcanically vented matter, and the hydrated rocks (mostly silicates, with water trapped inside) that make up Earth's mantle today store several times more water than in all the seas combined. But outgassing from within may not be the whole story since our planet's water has chemical (isotopic) subtleties implying that some of it may have come from beyond.

Debate swirls among geologists concerning the rate and timing of ocean formation. We are unsure whether Earth's mantle outgassed the global seas all at once early in our planet's history—known among geologists as the "big-burp" theory. Or perhaps the seas took some time to form, having secreted from Earth's interior in a series of volcanic events that occurred more gradually. A minority of researchers argue that some (perhaps even all) of the waters of Earth could have resulted from a rain of water-rich comets and meteorites that collided with our planet in great numbers during its first billion years. But here, too, there is a chemical anomaly: the makeup of water on Earth does not well match that trapped in interplanetary bodies. Three comets that recently bypassed Earth—Halley in 1986, Hyakutake in 1996, and Hale-Bopp in 1997—all emitted radiation that revealed a heavy-water (deuterium) content twice that in Earth's ocean.

Most likely—as with many other aspects of the cosmic-evolutionary story that are neither black nor white, neither solely this nor cleanly that—Earth's large bodies of sea water emerged from both within and without, and then, as the rate of outgassing and bombardment declined, a global recycling system began to operate. Water locked in rocks was expelled back into the ocean whenever the rocks were heated, such as those near volcanoes or suboceanic faults and ridges. To be sure, most of today's seawater is thought to have been recycled many times through the worldwide system of oceanic ridges, perhaps as frequently as every ten million years. Recently, water has been directly observed emanating from certain underwater vents. Whether this is "juvenile" water still originating directly from Earth's mantle and incrementally adding to the world's supply or merely existing seawater cycling through the vents is not yet known.

Earth's ocean and atmosphere gradually stabilized. As activity on the early planet subsided, the atmosphere cooled, enabling gravity to retard its further escape into space. Nitrogen partly reacted with other gases and partly remained free in the atmosphere, where it now forms the largest fraction of our air. Gaseous water vapor changed into liquid water, which further rained down into Earth's oceans. And discharged carbon dioxide reacted with silicate rocks in the presence of water to form limestone. Whatever pure oxygen gas existed on primitive Earth would have quickly vanished by reacting either with hydrogen to make more water or with surface minerals to form oxides such as rust and sand now found throughout the crust of our planet. Breathable oxygen and the protective ozone layer arose only much later, after plants had blossomed across the face of our planet.

Shaded by Earth's secondary atmosphere, some of its chemicals would have further interacted with one another. No coercion by outside influences was needed for the airy gases to collide, stick, and react, thus forming slightly more complex gases of ammonia and methane. Chemists verify these reactions almost every day in industrial and academic laboratories. And theorists well understand the electromagnetic forces among electrons that persuade these simple atmospheric atoms to combine spontaneously, thereby concocting stable gas molecules.

With time, the molecular products of these spontaneous reactions became the reagents of additional chemical reactions. These additional reactions, however, were not spontaneous. Laboratory experiments prove that the simple molecules of ammonia, methane, and water vapor require some energy in order to combine further. This energy is, in some sense, a catalyst that helps produce even bigger molecules. Actually, it's more than just a catalyst. The application of energy fashions a near miracle: it synthesizes molecules a good deal more complex than those likely to form by chance in a collection of free atoms and simple molecules. As we shall see in the Chemical Epoch, the molecules produced are among the very building blocks of life.

∞

The rocky surface, or lithosphere, of Earth—that part of our planet most familiar to us—is interesting not only because we live on it, but also aesthetically given its sheer beauty and scientifically given its occasional activity. Earth is geologically alive. Its interior boils and its surface erupts. Change still affects our home in space.

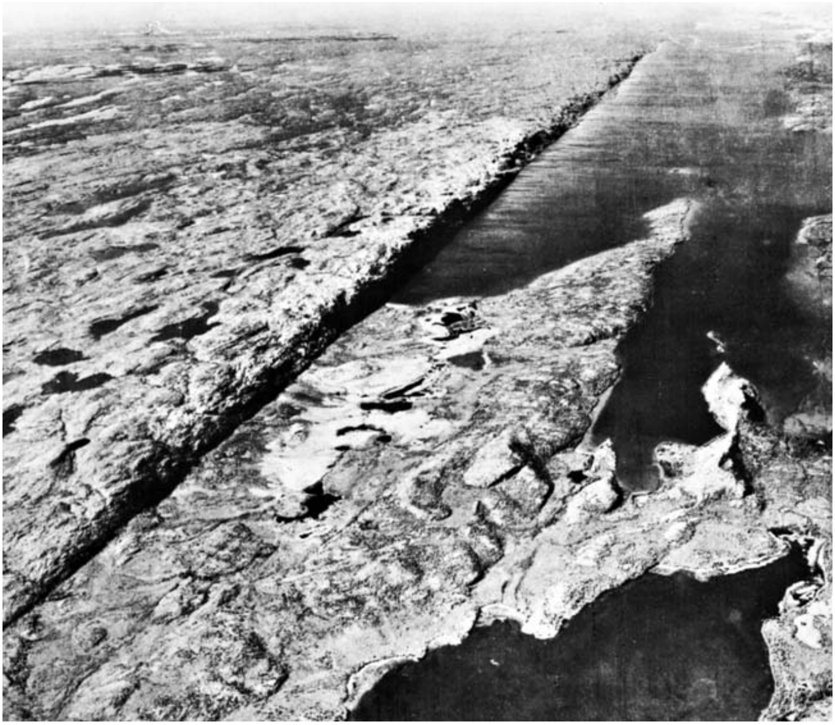
Volcanoes are especially clear indicators of lithospheric activity—sites where molten rock and hot ash upwell through fissures or cracks in the surface. Despite their scarcity nowadays—rare, yet sensational, enough to make them the lead item on news broadcasts—volcanoes are examples of present-day activity, as are the earthquakes that regularly occur when Earth's crust suddenly dislodges under great stress. Tracing events backwards, geological studies of rocks, lava, and related substrata imply that surface activity must have been much more frequent, and probably a good deal more violent, long ago.

Huge slabs of Canadian rock lie side by side yet differ in age by well more than a billion years; their juxtaposition alleges some large-scale jostling of surface rock. Cliffs along the coast of Scotland display evidence for horizontally layered rock much as expected from millennia of sedimentary growth on the ocean floor, yet above it lies nearly vertically layered rock; apparently some type of surface upheaval thrust one part of this rocky cliff onto the other. Other examples abound for past activity at or near the surface of Earth, despite erosion by wind and water that has clearly wiped away much of the evidence for truly ancient events.

A map of the most active current sites shows them unevenly spread across our planet. Indeed, those sites outline well-defined lines of activity, or faults of weakness, where crustal rocks dislodge (as in earthquakes) or mantle rocks upwell (as in volcanoes). Only a few decades ago did it become clear that these faults delineate about a dozen gigantic "plates" or slabs of Earth's surface. Called plates because of their size, their scale, and their shapes' resemblance to upside-down dishes, the huge horizontal extent of these slabs (typically thousands of kilometers across) usually dwarfs their vertical thickness (roughly a hundred kilometers deep).

Notably, one of the most prominent faults separates the North American Plate from the Eurasian Plate, continuing on down between the South American Plate and the African Plate. Throughout the midst of the entire Atlantic Ocean and extending without interruption all the way from Scandinavia in the north to the latitude of Cape Horn in the south, is the Mid-Atlantic Ridge—a thin, almost continuously submerged fault rising above sea level only at the subcontinent of Iceland. This ridge is the best known and most impressive of many smaller cracks and underwater trenches discovered in recent years by miniature submarines deployed from oceanographic ships.

Startling when first realized but now pretty well understood, the huge plates are slowing sliding around—literally drifting on the surface of our planet, hence the popular term "continental drift." In doing so,



Evidence of plate tectonics.

Surface activity on Earth is exemplified by this enormous crustal fracture dividing two huge slabs of rock in Hudson Bay. The rock to the right of the crack has been dated by radioactive methods to be nearly three billion years old, while that to the left is hardly more than one billion years old. Two vast segments of Canadian rock with greatly differing ages, yet laying side by side, clearly imply that large-scale jostling of surface rock must have occurred, probably caused by nameless continents adrift. *Source: Canadian Government.*

the plate movements have created the surface mountains, the oceanic trenches, and many other large-scale features strewn across the face of planet Earth. The plate motions have in fact shaped the continents themselves, hence the official term "plate tectonics"—tectonics deriving from the word "architecture," meaning to build or construct, in this case mountains and oceans via the movement of plates.

Not that the plates are moving fast, by any means. The ground beneath our feet rightly feels like terra firma. Though the plates are still now drifting billions of years after the hardening of Earth's crust, they do so today at extremely slow rates. Typical velocities of the plates amount to less than a few centimeters per year, or roughly the speed at which our fingernails grow. Still, even at this sluggish pace, each plate has had plenty of time to move large distances during Earth's history. For example, a drift rate of only two centimeters per year could separate two continents by some four thousand kilometers over the course of two hundred million years—which happens to be, not coincidentally, the width and age of the Atlantic Ocean. That's surely a long time by human standards, but actually only about five percent of the age of Earth.

As the plates drift around, collisions are routine, with whole continents bumping into one another. Yet, unlike two automobiles that collide and then stop, the surface plates, being so massive, have enormous momentum. Not easily stopped, they just keep crunching into one another for thousands, if not millions, of years. The Himalayan mountain range, of which Mount Everest is a member, is a notable example of two plates colliding; here, the Indian subcontinent is thrusting northward into the Eurasian landmass at an anomalously fast rate of five centimeters per year, not only lifting the towering Himalayas but also buttressing the Tibetan Plateau even at the present time. This collision between continents has been underway for nearly the past fifty million years. The Alps are another good case of continental collision, as northern Italy, which is mostly part of the African Plate, plows northward into central Europe; together, the two plates are causing vast amounts of wreckage, slowing and currently reshaping the ruggedly young Alps. By contrast, the rounded and heavily eroded Appalachians in the United States exemplify tectonic events that culminated some three hundred million years ago.

Not all plates experience head-on collisions; many slide or sheer past one another. The most famous active region in North America, the San Andreas Fault in California, serves to illustrate this kind of less severe plate interaction. This fault causes much earthquake activity as the Pacific and North American plates rub past each other. The two plates are not quite moving in the same direction and not quite at the same speed. Though they surely are in contact, like a poorly oiled machine their motions are not steady and smooth. Rather, jerky and sudden movements result each time the pressure to drift overwhelms the friction to stay put.

Several strands of evidence support the idea of plate tectonics on Earth, the most obvious provided by geography. Just looking at a map of Earth's major continents, we readily see that they seem to fit together like pieces of a puzzle, especially the facing coastlines of Africa and South America. The easternmost Brazilian coast meshes nicely with the Ivory Coast of Africa and its armpit near Nigeria. Farther north, West Africa fits nicely into the oceanic cavity in the Caribbean Sea and the Gulf of Mexico. Farther south, southwest Africa matches up with the southern coast of Brazil, Uruguay, and Argentina. Admittedly, the fit is not as good in the northern hemisphere, given the "debris" in the North Atlantic, including Iceland, Greenland, and the British lsles. Yet the western coast of Europe nestles nicely with the Mid-Atlantic and New England coastlines of North America. Furthermore, rock formations all along both sides of the Atlantic are very much the same.

Apparently, a gargantuan landmass must have dominated our planet sometime in the past. Knowing the plate vectors (both direction and magnitude), geologists can trace their movements back into the past. (It's not too different from measuring the speeds of galaxy recession and then mentally reversing them to model the early Universe, or reconstructing an automobile accident by studying the scattered wreckage.) And what the geologists have found is evidence for a single ancestral supercontinent, called Pangaea, which means "all lands" and must have contained almost all of the dry land on Earth. To the north was Laurasia and to the south Gondwana, each separated by a V-shaped body of water called the Sea of Tethys. The rest of the planet must have been entirely covered with water.

Exactly when did Pangaea exist? The current locations of the continents, along with their estimated drift rates, imply that it must have been the major land feature on Earth approximately two hundred million years ago. Dinosaurs, which were then the dominant form of life, could have sauntered from Russia to Texas via Boston without getting their feet wet. About twenty million years thereafter, Gondwana and Laurasia began dividing for reasons unknown, probably near the present-day Gulf of Mexico. About thirty million years after that—namely, about one hundred and fifty million years ago—Gondwana itself broke into various pieces: South America, Africa, and Australia, as we now know them. Shortly thereafter, Laurasia split, perhaps more violently, thereby producing North America, Europe, and the smaller fragments now in the North Atlantic Ocean.

The existence of such an ancient supercontinent and its subsequent breakup explain several heretofore peculiar findings. For example, when climbers first reached the summit of Everest fifty years ago, they discovered fossils of fish and old clamshells. Only plate tectonics can seemingly justify how marine fossils could get to nearly the highest point on Earth. Whatever caused Pangaea to come apart also set its continental fragments in motion. In particular, as India began its slow trek northward across the Sea of Tethys, fossils of marine life deposited at the bottom of that ancient sea were apparently pushed up alongside parts of the Eurasian landmass to form the Himalayas.

Plate tectonics have slowly reshaped the surface of our planet. In some cases, the seafloor has been literally thrust to the top of the world. In other cases, huge underwater mountain ranges have slowly emerged. In still others, entire subcontinents have apparently submerged. As humankind begins its in-depth exploration of the nearby planets, it will be interesting to see if those alien worlds were also reshaped so heavily by physical events that went far beyond mere surface erosion. So far, we think not, as noted later in this Planetary Epoch.

To give credit where credit is due, much of this geographic puzzle was solved early in the twentieth century by the German meteorologist Alfred Wegener, but few believed him. The idea that huge slabs of rocky crust could be literally drifting around on the surface of Earth was preposterous until the 1960s. Views changed rapidly, however, when several additional lines of evidence suddenly became available. In the mid-1960s, academic geologists who championed continental drift couldn't possibly have gotten tenure, whereas by the early 1970s those who didn't, couldn't.

The subject of paleontology—the study of the fossilized remains of dead organisms (the Greek prefix "paleo" meaning old)—further bolsters the idea of plate tectonics. For example, fossils of *Mesosaurus*—a freshwater reptile that lived nearly three hundred million years ago and has been extinct for the past two hundred million years—have been uncovered at only two locations on Earth. One place is a small part of what is now the northeast Brazilian coast, while the other is on the west coast of Africa near present-day Ghana. These two places are precisely where the continents would have dovetailed on the ancestral supercontinent of Pangaea. If Africa and South America were always separated by the great expanse of the Atlantic Ocean, these reptilian creatures could have hardly survived the long swim between coasts. Even if they did, the chances are slim that they would have departed and landed at exactly those parts of the two continents that geographically mesh. A more reasonable judgment is that Africa and South America were once joined and that this reptile lived in a small region in the midst of Gondwana. Likewise, fossils of identical snails turn up in New England and Scandinavia and those of marsupials on the west coast of South America and eastern Australia—in both cases, among other examples, animals that could not have swum across such wide and open oceans.

A third piece of evidence for plate tectonics comes from oceanography—the study of ocean dynamics and history, as well as its physical and chemical behavior. Many of the active sites submerged beneath the ocean form a giant system of undersea cracks—the Mid-Atlantic Ridge being the most prominent example, stretching for fifteen thousand kilometers right down the middle of today's Atlantic Ocean. Not only have underwater cameras lowered from surface ships mapped this vast fault, but robot submarines have also managed to retrieve samples of ocean floor at various places on both sides of its submerged mountain range. And what they found is that matter on the ocean floor closest to the underwater ridge is relatively young whereas that farther away is noticeably older.

These observations support the notion that hot matter upwells from cracks all along the Mid-Atlantic Ridge. In this way, some of the plates are literally pushed apart. The North and South American plates are moving generally westward while the Eurasian and African plates drift eastward—which is exactly the trend implied by the geographical fit noted above: the plates on either side of the Atlantic Ocean have presumably been drifting apart for the past two hundred million years. To be sure, oceanographers have never found any part of the Atlantic seafloor to be older than this date. Submerged rocks close to the east coast of the Americas and the west coast of Europe and Africa are radioactively dated to be nearly two hundred million years old; those closest to the Mid-Atlantic Ridge are only a few million years old.

Finally, a fourth bit of evidence also favors the idea of plate tectonics, although here the data are not as robust, their implications not as clear. This evidence is supplied by paleomagnetism—the study of ancient magnetism. Everyday experience tells us that iron is magnetic; in fact, any metal containing even small amounts of iron ore is usually magnetized. However, when iron is heated to temperatures around seven hundred Celsius, it loses its magnetic properties as individual atoms jostle freely (which is why magnetic thermometers often fall off the side of a roaring wood stove). Hot basalt—the dark, dense stuff of volcanoes—impregnated with traces of iron and upwelling from cracks in the oceanic ridges, is thus not magnetic. As the basalt cools, magnetism sets in as each iron atom effectively responds to Earth's magnetic field like a compass needle. When the basalt solidifies to form hard rock shortly thereafter, it fixes the orientation of the embedded iron, since the iron atoms align themselves with the orientation of Earth's field *at the time of cooling*. Accordingly, the ocean-floor matter has preserved within it a history of Earth's magnetism.

Consider the Mid-Atlantic Ridge again. When samples of ocean floor are examined from sites close to the ridge, the iron deposits are aligned with today's north-south field. This is the "young" basalt that upwelled in recent times. Samples retrieved far from the ridge, however, often have their iron deposits misaligned at odd angles relative to the usual north-south field. This is the "older" basalt that upwelled in earlier times and was then subject to the twisting, turning, and drifting of nearby plates ever since. Working backward, oceanographers use the embedded iron to infer the past positions of the plates, as well as the North and South Poles. Much like reorienting giant, mobile pieces of a very large jigsaw puzzle, they realign parts of the ocean floor to a common north-south direction and thereby infer the approximate drifts of the plates during the past two hundred million years. When this is done, these paleomagnetic data also tend to support the existence of a single ancestral supercontinent on our planet.

Paleomagnetic studies have advanced some remarkable additional findings in recent decades, all the while spawning yet another rich interdisciplinary interaction—this one between geology and astronomy, and possibly biology too. Foremost among these findings is the discovery that Earth's North and South Poles have flip-flopped many times over the years. Surprising as it sounds, the north magnetic pole, now in the Arctic region, has occasionally been in the Antarctic region, while the south magnetic pole has sometimes resided in what we now call terrestrial north. Such back and forth reversals have occurred irregularly hundreds of times during the past two hundred million years and possibly thousands of times before that. Although not a frequent happening on human timescales, magnetic reversals have nonetheless occurred at least a dozen times in only the past ten million years. The north magnetic pole has been in the Arctic region for the past three-quarters of a million years—currently about a thousand kilometers west of true geographic north, on Canada's Prince of Wales Island. During the twentieth century alone, Earth's global magnetic field decreased several percent, perhaps presaging another dramatic reversal in the upcoming millennium.

These facts we know from seafloor samples dredged up from near the Mid-Atlantic Ridge. Since oceanographers date the seafloor according to its distance to the east and west of the ridge, samples tested for magnetism mark when Earth's field changed its orientation. The seafloor resembles a giant tape recorder, with matter flowing from the central ridge to either side and laying down a record that is progressively older farther away. Hot matter first upwells through cracks in the ridge, after which it quickly cools, solidifies, and aligns trapped iron ore, preserving a record of Earth's magnetism at that time. The matter then spreads out from the ridge while pushing apart the plates. In this way, the underwater data reveal a history of the Atlantic Ocean's formation, growth, and alternating magnetism.

What could have caused such global reversals of Earth's magnetism? While researchers don't yet know for sure, apparently something occasionally upsets the steady spin of our planet's liquid iron-nickel core. The metal core is the probable source of our planet's magnetism—a geodynamo whose spin of an electrically conducting metal induces a magnetic field much like any electromagnet. Speculation has it that the culprit might be an inherent, convective instability in the usual rotation rate of the core, or perhaps even something more dramatic, such as collisions of Earth with cosmic objects like comets or asteroids. Such catastrophic events could upset the core's spin by sloshing around the liquid trapped there, thus playing havoc with the normal field and setting the stage for it to change.

Whatever their cause, magnetic reversals are probably not instantaneous events. Some time is probably needed for magnetism to weaken gradually and finally disappear. Likewise, some additional time—perhaps on the order of hundreds, even thousands, of years—might be needed for the magnetic field to become reestablished. If so, then each time a reversal occurs, the magnetosphere (including the Van Allen belts that shield Earth from radiation far above its surface) disappears, perhaps for a rather long time by any living standard. Without this protective "umbrella" to deflect or trap the charged cosmic-ray particles incident on our planet, biological systems *might* be harmed.

Circumstantial evidence argues for life having been threatened in this way in the past. Fossil records of ancient life-forms show that, every so often, once-abundant plants and animals suddenly became extinct. No one knows why they perished so rapidly, and in the Biological Epoch we shall note some alternative ideas. Even the dinosaurs, which reigned supreme on our planet about a hundred million years ago, seem to have vanished within a relatively short period of time. Asteroid impacts might well have triggered mass extinctions of living systems, but their ramifications likely included, in addition to climate and sea level changes, magnetic-field reversals. Some analyses of seafloor matter, though controversial, do imply a connection between the extinction of certain species of life and a reversal of Earth's magnetism. Magnetic turnabouts are possibly death sentences for *some* species, yet therein lies a positive note as well: other fossils found within the seafloor samples also show that other, wholly novel species newly emerged.

No one knows for sure what effect magnetic-field reversals have had (or might yet have) on the surface of our planet. But a general consensus seems at hand: While Earth's magnetosphere is collapsed, the influx of high-energy particles (mostly from the Sun) would likely destroy the ozone layer and increase the amount of radioactive atoms in our atmosphere. These atoms would then be absorbed by plants and in turn eaten by animals, including humans. Although higher levels of radiation affecting plants and animals are unlikely to kill any life directly (for the atmosphere does also provide robust protection), cancerous cells would surely increase, and the normal course of biological evolution would be disrupted. Some basic biological molecules, including genes, would suffer damage, causing reproductive errors, or mutations, from generation to generation in some living systems.

Contrary to popular belief, however, not all mutations are bad. Some are beneficial, enabling life-forms to adapt better to changing Earthly environments—in short, to evolve. Mutations act as a motor of evolution and without them life would not have complexified so rapidly, as we shall see in the Biological Epoch. And when the magnetosphere temporarily shuts down, that motor apparently accelerates.

We shall also come to realize later along the arrow of time that whatever species dominates life at any given period does so largely because that species enjoys a nearly optimum relationship with its environment. It's best suited for its natural surroundings, enabling it to survive, and sometimes thrive, quite nicely. An analogy might be a well-focused image illuminated through a slide projector. A mutation (or a slight change of the slide's focus) is likely to harm the dominant species (or the quality of the projected image). By contrast, lesser species (or slides out of focus) might profit from a period of increased mutations by finding themselves bettered. At any one time, it is the dominant species that is most likely to change for the worse. Humans are now the dominant species on Earth.

The Atlantic seafloor is still growing, pushing apart the North American and Eurasian plates, as well as the South American and African plates. Submarine expeditions prove that the Mid-Atlantic Ridge remains active, with hot basalt now rising through fissures and cracks encircling the globe for thousands of kilometers like seams on a giant baseball. Geologists assume that other plates on Earth's surface are also pushed around by matter upwelling from similar, yet only partially explored oceanic ridges. Oceanographers are now retrieving and studying rock samples from the bottom of the Pacific Ocean and Caribbean Sea, among other places where seafloor activity is exposed. During the upcoming decades, we should gain a better understanding of all the major underwater cracks through which hot matter oozes from Earth's interior.

We thus confront a central question of modern planetology: What drives the plates, however slowly they drift across our planet? In other words, what's the source of the seafloor spreading? Answers all point to the mechanism of convection—basically, the upward transfer of subterranean heat via circulation—in this case, giant cyclical patterns of molten matter in Earth's mantle. The conditions are perfect for this kind of physical process, which amateur geologist James Hutton foresaw more than two hundred years ago; yet he found few adherents until a generation ago partly because he was such a poor communicator of science.

Here's what we know today: The ocean floor is covered with a layer of sediment—dirt, sand, and dead marine organisms that have fallen through the seawater for millions of years. Below that sediment lies about fifty kilometers of mostly granite, the low-density rock composing the continents. Deeper still lies the mantle made of warm, partially molten basalt. And beneath that is Earth's core, whose temperature tops five thousand degrees Celsius—nearly equaling that of the Sun's surface! This is indeed the perfect setting for convection: warm matter underlying cool matter. The warm basalt wants to rise, just like hot air in our atmosphere or smoke from a fire. It does so through any cracks and faults in the largely granitic crust. Every so often, such a fissure opens in the midst of a continental landmass, producing a spectacular volcano such as Mount Etna in Sicily or a geyser like Old Faithful at Yellowstone National Park in Wyoming. However, most of the major crustal faults now known are submerged below water, probably for the simple reason that nearly three-quarters of Earth's surface is covered by water. The Mid-Atlantic Ridge is the premier example, yet even this long and impressive suboceanic system is only part of a sixty-thousand-kilometer-long submerged mountain chain that formed in progressive stages by the breakup of previously intact continents.

Not all the warm basalt in the mantle can squeeze through the cracks and fissures. Some gets pushed back down. In this way, ascending and descending matter form huge circulation patterns within the upper mantle. These convective cycles often extend as much as a thousand kilometers below the crust, or a fifth of the way toward Earth's center. They are also very sluggish, flowing at speeds of only a few centimeters per year and taking perhaps millions of years to complete one cycle. The basalt slowly circulating below a crack is, after all, a semisolid resembling warm asphalt and moving much less smoothly than gas in the air or liquid in a pot of boiling water. The result is that the upwelling matter eventually spreads out horizontally in the upper mantle, thereby exerting enormous drag (frictional force) on the thin surface plates and causing them to slide or drift across the face of our planet.

Much of Earth, then, acts like an engine—a heat engine that obeys thermodynamic laws. And the key to any such thermal device is a temperature gradient. Accordingly, energy naturally flows from hot to cold, that is, from inside to outside of any planet. Convective cycling allows Earth to cool and lose its heat to space in the fastest possible manner. Since rocks conduct heat so poorly—feel the cool bottom of a flagstone on a hot, sunny day—conduction of energy is not an option. In other words, tectonics is the most efficient way to remove internal heat by recycling rock from the interior to the surface over and over. In that way, Nature quite naturally seeks to destroy all gradients, namely, to attain the lowest possible energy state represented by a true equilibrium, at which time—surely billions of years into the future—planetary evolution will stop.

Where are the plates headed? Assuming that warm basalt continues to rise through the fissures and thereby power the plates on their steadfast journeys, can we predict the positions of the giant landmasses on Earth in the relatively near future? By extrapolating the plate's current vectors—again, both magnitude and direction of their drifts—geologists can reasonably anticipate where the continents will be in the years ahead. For example, in fifty million years, the Atlantic Ocean will have widened nearly another thousand kilometers. By contrast, the Pacific Ocean will have shrunk considerably; there being no large landmass on the Pacific Plate, this oceanic plate will presumably continue to be overridden by other continental plates such as the westward-moving South American Plate. Australia, which is actually part of the Indian Plate and moving fastest of all the plates at nearly eight centimeters per year, will continue its northerly motion toward the Eurasian landmass, destined for a massive collision and truly renewed landscape in South Asia several million years hence. India itself will continue to thrust northward, as now, building the Himalayas to possibly greater heights. The Mediterranean Sea is doomed, given the African Plate's northerly motion, at the same time guaranteeing great skiing in the Alps for millions of years (provided that the climate doesn't change much). And southern California, as part of the Pacific Plate, will be torn away from the North American Plate, with Los Angeles becoming a suburb of San Francisco in about twenty million years, before being dumped into the Aleutian Trench some fifty million years in the future.

∞

We stand to learn much about Earth's evolution by comparing its properties to those of some of the other Terrestrial Planets. The interdisciplinary subject of comparative planetology has recently come into its own—the study of the broad and contrasting properties among the diverse worlds in our solar neighborhood. What makes Earth so different from the other planets? How is it that our home alone has blue skies, liquid water, and a gentle climate? And why is Earth the only world in the Solar System (as far as we know) that is an abode for life?

Consider the land, for instance. The high-standing continents on Earth, set slightly above the sea, owe their existence to the long history of plate-tectonic activity—activity probably absent on any other nearby planet. We take the land for granted, of course, for that's where humans live. We even tend to focus on the land areas in those magnificent photos taken by astronauts of a distant Earth in space. But most of Earth is covered by water; a typical view from our planet's surface would show exclusively water in all directions—which may be why the astronauts call Earth, when looking back from orbit, "the big blue marble." In fact, the conditions needed to form the continents on Earth may be unmatched anywhere else in the Solar System. Those conditions led to active tectonics, the sign of a geologically lively planet, and life on the dry land is the beneficiary of it.

By contrast, Venus has recently expired, geologically speaking. Its closer proximity to the Sun might have shut down plate tectonics early, assuming it ever really got going. A difference of a few tens of millions of kilometers in the two planets' distances from the Sun might have been enough to turn Earth's nearest relative into a remote cousin. Extra solar heating seems to have thoroughly dehydrated Venus—its surface temperature today is a torrid five hundred degrees Celsius, nearly enough to melt lead—making its crust and upper mantle too dry and especially too buoyant to sink back down into its interior. The great upwelling of lava, outgassing of chemicals, and jostling of crust that accompany tectonics on Earth probably never much affected the Venusian surface. Robotic radar observations of this totally enshrouded planet, especially those made by the *Magellan* spacecraft in the 1990s, show little evidence for recent faulting, ridges, or volcanism (though ancient, now dormant volcanoes are visible). Surface features comparable to Earth's continents, such as the highland landmasses Ishtar and Aphrodite Terra, have apparently not wandered around much, if at all. Parched Venus seems to have been inactive for the past half-billion years or so (as dry rock is stronger than water-bearing rock) and to have encased itself in a single, thick shell—yet a shell that probably preserves a record of its last attempt at crustal deformation, such as that which produced its most dramatic topographic feature, the Maxwell Montes mountain chain, which exceeds the height of Mount Everest on Earth. Planetologists surmise that Venus, being slightly smaller, is aging more quickly than Earth, and as such its recent past may portend our future. Ironically, the volcanic surface of Venus, repaved as recently as seven hundred million years ago, before the planet went dormant, most likely resembles a young Earth that began to solidify, and therefore it might also tell us something about our planet's distant past.

As for Mars, the red planet has been geologically dead for a long time. Its store of internal heat ran down billions of years ago, shutting off all surface activity except at a few volcanic sites. The problem here is size; Mars is a good deal smaller than Earth. Most people regard Mars as comparable to Earth in size and scale, but it actually has ten times less mass. Venus is more properly labeled "Earth's sister planet" than Mars. Consequently, Mars never did heat up enough to melt its whole interior, to generate global magnetism, or to drive much (if any) plate tectonics. It seems to have been a one-plate planet since the end of the heavy bombardment period nearly four billion years ago. Mars's topography has probably been locked (and maybe frozen) in place for more than three billion years—which is why some of its fixed lava sites that were previously active, such as Olympus Mons and the Tharsis rise, are so much more extensive (several thousand kilometers across) than volcanoes on Earth. Hence, the Martian surface might also inform us about early planetary evolution of Earth—before our plates began moving—a time domain for which firm knowledge of our planet is sadly lacking.

Another good example of comparative planetology is provided by atmospheric gases—again of Venus, Earth, and Mars. Although almost certainly endowed at birth with similar amounts of hydrogen, carbon, and oxygen, each of these Terrestrial Planets has evolved differently. Their differences derive largely from their varying masses and distances from the Sun. When it comes to real estate value—terrestrial or extraterrestrial—the bottom line is much the same: size, location, and timing.

Of these three neighbors, our inward sister Venus receives the most solar energy, in fact roughly twice as much as Earth. Although liquid water is nowhere to be found on this planetary hothouse today, early in its history, when the Sun shone less brightly (about two-thirds of its present luminosity four billion years ago), Venus might have had widespread oceans, lakes, and rivers. As the Sun slowly increased its output in the normal course of its stellar evolution, the planet gradually heated and its water boiled off. In the meantime, Venus's volcanoes continued to vent much carbon dioxide into its atmosphere. And without the water to change carbon into rocky carbonates such as chalk, limestone, or coral (as is the case on Earth), the carbon dioxide gas levels on Venus rose unchecked—in short, most of Venus's carbon stayed in its atmosphere. The result was a "runaway" greenhouse effect, allowing solar energy to penetrate the thickening atmosphere yet blocking some of its outgoing infrared radiation, all the while making the surface of Venus too hot to support even primitive life.

By contrast, Earth is far enough from the Sun to have retained its liquid water. As water vapor rises in our atmosphere, it cools, condenses into droplets, forms clouds, and rains back down—the "water cycle." Earth is furthermore able to recycle its carbon through plate tectonics, an action probably untenable on Venus. Even today, carbon dioxide outgases from Earth's volcanoes, such as those along the Cascade Range in Oregon, but it weathers on land and dissolves in the sea, forming carbonic acid that eventually reacts with oceanic rocks to help form a limestone crust that, in turn, releases carbon dioxide yet again some tens to hundreds of thousands of years later—the "carbon cycle"—all of which checks the buildup of this greenhouse gas. The prolonged presence of water enabled the evolution of marine organisms, which then, as now, served as an effective means to further remove carbon dioxide from the air by making shells and skeletons, which later fall to the seafloor and compress into yet more rock—the most famous such geological feature being England's White Cliffs of Dover. Long ago, an atmospheric steady state—a chemical and thermal balance of sorts—was apparently reached: volcanism regularly vents carbon dioxide to the atmosphere, whereupon it's trapped in plants and rocks. A small percentage of carbon dioxide gas in our air does manage to drive a weak greenhouse effect, thereby raising our average surface temperature above the freezing point of water. Our climate is thus more moderate than Venus's, although humans are beginning to tinker with the delicate balance. We are industrially polluting our air as well as deforesting the land, in the process causing both the carbon dioxide content and the global temperature to rise; these are measured facts.

Mars, too, probably once had a moderate, wet climate, with liquid water on its surface. Ample photographic evidence virtually proves that substantial amounts of water flowed through several large channels and smaller tributaries, perhaps even inundating a third of the planet in huge lakes and Martian seas. The robot spacecraft *Spirit* and *Opportunity* that landed on Mars in 2004 confirm the idea that the landscape was once likely flooded with shallow seas of salt water. But with only one-tenth the mass of Earth, Mars had trouble holding onto its original atmosphere. And given that its tectonics never really got going (nor is it likely they ever will), Mars couldn't generate much of another atmosphere. The result, despite the high percentage of CO_2 gas, was slight greenhouse warming at best. Unable to "hold a lid" on its water, much of it dissipated to space. Water that didn't escape early on is now completely frozen at the poles and in permafrost, as Mars seems in the grip of a perpetual ice age.

Surprisingly, some of the Jovian moons might also grant added insight into early changes in the planetary evolution of Earth. For example, Jupiter's moon Callisto has a thick icy shell deeply pitted with impact craters that date back some four billion years to the burgeoning days of the Solar System. Since it has no source of internal energy, nor is it close enough to Jupiter to be affected (cracked or heated) by tides, Callisto has not been repaved with fresh, upwelling matter. That makes this scarred and battered object the oldest known surface anywhere, and as such it might tell us something about conditions shortly after the Solar System formed. By contrast, another of Jupiter's famous Galilean moons, Io, orbits so closely to the planet as to incur huge tidal forces that cause unceasing volcanism that wipes clean its surface and thus any clues about its past.

The contrast between Earth's early atmosphere and that on Titan today is also instructive. The largest moon of Saturn (in fact, bigger than Mercury and Pluto) is rich in methane and nitrogen gas as well as in several carbon-based compounds. Under the action of sunlight, these gases undergo a complex series of chemical reactions, producing a hazy, hydrocarbon smog. Perhaps most notably, these chemical reactions and the organic matter they yield are thought to resemble those produced in Earth's atmosphere billions of years ago, before the advent of living things and oxygen-rich air. Titan seems to be a chemical "factory" that might provide a wealth of information about the vital prebiological steps that led to life on our planet long ago. This is one task for the premier planetary mission recently en route: The multi-billion-dollar *Cassini* mission left Earth in 1997 and arrived at Saturn in 2004. While orbiting in and amongst Saturn's moons, the *Cassini* mother-craft dispatched a small probe called *Huygens* into Titan's atmosphere, seeking to unlock secrets of its—and perhaps our—past. Though the experiment is still underway, the early results show a pale-orange landscape interspersed with icy valleys laden with hydrocarbon sludge.

Despite our inability to explore much of our own planet's early history, we are in the ironic position of being able to study better the earliest phases of other, alien worlds. Some of the planets and moons have become virtual fossils, or relics, telling us things about our origins that our own planet cannot. As we acquire more data regarding the vast range of their physical and chemical properties, we gradually gain a better understanding of whatever did happen here approximately five billion years ago. This is the way the scientific method most commonly operates: groping unsurely and probing ever so slightly into uncharted territory—both real interplanetary turf as well as theoretical ground—as planetologists make incremental yet real progress in our quest to better approximate the reality that was once Earth's primal history.

∞

The benefits of comparative planetology can also sometimes be overdone—hyped beyond its ability to deliver. A few decades ago, the Solar System was judged a simple place: commonalities among bodies prevailed—and great promises, too, for us to learn more about global warming on Earth by deciphering the chemistry and climate of Venus or to better forecast weather on Earth by monitoring and modeling Jupiter's great red spot or generally to know more about Earth by looking to other planets for answers. Yet, although some common features do pervade the many planets and moons of our Solar System, spacecraft exploration of the past decade has shown, if anything, the great diversity of objects in our celestial neighborhood. Our expectations have lowered somewhat, as the focus has shifted from similarities to differences among alien venues: Mars is locked in a permanently frozen ice age; Venus is a hellhole wrapped in an immobile shell; the Jovian planets and their moons are a panoply of bizarre balls of matter that resemble neither Earth nor Moon. Much as with life-forms on our planet, we now ponder the staggering variety of material worlds beyond. Simple classification and great insight sought using comparative planetology has partly receded, given the rich detail of recent observations of the planets and their moons. What we need now are better statistics for whole families of planets, and that means finding more of them—which is exactly what has happened in recent years as astronomers have begun to discover scores of alien worlds beyond our Solar System.

The condensation model holds that the events that produced our home in space are not at all unique. Even considering the role of chance in the mix of changes that occurred long ago, astronomers can still clearly identify an underlying, deterministic sequence of events that led naturally to the birth of our Sun and its family of planets. Furthermore, we have no reason to expect that similar events, in general, would not have happened elsewhere. Many stars are expected to have planetary systems of some sort, and if we could find and map them then real statistics would bolster the subject of comparative planetology. Even if only one percent of all the stars in the Milky Way Galaxy have planetary systems, that still leaves billions of stars with planets. And each star, of course, would likely have more than a single planet orbiting about it.

Theory is one thing, yet observation quite another. Until the mid-1990s, astronomers had no reliable evidence of planets orbiting other stars. The scientific literature of the last many decades is littered with claims for "extrasolar planets" (called "exoplanets" for short) circling stars beyond our Sun. But none of those early claims could be confirmed and most were eventually retracted. Despite a strong desire and effort to prove that our planetary system is not alone in space, astronomers were unsure until recently about the plurality of planets elsewhere. All that quickly changed as new telescope technology and powerful computers made it feasible to detect, indirectly though unambiguously, the presence of planets around some nearby stars. At the start of the twenty-first century, we now know of more than a hundred such exoplanets orbiting stars beyond our own—that's more than ten times the number of planets in our own Solar System. And although the alien planets discovered to date seem to have properties quite different from those in our own system—none are even remotely Earth-like—they do provide examples of worldly systems against which to test our ideas about the origin and evolution of planets in general.

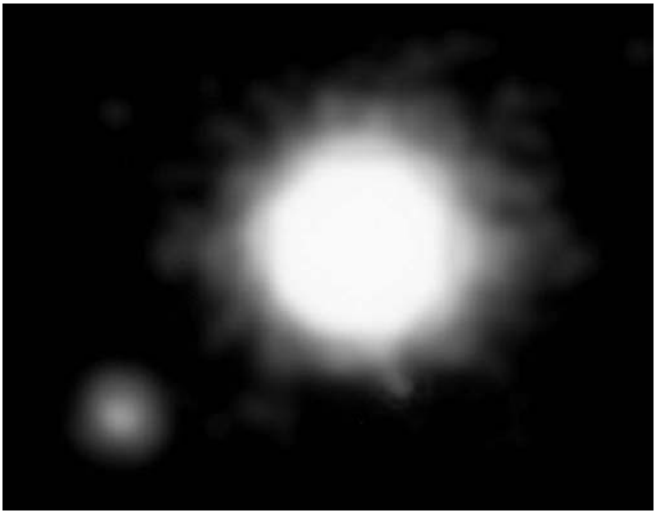
Few extrasolar planets have yet been imaged directly. Their presence to date is mostly based on inference. Astronomers have few photographs of them, and even the best ones are faint and fuzzy. Most planets orbiting other stars would appear too dim and too close to their parent stars for today's telescopes to resolve them. Instead, the techniques used to find distant planets are based on studies of light emitted from their parent star, not of reflected light from the planets themselves.

As a planet orbits a star, gravitationally pulling first one way and then the other, the central star tends to "wobble" slightly. The higher the planet's mass, or the lower the star's mass, the greater the induced motion. However, even this wobble cannot be seen directly in the movement of stars across the sky. Instead, the presence of the new planets is inferred, as are their masses, by careful observations with moderate-sized telescopes equipped to monitor the spectral shift in the light of a star while it moves back and forth along our line of sight. This is again the famous Doppler effect, much like that used to track the recession of the galaxies or to catch a speeder on the highway. Here, light emitted by a star in motion has its wavelength shortened or lengthened, owing to motion toward or away from us, thus betraying the presence of unseen objects orbiting about the star.

The first such star system to have had a planet found and confirmed in this way, 51 Pegasi, is a near-twin to our Sun yet forty light-years away and barely visible with the naked eye just outside the great square of the Pegasus constellation. Analysis of this star's radiation implied a planet having about half the mass of Jupiter and orbiting with a period of only a few days. That's an extremely short period, meaning, according to Newton's law of gravity, that this foreign planet must be very near its parent star, in fact well inside the equivalent orbit of Mercury, which takes eighty-eight days to orbit the Sun. Thus, the first such exoplanet discovered was odd to say the least and quite unlike anything in our Solar System—a surprisingly massive planet in a highly eccentric orbit and almost right on top of its parent star. Even more surprising, this trend has continued in the past few years as more planets were found around other stars: "hot Jupiters," as they are called, are massive planets orbiting in close proximity to their central stars. Inward orbital migrations, especially for the giant planets whose large tidal interactions might cause them to spiral toward their parent stars, might be a common dynamical feature in all planetary systems and may have already occurred in our own Solar System.

One of the most interesting extrasolar systems discovered to date is Upsilon Andromedae. Here, a triple-planet system orbits a single star much like our Sun, just forty-four light-years away. All three of the suspected planets have Jupiter-sized masses, and all three are well inside the equivalent orbit of "our" Jupiter. Clearly, this family of planets doesn't resemble our own Solar System much at all. But which is the "normal" system, them or us? Virtually all the recently found planets seem highly peculiar by our home standards, yet who has the right to claim our planetary system as the standard? Only about five percent of all stars surveyed to date show evidence for exoplanets and only nearby stars at that (within a few hundred light-years), so the planetary properties of the vast majority of stars remain a mystery.

These new planetary findings, however strange and unexpected, almost surely suffer from observational bias. The techniques used to detect extrasolar planets are most sensitive to massive planets, so it's not too odd that the earliest results favor the gas giants. Those techniques are not yet accurate enough to have found smaller, less massive planets—assuming the smaller, probably rocky planets do exist. What remains



Evidence of extrasolar planets.

Few direct photographs of planets beyond our Solar System currently exist. Most of those planets have only recently been found using indirect methods that track subtle gravitational tugs on their parent stars. In this remarkable image, taken in the heat-sensing infrared part of the spectrum by the Very Large Telescope in Chile, a faint planet (lower left) is seen nearby its larger parent star cataloged as 2M1207, which is a dim and distant brown dwarf some 200 light-years from Earth. All the extrasolar planets found to date, like this one having five times the mass of Jupiter, resemble our Jovian planets, and since most of them orbit close to their parent stars, they are unlikely to be hospitable for life as we know it. *Source: European Southern Observatory.*

puzzling is that so many of the new Jovian-sized planets are so close to their parent stars. Not inconceivably, some of them might eventually prove to be brown dwarfs, or “failed stars,” and not genuine planets at all. That is, some could be double-*star* systems, their current status as planetary systems misidentified. While the dividing line between a planet and a Sun-like star is about seventy Jupiter masses, that between a planet and a dwarf star may be as little as a dozen Jupiter masses—and some of the newly found objects are close to that latter mass value. Even so, the consensus in the astronomical community today is that not all the new objects are likely to be brown dwarfs. At least some of them must be bona fide planets. Our Solar System is not alone in space!

Jupiter-sized exoplanets are intriguing, but we naturally (and chauvinistically) wonder: Are other “Earths” out there? How unique is our home planet? Alas, that question remains as difficult as ever to answer. As noted previously in the Stellar Epoch, astronomers have not yet invented the equipment needed to inventory small, compact, dark bodies residing in even nearby space. To stress an annoying limitation, as ironic as it is: We can detect objects as large as stars, for they glow of their own accord, making themselves visible. And we can detect objects as small as atoms, largely by means of spectral radiation they emit and absorb. But we have a hard time detecting anything midway between these two extremes, unless they are very nearby, like asteroids. Faraway objects having sizes between stars and atoms go mostly unseen, and unless they tug on some other nearby object gravitationally they are virtually undetectable—which is the case so far, as Earth-sized planets are too small to cause their parent stars to wobble enough for us to see.

It does seem likely that small, rocky planets are absent from those several score stars where the massive exoplanets have been found to date. In our own Solar System, the presence of Jupiter in a nearly circular orbit well out from the Sun is judged a stabilizing influence. Jupiter helps to dynamically regulate the orbits of Earth and the other Terrestrial Planets; its gravitational tides tend to damp large-scale orbital eccentricities, causing planetary paths to circularize more readily. Jupiter also helps to protect the inner parts of our Solar System from huge rocks wending their way toward the Sun. This big, outer planet literally acts like a vacuum cleaner, using its ample gravity and large cross-section to sweep our planetary system relatively clean, and thereby prevent too many impacts from badly whacking the inner, Terrestrial Planets. By contrast, for the newly discovered extrasolar systems, having inwardly migrating Jupiter-sized planets plowing through their inner parts on elliptical orbits means that any small, Earth-like planets were likely destroyed in place or ejected from the system long ago.

The upshot is that astronomers are now finding clear, undeniable evidence for planetary systems beyond our own. That's the good news, for it does bolster the idea that planets form everywhere as natural by-products of star formation. But the new results are also unsettling, since they don't even remotely resemble our Solar System. Perhaps it was silly to think, even with the condensation scenario operating in many nooks and crannies of our Galaxy, that all such alien systems would look like ours—though that's what the theorists thought just a few years ago. Now, armed with real data, astronomers are rapidly changing preconceived attitudes, fine-tuning their models to match the real worlds.

Despite the avalanche of incoming data in this fast-breaking field, no one is about to abandon the intricate scenario that explains so well so many of the gross features of our own Solar System. The condensation model remains the most viable explanation for our home planet's origin. That said, some of the latest computer models suggest that planetary systems perhaps ought to look more like those implied by the new extrasolar data, with big Jupiter-sized planets in tight, eccentric orbits. Maybe our Solar System is the unusual case after all. This new field is data driven; only more observations will tell for sure.

We are left with the notion that, if Earth-like planets in stable orbits do form by condensing out of cooling gas and dust, then space should be teeming with them, just as it bristles with stars and galaxies beyond. But the feeling is an uneasy one, for we don't yet know for sure. Just how common or rare—or special—are Earth-like planets in the Universe?

∞

The current generation of planetologists has discovered more about the Solar System than in all of prior recorded history. Like earlier adventurers of Renaissance times, we are now living in another golden age of exploration—in this case, exploration of alien worlds and unearthly environments quite foreign to our own. Much like the pathfinders who ventured forth in the great sailing ships to map the New World on planet Earth several centuries ago, today's scientists vicariously crew robotic probes that trek around the Solar System. The effort is still a work-in-progress, but the results thus far have revolutionized our knowledge of both our present cosmic neighborhood and our planet's natural history.

Our Solar System harbors a vast array of material objects beyond the central Sun. Planets, moons, comets, and asteroids are all well-known, if not completely understood inhabitants of our minuscule niche in the suburbs of the Milky Way. The wide range of physical and chemical properties among the peculiar planets and their motley moons yields the impression that our Solar System is full of debris, or at least great diversity—the remains of a more violent, yet simultaneously more formative era in the history of our local interplanetary environment. Can we realistically expect to identify all the pieces of this celestial puzzle and thereby decipher the full mosaic of our planetary origins? The answer, we think, is yes.

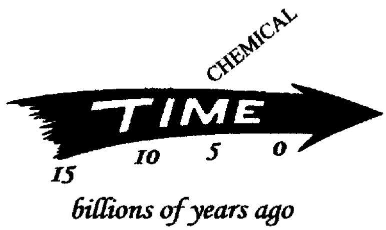
Each planet contributes knowledge that widens our appreciation for planetary evolution, much as diverse stars add to our understanding of the stellar life cycle. Most planets and moons are now in different stages of development, much as red giants and white dwarfs represent distinct phases of stellar evolution. The Jovian Planets are galactic fragments frozen in time, not massive enough to have become stars, yet too massive to have condensed rocky surfaces. To varying degrees, these gassy worlds, having originated mainly via gravitational instabilities, preserve the pristine properties of the early Solar System. By contrast, the less massive Terrestrial Planets, formed mostly via accretion, have evolved a great deal, cooling and crystallizing hard rocky surfaces while outgassing atmospheres and sometimes oceans. At least one of these small planets has spawned life.

Regard the Solar System, then, as not just a collection of planetary refuse. Every planet and its family of moons have something to tell us, something about their origin and evolution. Each time a new space probe reconnoiters a planet—and more robots are on their way right now—we learn a little more about bizarre landscapes, alien atmospheres, and comparative planetology. We learn how each planet fits into the overall architecture and general history of the grand Solar System, thereby helping us frame a planetary heritage of which humankind can be proud.

Planet Earth, in particular, is large enough to have remained warm inside and thus continues to experience some surface activity. Yet its outside has cooled enough to allow gravity to bind air and water to its surface. Earth thereby still evolves, though slowly and subtly. Our distance from the Sun and our atmospheric blanket combine to keep Earth's surface temperature suitable for water to remain liquefied—an apparently vital factor in any environment hospitable to life as we know it. To be sure, many of the key attributes of our planet—and they are key, if not special or unique, such as tectonic dynamism, liquid water, free oxygen, and life itself—depend on a planet's size, location, and timing in its planetary system. Without any anthropocentrism meant or implied, Earth seems to be the right body at the right place at the right time.

5. CHEMICAL EPOCH

Matter Plus Energy



NEARLY EVERYTHING ON EARTH is made of elements heavier than hydrogen and helium. We need not be clever chemists to realize that the air, land, and sea are all partially made of matter cooked in the hearts of stars. The heavies are an essential prerequisite for the continued evolution of complexity in the Universe. Although it sounds poetic to claim that much of everything around us—and within us—is "starstuff," it happens to be true.

The bluish oceans of liquid water consist partly of a heavy element, for water is after all not just two parts hydrogen but also one part oxygen. Only a fortuitous combination of temperature and density, unlike that on any other known planet or moon, allows large quantities of water, enough to cover nearly three-quarters of our planet's surface, to remain in the liquid phase. Tidally pulled this way and that, and occasionally evaporated into the air only to condense back down as rain, water is an integral part of our changing terrestrial environment.

The air we breathe and the land we walk are also rich in heavy elements. Composed mainly of nitrogen and oxygen, Earth's atmosphere is constantly yet unevenly subjected to solar heating and variable weather, causing billowy clouds to dominate our planet from afar. And the brown and gray tracts of mineral-rich soil and silica-rich rock, looking immutable from space and feeling stable beneath our feet, drift imperceptibly across our globe. Surprisingly, iron is the most abundant element in all of Earth, most of it having sunk to the core long ago.

Oceanographic change, meteorological change, geological change: all these changes are up and running in our home in the Universe. Chemical change functions no less, for eventually the heavy elements begin to interact, react, and complexify.

The most remarkable heavy-element assemblage on Earth is life. Plants, animals, fungi, and bacteria are widespread in our biosphere—on the land, in the sea, and throughout the air—though only within the past ten percent of our planet's history have some of them become bigger, mobile, and sentient. Of particular note and telling perspective, the heavy-element concoctions known as men and women have existed for well less than one-tenth of one percent of Earth's history.

To connect with previous epochs in time, not only our planet but also our bodies are peppered with the heavy elements fused in ancient supernovae, and all of them as well with the ancestral element hydrogen of the early Universe. The principal epochs of cosmic evolution are overlapping; our story is becoming more interdisciplinary. The old adage that life is mainly made of stardust is never more apt than while noting the celestial roots of our chemical makeup. Literally, many stars have died so that we might live.

Life everywhere now seems biologically adapted to planet Earth, but adaptation is a never-ending effort. Change is inevitable. Nothing is immutable, nothing at all. The climate alters. The Alps build. The Atlantic widens. Even the rock-solid aspects of our sturdy planet quake and drift, evolving over timescales immense compared to human life spans. What we can't see is tough to believe, but we are witness to so brief a time interval. Even the ten-thousand-year duration of human civilization is a mere wink in the grand spectacle of eternal change.

It's that time along the arrow of time (on Earth anyway) to survey the salient changes of the Chemical Epoch that originally led to life long ago. Here, the pace of change ramps up, with novel ordered systems coming forth in greater numbers, greater diversity, and greater spectacle. Additional changes, either natural or legislated, among today's life-forms could either yield smarter, perhaps even wiser creatures or someday conceivably render planet Earth uninhabitable. Our fate once more mixes chance with necessity, nothing being preordained in this business.

∞

Having encountered many of the astronomical and physical (or "astrophysical") concepts needed to appreciate the origin and evolution of matter, we now consider some of the biological and chemical (or "biochemical") ideas central to the origin and evolution of life. It is the synthesis of these two vibrant interdisciplines that create the essence of cosmic evolution.

We are moving across the threshold connecting matter and life. And although life itself is closer to us in space and time—indeed, we *are* life—that doesn't necessarily mean that we understand it any better than matter. The reason is that living systems are so much more complex than any inanimate objects; a potted plant is more complicated than the most splendid galaxy. Much as some missing links hamper our current knowledge of distant stars and galaxies, gaps also plague our understanding of the history of life right here on Earth.

Yet each day brings new discoveries, tests, and refinement of our modern ideas of chemical, biological, and cultural evolution. And with these advances comes greater objectivity, and progress too, in our search to know reality. In this, the Chemical Epoch, we explore the ways taken and means used for the building blocks of life eventually to become life.

Let's first ask, what is life? And immediately we are stumped. By contrast, physicists throughout the world, regardless of country or creed, agree on a definition of matter—anything that has mass and occupies space. Matter is among the basic stuff of the Universe, and we have a reasonably good idea how it (at least detectable, normal matter) behaves from quark to quasar. But biologists are hard-pressed to offer a clear, concise, standard definition of life. At issue, again, is life's complexity. Life is so intricate, it's hard to describe even though we ourselves are examples of it! Frankly, the biological community has been unable to reach a uniform consensus about life's true character.

Usually, biologists attempt to define life operationally by appealing to its practical properties. By noting a few of life's attributes, we can begin to know its important features, especially those that distinguish life from matter. We might, for example, suspect that living systems differ from nonliving systems because the *whole is greater than the sum of the parts of which it's made.* An individual cell dies when removed from a living organism of which it was a part, since the interactions of that cell with other parts of the whole living organism are vital to the cell's health. On the other hand, should a single cell be nourished in a comfortable laboratory environment having optimum temperature and density—a so-called culture medium—such a cell could once again flourish outside its original living organism.

At first glance, then, we might take the italicized property above to be a peculiarity of life. But on second thought, this property is not at all restricted to life, for it's also a property of matter. To see this, imagine removing a small part of a star normally fusing hydrogen into helium. The extracted chunk of matter would no longer release nuclear energy, since it would immediately disperse into space and grow cold. Yet if that chunk were surrounded by additional matter having an appropriate temperature and density, it would once again shine as brightly as before.

These statements are not meant to suggest that stars are somehow "living": quite the contrary. It is precisely because we can be sure that incredibly hot stars cannot possibly be alive that this comparison demonstrates how tough it is to define life. Thus we cannot claim that the "whole being greater than the sum of its parts" is a property solely of living systems. This property applies equally well to many objects that are not living, as in a watch, for instance, which is surely more than the sum of the gears and springs (or silicon chips and integrated circuits) of which it's made. A watch's structure is made of atoms, but its function tells the time!

Biologists often claim that the *ability to heal itself* is a peculiar property of a living system. A shallow cut on a finger, for example, usually heals quickly and the system goes on living. On the other hand, the aforementioned star from which a small chunk of matter was extracted would also eventually "heal" itself. The star would adjust a bit, eventually attaining a new balance between the inward pull of gravity and the outward pressure of heat. Having resumed its original spherical shape, the star would then go about its business of shining as a perfectly normal, though slightly smaller star.

We might say that living systems have a special property that allows them to *react to unforeseen circumstances.* But a star wouldn't expect to have a small part hypothetically removed, yet it would react quite adequately to this unexpected occurrence. Stars can react, and adapt, to new states too.

The *ability to reproduce* is clearly a special property of living systems. Still, we could imagine a contracting protostar which, because of faster and faster rotation, divides into two separate protostars. In this way, angular momentum is sometimes judged an agent of replication, or at least subdivision. Admittedly, this example probably occurs rarely, yet it has undoubtedly happened many times in the billions upon billions of years since the start of the Universe. Some of the binary stars in our Milky Way Galaxy may well have been formed in just this way. A better example of "replication among the stars" might be the process of sequential star formation proffered toward the end of the Stellar Epoch, whereby the concussive deaths of some stars naturally lead to the births of others. Furthermore, mules don't reproduce and neither do sterile men, so perhaps reproduction is not such a definitive, unique quality of life.

Surely, some property must be associated with life and only life. Bioscientists often raise the possibility that living systems can *learn from experience.* Most living organisms do have a memory of sorts. Yet some nonliving systems can also remember, and even learn from experience, such as chess-playing computers. When a well-programmed computer makes a mistake, it doesn't forget it. These so-called neural networks can store mistakes in their hardware memory, never to be made again under the same circumstances. Accordingly, few humans can beat our best computers at chess and no one can beat them at blitz-chess (when the timescale for moves is much shortened). So some of our more advanced machines, which are still merely clusters of matter, can seemingly learn from experience, much like living systems.

Finally, life is often operationally defined as having an entire *hierarchy of functions.* Much of the activity of living systems is controlled by chemical hormones; hormones in turn are controlled by secreting organs called glands; glands by brain cells, and so on. Such hierarchies characterize all living systems from simple amoebas through advanced humans. Similarly, though, we can regard nonliving matter as being controlled by a hierarchy of functions: the motion of the Moon is dictated by Earth; Earth's motion in turn is directed by the Sun; the Sun by the Galaxy, and so on through the galaxy superclusters. Many material systems have governing hierarchies that resemble those of living systems.

The point worth stressing is that we cannot easily specify any property applicable to life and only life. Apparently, under some circumstances, common properties of life can also apply to matter. In short, there seems to be no clear dividing line between what's alive and what's not—no obvious distinction between matter and life.

All this back-and-forth discussion reinforces the notion that life is surprisingly difficult to define, even operationally. The old saw, popular even among biologists, that "I know life when I see it," is cute but not useful in a scientific context. An idiosyncratic definition of life will be offered toward the end of this, the Chemical Epoch.

Living and nonliving systems, then, do not seem to differ *in kind.* Their basic properties cannot be readily distinguished. However, living and nonliving systems do differ *in degree.* All forms of life are more complex than any form of nonliving matter.

As a result, we could reasonably postulate that life is merely an extension of the complexities of matter. If correct, then everything around us—galaxies, stars, planets, and life—might well constitute a grand interconnected spectrum of all known objects in the material Universe, including ourselves. This is the crux, the very heart and soul, of the interdisciplinary subject of cosmic evolution.

A central issue for the Chemical Epoch is Nature's path from simplicity to complexity: Does it always lead from matter to life? In other words, is life's origin merely a natural event, or is it perhaps inevitable? Given the laws of physics and chemistry, as well as the proper ingredients and much time, the subject of biology would seem to arise naturally. The path from atoms to molecules to life seems straightforward enough, based on what is known of modern science today. But it's not a certainty.

An important, though as-yet-unanswered question concerns the direction and nature of the path from matter to life. Is there only one way that complex matter eventually becomes life? Or, can molecules cluster in many ways to create life? Both of these choices are consistent with the basic ideas of cosmic evolution, yet the chances for life elsewhere in the Universe depend critically on which of these two cases pertain.

One case depicts a single path from matter to life. Provided that the environmental conditions are not adverse and time is abundant, matter becomes increasingly complex until eventually some system resembling a single cell originates. We have no way of knowing how long this process of chemical evolution usually takes. The timescale probably depends mostly on the surrounding physical and chemical resources. Temperature, density, energy, and raw materials all play key roles in the origin of life. Furthermore, many false starts are likely wherein life almost forms (or even does so temporarily) only to be destroyed quickly thereafter.

If life definitely emerges once a certain complexity of matter is reached—a critical threshold—we can be reasonably sure that life is not only a *natural* consequence of the evolution of matter but an *inevitable* one as well. A rather direct evolutionary path from matter to life greatly increases the chances that life resides elsewhere in the Universe.

On the other hand, if matter can become complex in many ways, only one (or few) of which leads eventually to life, we cannot justifiably claim that life is inevitably produced from matter. Life would indeed be a natural consequence of the evolution of matter, but not an inevitable one. This second case implies that vastly complex clusters of matter might form without ever crossing a threshold to become a system rightfully judged "living." If so, the prospects for extraterrestrial life are poor.

The truth might also lie between these two extreme cases, with the likelihood for life's origin being moderate and the prospects for extraterrestrial life neither good nor bad. To repeat a theme of this book: Nature is often neither black nor white, but more like shades of gray throughout.

∞

The question of life's origin has engaged the minds of humans since they first contemplated our place on Earth and in the Universe. The subject often elicits emotion—first because it involves ourselves, and second because biochemists do not yet have a comprehensive account of the specific steps that led to life on our planet.

Many people have been raised to accept unquestioningly certain principles, one of which is that life originated by means of a God or gods. The theological or philosophical idea that life resulted from such a *super*natural process is a *belief.* Admittedly, it might be a perfectly good belief, but it remains just that—a belief—for no unambiguous information, acceptable in a laboratory of science or a court of law, confirms the creation of life by a supernatural being or beings. Scientists have no clear data whatsoever supporting the idea that someone or something deposited already-made life on planet Earth long ago. Furthermore, we have no known way to test experimentally the idea that divine intervention created life.

Science is agnostic when it comes to God—not atheistic, as some people prefer to read that laden word wrongly—just agnostic. Aside from personal feelings or cultural persuasions, most professional scientists just don't know what to make of a God or gods. We simply have no bona fide *data* on which to base a judgment.

The belief that life suddenly arose by means of some vitalistic process is outside the realm of modern science. Today's scientific method, which is a philosophy of approach based on reasoned logic bolstered by experimental and observational tests, cannot be used to study supernatural ideas for the origin of life. Accordingly, such ideas, unprovable even in principle, seem destined to remain beliefs forever, hence beyond the subject of science.

Several alternative theories for the origin of life do not require the help of supernatural beings. Each of these theories relies on natural principles and each can be tested experimentally. These theories are thus based on science rather than on theology, and only one of them has thus far survived the test of time, criticism, and debate.

First, life might have originated on Earth by means of panspermia, meaning "germs everywhere." This idea, also called exogenesis, maintains that microscopic living organisms came to our planet from outer space. An asteroid or comet, perhaps containing primitive cells or simple bacteria, could have fallen to Earth at some time in the past, after which they evolved over billions of years into the more advanced forms of life now spread across our planet. That said, no meteorites—the landed debris of asteroids and comets—have ever been shown to harbor bona fide life.

The basic tenet of panspermia is that primitive life, which originated someplace else, was deposited on Earth's surface by means of a collision with some other object that already harbored life. However, most space scientists argue that unprotected simple life would not likely survive the harsh environment of outer space or the fiery plunge into our atmosphere. High-energy radiation and high-speed particles in interplanetary and interstellar space, as well as violent friction and intense heat while moving through air, would almost surely destroy any form of life riding on the backs of small celestial bodies. On the other hand, microscopic spores might survive such alien conditions, provided they're deeply embedded within the incoming rocks. If biologists have learned anything new about life recently, it's that life is very hardy and often capable of surviving in extreme environments.

(Outlandish versions of the panspermia idea abound, perhaps the strangest of them being that life on Earth arose from the garbage dumped here eons ago by extraterrestrial voyagers! Likewise, extraterrestrials might have deliberately seeded our planet, if only because of missionary zeal. These and other bizarre variants of the panspermia theory have fueled science-fiction writers for decades, but working scientists are content with regarding them as truly "garbage theories.")

A related aspect of panspermia has recently become popular—some call it "weak panspermia"—whereby only the ingredients for life, but not life itself, are delivered to Earth from space. With the rash of discoveries of organic molecules in interstellar space during the past few decades, as noted earlier in the Stellar Epoch, some researchers have proposed that not necessarily life itself but the basic chemicals needed for life might have arrived on Earth embedded in comets or asteroids. These molecules could have then acted as seeds that gradually spawned life by natural chemical means—endogenesis, as explained below. It is true that some meteorites, particularly the carbonaceous chondrites known to contain much carbon and to derive from the most ancient asteroids, house an array of chemicals including life's building blocks that apparently did survive the joy ride through Earth's atmosphere. The Murchison meteorite, which fell near Murchison, Australia, in 1969, is the foremost example of this kind of bolide containing raw materials capable of kick-starting life on Earth several billion years ago. Other meteorites have been shown to contain bubblelike organic globules similar to those produced in laboratory simulations of life's origin described later in this epoch, the most recent one having landed in Canada's Yukon Territory just days into the new millennium. What's more, simple organics have been clearly detected in some well-studied comets that recently graced our skies while visiting the inner Solar System, such as Halley, Hale-Bopp, and Hyakutake. At the very least, these findings show that the molecules needed for life can conceivably form in an interplanetary or interstellar environment and that they might have reached Earth's surface unscathed after their fiery descent.

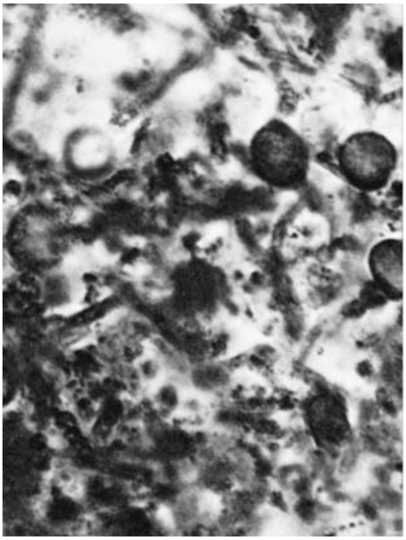
On the other hand, many biochemists argue that organic chemicals could have formed just as easily (and perhaps more so) indigenously on Earth, without looking to outer space for answers to terrestrial puzzles. Even if the notion of panspermia someday becomes a more promising idea for the origin of Earth's life, it does not qualify as a valid theory for the origin of life itself. "Strong panspermia" (whereby intact life falls to Earth like manna from heaven) merely defers the question of life's origin, transferring it to some other, unknown locale in the Universe.

Another theory of life's origin—one that directly addresses the ultimate origin of life itself—goes by the name spontaneous generation. Here, life is thought to emerge rather suddenly and fully developed from peculiar arrangements of nonlife. This idea was popular as recently as a century ago, yet only because people were misguided by their senses. For example, small worms often appear on decaying garbage and mice sometime seem to squirm spontaneously out of dirty linen. Such phenomena were once claimed as evidence for the spontaneous generation of new life from the decayed remains of old life. However, although the observations were correct, the interpretations of those observations were not. Hardly a century ago, most naturalists just didn't realize that flies often lay eggs on garbage, after which the eggs hatch to become worms. Similarly, mice don't originate in soiled sheets, though that indeed may be where they like to hide.

The theory of spontaneous generation was proved incorrect when scientists began carefully monitoring laboratory experiments. The nineteenth-century French chemist Louis Pasteur, in particular, was one of the first researchers to conduct experiments under sterilized conditions. By using specially designed equipment, he was able to show that any parcel of air contains microorganisms, among other unseen contaminants. Without special precautions and close inspection, living matter often comes into contact with nonliving matter, giving the illusion that life originates suddenly in places where no life had existed before. However, by heating the air and thus destroying the microorganisms, Pasteur thoroughly disproved the idea of the spontaneous generation of life. Once sterilized and isolated, air remains free of life, even microscopic life, indefinitely.

A third theory of life's origin is known as chemical evolution. In this idea, prebiological changes slowly transform simple atoms and molecules into the more complex chemicals needed to produce life. Favored by most scientists today, the central premise of chemical evolution stipulates that life arose naturally from nonlife. In this sense, the theories of chemical evolution and spontaneous generation are similar, but the timescales differ. Chemical evolution doesn't occur suddenly; instead, it proceeds more gradually, eventually building complex structures from simpler ones. This modern theory then suggests that life originated on Earth by means of a rather slow evolution of nonliving matter. How slowly and when precisely we are unsure.

Estimates of the timescale over which chemical evolution occurred can be inferred by studying fossils—the hardened remains of dead or-




Evidence of ancient cells.

The photograph at left, taken through a microscope, shows several fossilized cells found in Canadian rock radioactively dated to be nearly three billion years old. The remains of these primitive organisms display concentric spheres with semi-permeable walls and smaller attached spheroids. The image at right shows a magnified view of one of these more clearly. The fossil's inner wall is about a thousandth of a centimeter (or ten microns) across. Many resemble modern blue-green algae. *Source: Harvard.*

ganisms whose skeletal outlines or bony features are preserved in ancient rocks. For example, sedimentary rock, when magnified many times, shows clear evidence for the fossilized imprints of ancient individual cells—the simplest known form of life. Radioactive testing proves that the age of the rock is typically between two and four billion years. This is taken to be the duration of time that the fossils have been buried, presumably having been trapped in the rock while it was solidifying, thus making them some of the oldest fossils ever found.

Knowing that Earth originated four and a half billion years ago and that the oldest rocks crystallized from their early molten state about four billion years ago, we conclude that life likely originated roughly a billion years after Earth formed and no more than a half-billion years after Earth's crust cooled enough to support life. Since even older, as-yet-undiscovered fossils probably lie buried somewhere in Earth's rocks, we surmise that the most primitive forms of life may have taken hardly more than a few hundred million years to evolve chemically from nonlife. Conceivably, they might have taken even less time, even as short as millennia or centuries. Clues to the history and tempo of life's origin are likely written not only in the fossils but also in the cells and molecules of existing organisms.

∞

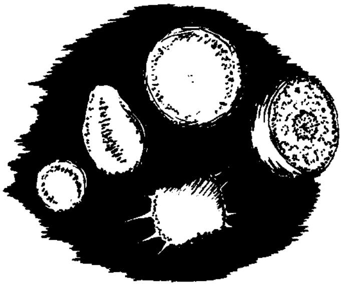
Life may seem biologically, socially, and culturally complex, but physically and chemically it's rather simple. When reduced to its component parts, the basic ingredients of life—any life, from bacteria to whales, and including humans too—are hardly more exotic than two dozen molecules. So to understand the essential properties of life, we need not examine an organism as messy as the entire human body; the molecular nature of contemporary life will do.

All living systems are made of cells, the simplest form of material substance having the common attributes of life—birth, metabolism, and death. From primitive microbes to intelligent humans, the basic unit is the cell. To appreciate chemical evolution—the changes that occurred among atoms and molecules in order to produce life in this, the Chemical Epoch—we need only consider the construction of a cell.

Cells are minute, about a hundred times smaller than a millimeter (or ten microns across), thus invisible to the naked eye. Nearly a thousand such cells would fit within the period at the end of this sentence. A microscopic view shows a central nucleus to be the most complex part of most (but not all) cells. Containing trillions of atoms and molecules, such biological nuclei should not be confused with the much smaller atomic nuclei produced in the cores of stars. Resembling the yolk of an egg, a cell's biological nucleus is surrounded by a thick, fluidic cytoplasm of less complexity. The whole unicellular life-form is encased within a semipermeable membrane through which atoms and molecules can pass in and out.

Cells, then, are the simplest form of life—the "bricks" of anatomical structure. However, they are vastly more complex than the simplest form of matter—elementary particles within atoms. In fact, it's worth stressing that simple cells are more complex than any known type of inanimate matter, lending credence to the evolutionary progression from simplicity to complexity, from matter to life, along the arrow of time.

One of the primary creatures of all, the amoeba, consists of only one cell. More advanced organisms usually contain many additional cells,



. . . simple cells are more complex than any known type of inanimate matter.

often huge clusters of them. A grown human, for example, harbors about a hundred trillion microscopic cells in the guts, skin, bones, hair, muscles, and every other part of our bodies (though only one-tenth, or ten trillion, of these are true human cells, the other ninety percent being bacterial cells that crawl around in our bodies). Each one of these cells furthermore contains very large numbers—trillions or more—of atoms and molecules. The density of basic matter in advanced life-forms is indeed great—roughly two hundred million cells per cubic centimeter—which is why some researchers regard cellular compactness as a rudimentary measure of complexity. But surely there is more to the idea of complexity than just structural density.

Over the course of time, even as brief as a second, large numbers of cells are destroyed owing to the normal process of aging and death. All living systems are nonetheless able to maintain a reasonably constant size and appearance throughout adulthood. Thus, while some cells are dying, others must be forming. Our bodies and those of all other living creatures continually manufacture cell nuclei, cytoplasm, and membrane to sustain themselves throughout life. They do so by means of a curious interaction between the two basic building blocks of life.

The dominant, foundational ingredient of the cytoplasm in any living system is protein, a word deriving from the Greek and meaning "of first importance." Not the name of a particular substance but rather the term for an entire class of molecules, proteins contain large quantities of the element carbon. In fact, fifty percent of the dry weight of our bodies is carbon, largely because each of us harbors tens of thousands of proteins.

Such animate, "organic" substances strongly contrast with things obviously not living, such as a slab of concrete or a pinch of salt. Those things are said to be "inorganic," for they are made mostly of minerals. Inanimate objects have no proteins and their carbon content often amounts to less than about a tenth of one percent of their total weight.

So, carbon atoms play a prominent role in living systems. They play a vital role in the construction of proteins. Unquestionably, carbon is the single most important element in our lives.

Of what are proteins composed? Besides having lots of carbon, is there a common denominator among the myriads of different proteins found throughout the wide spectrum of cells alive today? The answer is yes, for experiments have shown that proteins are made of a rather small group of molecules, called amino acids. Although chemists have synthesized many such acids artificially, only twenty (plus two rare ones) of these structural units compose the millions of proteins found in Earth's life—not just human life: all life. Amino acids are one of the two basic building blocks of life.

Amino acids are not overly complicated substances. The simplest, glycine, is a molecular cluster of five atoms of hydrogen, two of carbon, two of oxygen, and one of nitrogen. Each of these atoms is held to the others by electromagnetic forces of attraction—cohesive chemical bonds involving electric charges. The most complex amino acid is tryptophan, composed of twelve hydrogens, eleven carbons, two oxygens, and two nitrogens.

In principle, the simplest possible protein should theoretically be the combination of two glycine amino acids. An electromagnetic link would couple them, provided that a hydrogen atom is removed from one glycine, and oxygenated hydrogen from the other. This amounts to an extraction of a water molecule (a process called "dehydration condensation"), and guarantees a strong chemical bond between the two glycines.

In practice, life is more complicated, and biochemists are unaware of any real protein as simple as this, for such a two-acid molecule (or "dipeptide") exhibits none of the functions—or job assignments—normally associated with proteins. One of the smallest known proteins in real life is insulin, with fifty-one amino acids linked together like pearls on a necklace. In comparison with the atoms featured in earlier epochs, this simplest amino acid has a mass equal to several thousand hydrogen atoms.

Another well-known protein is hemoglobin, a key component of human blood cells. Containing nearly six hundred amino acids, hemoglobin incorporates in its structure all but one of the twenty different types of amino acids that normally participate in life. The biochemical function of hemoglobin (as well as all other proteins) is highly specific: we know from experience with blood transfusions that blood of one type cannot serve as a substitute for blood of another. The differences among various blood types result partly from the ordering of the amino acids along the protein. Thus, the physical and chemical behavior of a protein—just a long, stringy accumulation of amino acids—depends not only on the *number* of amino acids, but also on the *order* of the acids comprising that protein.

On a larger scale, proteins give some function to cells, and cells, in turn, to entire living organisms. Ultimately, the overall character of life depends on the kind and sequence of amino acids. Only this numbering and ordering distinguishes a human from a mouse, or a duck from a daisy. Since the amino acids are few and relatively simple, the basic nature of life itself cannot be overly complex—at least at the microscopic level.

Mindful of the molecular structure of proteins, we return to our original concern: How are proteins made within organisms in order to keep them alive? Specifically, what chemical process serves to combine amino acids in order to replenish the dead cytoplasm in all living systems? Whatever that process, it must be of central importance, since the production of protein is absolutely vital to an organism's well-being—not any random collection of proteins, but exactly the right kinds of proteins, with their amino acids strung along in precisely the right order. To appreciate how protein is constructed with the correct numbering and sequencing, we defer to the nucleic acids, another of life's basic ingredients.

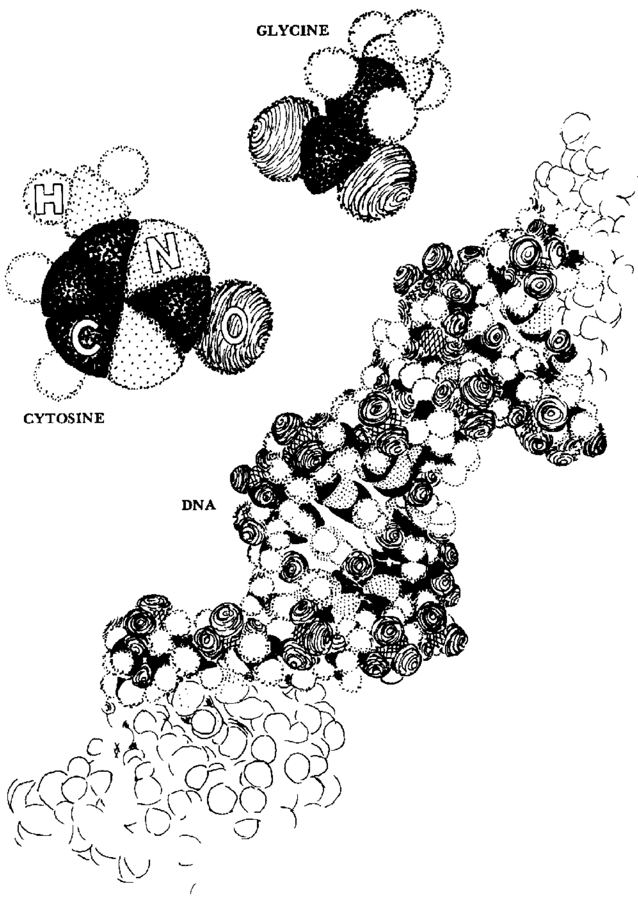
Nucleic acids, like proteins, are long chainlike groups of molecules, most of them also rich in carbon. Their name derives from the fact that these acids were first found in the biological nuclei of cells. Though chemists know of a large variety of them, the nucleic acids, again like proteins, are made of only a small number of key compounds. Called nucleotide bases, these are the second group of life's building blocks. The biochemical role of the bases is best illustrated by the most famous of all nucleic acids—deoxyribonucleic acid, alias DNA.

Most of the DNA molecule is made of a long string of four fundamental bases—adenine, cytosine, guanine, and thymine—that repeat over and over. A fifth nucleotide base, uracil, is used in the construction of other nucleic acids, though not in DNA. These five kinds of bases play much the same role for nucleic acids as do the twenty kinds of amino acids for proteins. Each nucleotide base is only slightly more complex than the amino acids, also being a molecular assemblage of carbon, hydrogen, oxygen, and nitrogen atoms—"CHON," for short. Parts of the bases bend around and attach to themselves in a ring, thereby becoming a little more stable.

Although, at first sight, DNA seems an elaborate molecule, it's really hardly more than a chain of the four types of bases, which form the "rungs" of an extended structure resembling a twisted backbone, or "ladder." Each rung of a DNA molecule consists of two interconnected (or paired) bases, giving this nucleic acid its famous double-helix structure. Experimental evidence shows, however, that all four bases do not bind together equally well. Cytosine always links with guanine, forming one of the two possible base pairs, while adenine links only with thymine, forming the other—C with G, and A with T, that familiar little jingle memorized by all beginning biology students. The structure of the ring-shaped molecules and especially their electromagnetic forces render incompatible any other combinations. In addition, two side uprights, or strands, of the DNA ladder, made partly of sugars (carbon-hydrogen as well as phosphorous-oxygen compounds) that link the base pairs, help shape the DNA molecule.

DNA is only one of many different kinds of nucleic acids, but it stands above all the rest because of one remarkable capability: DNA can copy itself—in effect, replicate. Just prior to the division of a cell, the DNA molecule splits apart by unzipping right up the middle of the ladder. Nucleotide bases floating freely in the cell nucleus then link (with the help of a catalyst called an enzyme) with each of the broken strands. The result is two DNA molecules, where formerly there was only one. The fact that cytosine can bond only to guanine, and adenine only to thymine, ensures that the two "offspring" replicas are identical to the original "parent" DNA molecule. The newly assembled DNA molecules then retreat to opposite sides of the cell nucleus, after which the cell divides into two, with each new cell housing a complete set of DNA molecules.

Preservation of the exact structure of the original DNA molecule is the most important feature of replication. All the information about the specific duties of that type of cell—whether a blood cell, hair cell, muscle cell, or whatever—passes from an old cell to a newly created one. Ac-



. . . in man, mouse, or microbe, the genes mastermind life and the proteins maintain its well-being.

cordingly, the biological function of the "daughter" cell remains identical to that of the "parent" cell. In this way, DNA molecules, whose functional units are the genes, are responsible for directing inheritance from generation to generation.

Just as for amino-acid sequences in proteins, the *order* of nucleotide bases as well as their *number* is paramount in the construction of nucleic acids. The sequence of bases along a nucleic-acid molecule specifies the physical and chemical behavior of that particular gene. In turn, all the genes of a living system collectively form a genetic code—an encyclopedic compendium of the physical and chemical properties of all the system's cells and all their functions. In a very real sense, the two most important features of a living organism—structure and function—depend chiefly on the nucleic-acid molecules in the many nuclei of its cells, for these are the material entities that are passed on, or inherited, from one generation of cells to the next.

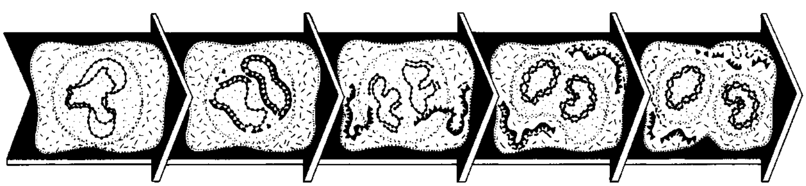
In analogy with another type of information storage—this book, for example—the individual bases can be considered words, the base pairs a sentence, and the whole DNA molecule a book of instructions. The words and sentences must be in the right order to give meaning to the book. An entire library of such instructional books is then comprised by the genetic code for all the varied functions performed by any living organism. In short, a full set of DNA molecules is really information—a blueprint, or master plan, for every life form.

The nature of all living creatures is ultimately prescribed by the structure of their DNA molecules. These molecules specify not only how one type of organism differs from another in both makeup and personality but also how the physical and chemical events inside a cell properly coordinate so that the overall activity of the cell is as it ought to be.

At first glance, it would seem impossible that one molecule could do all this—namely, dictate the behavior of all the myriad life-forms in the world today. After all, DNA has only four types of nucleotide bases. But DNA is the largest molecule known. In advanced organisms such as vertebrates, a DNA molecule can have as many as a hundred million bases or ten billion separate atoms, making the molecule nearly a meter (three feet) long if extended end to end; in humans, about two meters of DNA are squeezed into every human cell, and if all the DNA in a single person were unwound it would stretch about a billion kilometers, or several roundtrips between the Sun and Earth. In the above analogy where a DNA base equals a word, a single DNA molecule would resemble a hundred-page chapter. Consequently, huge numbers of possible combinations of bases guarantee a vast array of diverse living creatures, each with a different appearance, style, and personality. Yet, at the microscopic level, all creatures—without exception—are basically made of the same two dozen or so acids and bases, the very building blocks of life as we know it.

The common molecular content pervading all life on Earth is our best evidence that every living thing dates back to a single-celled ancestor—the so-called LCA, or Last Common Ancestor—billions of years ago.

Regarding our earlier query of protein synthesis, continuous production of the cytoplasm's proteins makes heavy use of the cell's nucleic acids. The sequence of events typically goes as follows: Just prior to cell division, the DNA molecule sends a related RNA molecule out of the biological nucleus and into the cytoplasm. RNA stands for ribonucleic acid—a smaller, single-stranded version of the normally double-stranded DNA (wherein, for RNA, the thymine nucleotide base is replaced by the uracil base). The RNA molecule acts as a messenger, carrying instructions from the DNA molecule. Once in the cytoplasm, single-stranded RNA attracts freely floating amino acids to its uncoupled bases. Only certain amino acids can successfully attach to the RNA bases, since the electromagnetic forces of RNA's bases attract some while repelling others. After some time—usually a few microseconds—the single-stranded RNA molecule fills its strand, partly by accidentally colliding and sticking, with an entire complement of proper amino acids. It's not all chance, however; these acids also attach to one another by means of their own electromagnetic forces, which are governed by physical law. Finally, when the long chain of amino acids is fully assembled along the entire length of the messenger RNA molecule, the chain detaches and drifts off into the cytoplasm of the cell. A protein has thus formed. But, again, this is no ordinary protein created entirely at random. Rather, it's a specific protein formed according to the instructions provided by the RNA molecule, taking into account both chance (the random collisions) and necessity (the linking forces). In this way, RNA acts as a prescription or template on which protein molecules are built—a template originating in the cell nucleus with the DNA molecule itself.



. . . production of the cytoplasm's proteins makes heavy use of the cell's nucleic acids.

This, then, is a highly simplified account of the way that proteins are continually replenished in living organisms. All living systems grow and eventually become biologically stabilized in this same way. The whole process is occurring repeatedly in our bodies right now. Of course, different organisms have different genes and therefore manufacture different proteins—except for identical twins, which do have the same DNA structure. In reality, life is much more complicated, as a single gene does not encode only one protein. One gene often yields a variety of proteins, which partly explains why humanity has hundreds of thousands of different proteins even though the so-called human genome—the sum of all our genes—contains only about thirty thousand genes.

Genes and proteins: The first directs reproduction, the passage of heredity from one generation of life to the next. The second directs metabolism, the daily flow of incoming food (which is high-grade energy) and outgoing wastes (low-grade energy). While genes surely contain the recipes for making proteins, it's the proteins that comprise the (structural) bricks and motor of cells and that do most of the (functional) work. Whether in man, mouse, or microbe, the genes mastermind life, and the proteins maintain its well being.

∞

A central puzzle in modern biochemistry is life's chirality—that is, a tendency for life's molecules to have a certain preferential orientation, or "handedness." Much of life is said to be inherently left-handed, especially its amino acids. No one has ever been able to explain satisfactorily how life became so asymmetric. Yet broken symmetry seems as central to biology and life on Earth as it is to physics and the behavior of matter in the early Universe. Asymmetry may well be an essential prerequisite for the origin and evolution of complexity throughout all of Nature.

Many molecules display two kinds of structures that are mirror images of each other. Their chemical formulas are the same in both cases, but the orientations of some of the molecules' atoms are reversed left for right and right for left. For example, two forms of the alanine amino acid are possible; each is a mirror image of the other, much like our left and right hands are mirror images, as are left- and right-handed wood screws.

A molecule's orientation can be determined by watching the behavior of polarized light passing through it. This is a type of light having an aligned plane of vibration, which can rotate right or left when encountering a material substance. Key molecules of life—especially the usual twenty amino acids that form the structure of all proteins—are almost exclusively of the left-handed variety, since light moving through them rotates left. By contrast, the nucleotide bases and sugars that make up RNA and DNA tend to be right-handed. "Handedness" is one of the most striking properties of life on Earth. But we don't understand it.

Life's amino-acid preference for left-handedness is particularly puzzling because such molecules, when artificially produced in the laboratory, invariably show an equal mixture of left- and right-handed configurations. Furthermore, should a right-handed amino acid drift into a living organism, the catalysts that control protein production will quickly destroy it. Not only that, when a living organism dies and decays, thermal fluctuations change molecular shapes randomly, so that eventually an even left-right mixture results. Why terrestrial life employs only left-handed amino acids (or right-handed nucleic acids) is one of the great unsolved mysteries of chemical evolution.

One possibility is that the first organism just happened, simply by chance, to be left-handed. If life arose only once on Earth, all its descendants would then also be left-handed; the continuity of life is a mere copying process. An alternative, less chancy possibility is that both left- and right-handed organisms originated, perhaps on different occasions billions of years ago, but that left-handed life proved advantageous in eliminating all competitors. The ability to make an extra amino acid or a healthy vitamin, for instance, might have provided such an advantage. If minerals acted as catalytic templates for life's origin, as posited later in this epoch, some crystals might have attracted left- and right-handed amino acids differently. For example, the rock calcite (a common mineral that forms limestone and marble) does display this kind of asymmetric property, possibly acting as a determined selector and not a purely chance event, which could explain why much of life is preferentially left-handed.

Yet another intriguing idea interfaces physics with biology. In brief, life's left-handedness might result from one of Nature's basic forces, once again invoking determinism more than chance. The weak nuclear force operates on size scales smaller than nuclear dimensions and is thus often dismissed as unimportant to atomic physics, let alone molecular biology. However, as noted earlier in the Particle Epoch, the weak force has now been merged with the electromagnetic force, which biologists often refer to as the "life force." And since some weak-interaction events studied in nuclear laboratories do show a preference for one handedness over the other (more elementary particles spin clockwise than counterclockwise during weak-force radioactive events), there might well be a very small (thus far undetected) difference in total energy between left- and right-handed molecules. If true, the left-handed amino acids could have been advantageously selected, as this is the lower-energy state preferred by Nature.

Polarized radiation, wherein waves of energy have specific orientations, is another possibility since energy is needed to drive the production of organic molecules. It is true that circularly polarized light has been detected from distant supernovae, their radiative beams perhaps emitted by collapsed neutron stars, though such specialized radiation has not been noticed coming from the Sun. These light waves move in corkscrew fashion, spinning either clockwise or counterclockwise while traveling through space. Researchers have shown that such light can skew chemical reactions toward producing one type of chiral molecule at the expense of its twin—a preference that could have affected the origin of life's biomolecules on early Earth. Closer to home, star-forming regions, such as the Orion Nebula, emit polarized infrared radiation that also might have favored left-handed interstellar organic molecules capable of reaching Earth while embedded in comets, meteors, or interplanetary dust.

Though these ideas mostly amount to speculation, they well exemplify frontier science at the intersections of physics, chemistry, and biology. Such interdisciplinary efforts will almost surely be increasingly needed, as specialists cross over into one another's disciplines in order to crack the case of Nature's uneven handedness.

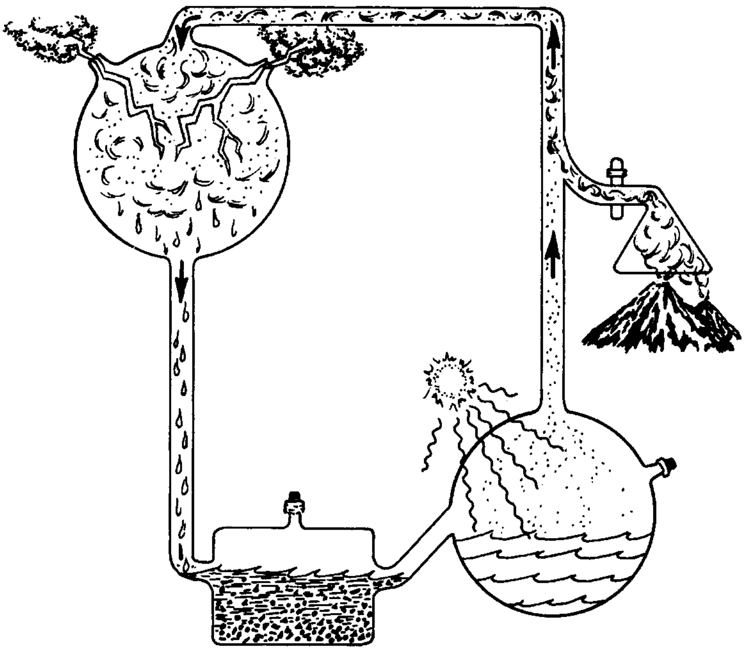
∞

Appreciating contemporary life is one thing, but understanding how it might have arisen from nonliving matter billons of years ago is quite another. Can we be sure that the basic ingredients for life were present, or would have naturally emerged, on primordial Earth? Furthermore, is it likely that those nonliving building blocks could have fashioned a simple living cell given the harsh conditions on our planet billions of years ago? These questions can best be studied in the laboratory, for the atmosphere and surface of today's Earth differ greatly from those of the early Earth. Results of modern chemical experiments that mimic the geophysical environment on our young planet imply affirmative answers to these questions.

First, imagine again the setting on primeval Earth nearly four billion years ago. Physics had done its job to form the planet, and chemistry was in high gear, but biology had yet to begin. As noted previously in the Planetary Epoch, terrestrial gases interacted with one another, as well as with energy, thereby synthesizing bigger molecules. Nothing magical attends this rise in complexity, provided the environmental conditions were not overly adverse and the strength of energy reasonable. Chemistry in action can naturally yield the building blocks of life.

With a test-tube-like contraption capable of holding water and some gases, laboratory gear can be built to simulate Earth's early ocean and atmosphere. The gases—usually a mixture of ammonia, methane, hydrogen, and sometimes carbon dioxide—are meant to match the composition of the secondary atmosphere. Though toxic to present-day life, some blend of this gas was apparently just right for the origin of life. Likewise, the flask of liquid is meant to resemble the primordial seas or some such pool of water. Upon heating this "ocean," its water vapor rises to mix with the other gases in the "atmosphere," whereupon it eventually condenses and "rains" back down with any newly formed chemicals—all of it reminiscent of the familiar evaporation-condensation-precipitation sequence happening every day now on Earth. When the equipment is shut tight, allowing the gases to cycle endlessly without escaping—an "isolated system"—nothing much happens. In the absence of energy, these gases just cycle through the machine unchanged, refusing to react spontaneously with one another. For example, molecules of methane and water vapor, even on direct contact, need a little help in order to react chemically. And that help, that catalyst of sorts, is energy. When energy enters the experiment—an "open system"—it breaks some of the bonds within each of the small molecules, allowing the liberated atoms and molecules to reform as larger, more complex molecules.

In order to hasten the reactions, chemists often employ gas abundances larger than those thought present on Earth long ago. Or they sometimes increase the intensity of the energy above the amount presumed present billions of years ago. In this way, the molecules' likelihood of colliding with one another improves greatly, allowing the experimental simulations to be completed in a few weeks. This does introduce some unrealism and hence some controversy, but quite frankly, researchers with finite careers and one-year grants cannot afford to wait several hundred million years to determine the outcome of their experiments.



. . . no worms or maggots crawl out of this primordial soup—not yet anyway.

After several days of energizing the gases, a thick, reddish-brown, soupy material collects in the trap at the bottom of the apparatus. Chemical analyses show that this slimy product—called "gunk" by some, "pond scum" by others—contains molecules indeed more complex than the initial reactants at the start of the test. Be assured, no worms or maggots crawl out of this primordial soup—not yet anyway. Nor has a simple cell, or even a single strand of DNA, been made under test-tube conditions. But many of the molecular products that are made are among the known precursors of life. They include several of the amino acids and nucleotide bases that constitute the building blocks of all modern life. Although not all the acids and bases common to terrestrial life have yet been identified in the gunk, this "warm little pond," much as theorized by Charles Darwin in the mid-nineteenth century, is regarded as a pretty good approximation of Earth's early ocean into which heavy atmospheric molecules would have fallen, pulled down by relentless gravity.

The recipe for the successful construction of prelife acids and bases is not a very stringent one. This experiment could essentially be done in a household bathtub, though it makes one hell of a mess and is not recommended. The gas mixtures, energy sources, and "cooking" times have been widely varied by chemists throughout the past few decades. The result is invariably the synthesis of complex organic molecules, provided no free oxygen is present. With even small doses of oxygen in the test tube, the gases oxidize, the concoction becomes unstable, and no organic molecules are produced. Ironically, although much of Earth's established life today requires oxygen, this gas was apparently toxic during the formative stages of that very same life. This is why we see no new acids and bases floating in the oceans of today's planet; there's too much oxygen around now.

A critical concern here is the amount and kind of energy used to power these experimental tests. Is it reasonable to suppose that enough of the right type of energy was present on the early Earth? In the laboratory, the simulations are often driven by energy provided by electrodes sparking the gases in the test tube. In reality, those electrical flashes would have been provided by prehistoric lightning storms. Spark discharges can also mimic several other types of energy undoubtedly present on Earth long ago. Besides lightning, plenty of volcanic activity and natural radioactivity were surely present, both of which produce heat. Cosmic rays, fast-moving particles probably among the debris of distant supernovae, also would have energetically belted our planet then much as they do now. Even thunder yields enough energy to have powered, in Earth's early atmosphere, some of the chemical reactions known to occur in the laboratory experiments; if thunder can shatter windows, it can also break (and help reform) chemical bonds. Meteoritic bombardment is a further source of energy; as huge rocks plow through the atmosphere, their friction often generates enough heat to ignite chemical reactions, and their crash landings even more so.

Most of these energy sources are localized, hence were sufficiently intense to make or break molecular bonds only at isolated places on early Earth. Solar energy, however, was widespread, reaching nearly every nook and cranny on the surface of our planet. While ordinary sunlight is not energetic enough to trigger many chemical reactions, solar ultraviolet radiation is. And without oxygen on prelife Earth, an ozone layer would not have surrounded our young planet, thus allowing ultraviolet radiation to have easily reached Earth's surface. Apparently, much the same solar energy that clearly sustains life now was also active in helping to create life billions of years ago.

Laboratory experiments like these are significant because they demonstrate conclusively that the molecular building blocks of life could have been made by strictly nonbiological (i.e., chemical) means in any one of many different ways using raw materials readily present during the early history of planet Earth. These basic ingredients, however, are not life itself. To repeat, the organic molecules found in the gunk are still much simpler than a single cell. The synthesized amino acids and nucleotide bases are in fact much less complex than even the proteins and nucleic acids essential to contemporary life. How, then, were the acids, bases, sugars, and salts in this primordial soup initially assembled into proteins and nucleic acids? The answer is that this dilute organic slime must have been further concentrated so as to permit stronger and drier interactions.

As noted earlier, two amino acids can be linked to reach the next stage of complexity, provided a water molecule is removed. Such a dehydration condensation of many amino acids can then build up chain molecules into complex proteins. Successive linkages of nucleotide bases and energy-rich sugars can likewise produce lengthy nucleic acids.

Heat, for example, could have evaporated some water from clusters of acids and bases, especially along the shoreline of an ancient ocean or a lagoon inlet. Repeated in-and-out sloshing of tides in shallow waters might have led to a daily cycle of solar desiccation of molecules in a temporarily dried tributary during low tide, followed by further interaction among those molecules when washed into the open ocean at high tide.

The opposite condition—cold—can also effectively remove water molecules from an organic mixture. The freezing of water transforms it from liquid to ice, thus allowing the acids and bases to become more concentrated and hence linked together. Regular freezing and thawing could have allowed the buildup of progressively larger chain molecules.

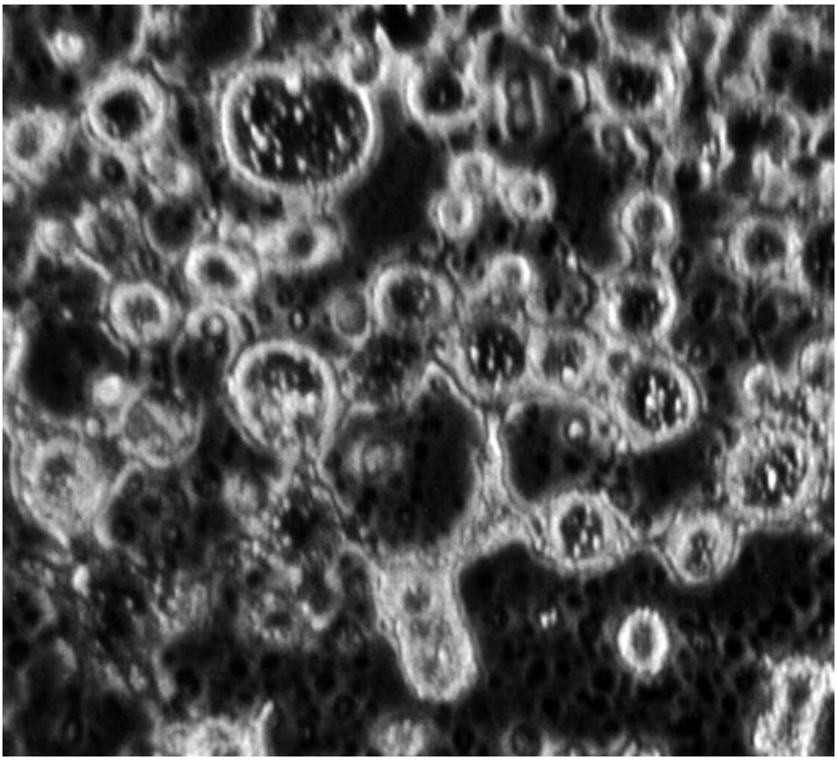
A third mechanism can actually remove water while reagents are still in the presence of water. Although this sounds impossible, it happens all the time in living organisms; composed mostly of liquid, the cells in our bodies routinely manufacture protein. They do it by using catalysts—third-party molecules that act like brokers by speeding up the process. Although the catalysts that now promote condensation reactions in today's life-forms were probably absent in the primordial ocean, chemists speculate that other catalysts likely existed four billion years ago. Certain kinds of clay commonly made by the weathering of rocks, for example, are thought by many researchers to have been the scaffolding needed to make larger organic molecules along the edges of oceans, lakes, and rivers. Clays, having layered, charged surfaces, could have potentially acted not only as tiny compartments to shelter acids and bases, but also as templates to assemble them into long, stringy substances.

Chemists are unsure if the first complex proteins and nucleic acids really did originate in any of these ways. The fossil record will probably never show the precise path whereby prelife molecules gradually coalesced into something that might be genuinely called life. Nonetheless, heating, freezing, and catalyzing are all plausible agents for the assembly of small amino acids and nucleotide bases into larger proteins and nucleic acids. Some of these, in turn, owing to their hydrophilic (water-loving) and hydrophobic (water-repelling) properties, coiled up and folded over. As such, they became microscopic bags of chemicals enclosed by thin membranes. Somehow they assumed forms that look like cells.

A single cell is astonishingly more complex than any of these prelife molecules. To reach this very root of the evolutionary tree, biochemists currently seek to understand how proteins and nucleic acids were able to forge more intricate combinations of biological significance. Understanding in this area is limited, however. Researchers have only been able to surmise that persistent interactions among the many molecules on early Earth could have eventually produced something resembling today's proteins, DNA, and simple cells.

More advanced laboratory experiments in recent years support this view. Repeated energizing and dehydrating of the simulated environment of primordial Earth produce organic molecules more complex than the amino acids and nucleotide bases. Of special interest are minute clusters of up to a billion amino acids united under the influence of heat. These "proteinoid microspheres" (also called "coacervate droplets") do resemble proteinlike substances that resist dissolution in water. Only about a hundredth of a millimeter (or ten microns) across and hence requiring a microscope for observation, these are not well-known proteins such as insulin or hemoglobin, but simpler, proteinlike compounds whose relevance to the origin of life is unclear. Chemical analysis confirms these microspheres to be dense little sacks of organic matter floating in a watery, mostly inorganic fluid. A view through a microscope shows them shimmering like globs of oil on the surface of water, or grease that bonds together like droplets on the surface of cooled chicken broth. Some chemists regard such microspheres as bona fide proteins; others aren't so sure.

Remarkably, the proteinoid microspheres made in laboratory experiments behave to some extent like true biological cells. The microspheres have a semipermeable membrane through which small molecules can enter from the outside as a kind of "food," but through which most larger molecules created within cannot get out. Some discharge of "waste" is noticeable through a microscope but, by and large, these proteinoids display a net intake of matter, in some ways mimicking today's



Evidence of laboratory-made organics.

These oily, hollow droplets rich in organic molecules were made by exposing a freezing mixture of primordial matter to harsh ultraviolet radiation. When immersed in water, these curious little blobs display cell-like structure; most span about ten microns across (or a thousandth of a centimeter). Although not alive, they bolster the idea that at least life's building blocks could have come from extreme environments, such as underwater vents or interstellar space. *Source: NASA.*

biochemical cells. Indeed, these curious little bags of chemicals are actually seen to become larger in the process.

We can therefore loosely imagine the proteinoid microspheres as eating, growing, and excreting—a primitive metabolism, perhaps. Not only that, when the experimental gear is jostled to create some turbulence in the fluid—the analog of early oceanic wave action—some of the larger microspheres fragment into smaller ones, suggestive of a primitive form of replication. Some of these smaller, second-generation microspheres disperse, an apparent "death." Others enlarge like their "parents," only to be ruptured by another act of "replication" (although these microspheres surely lack enough information to direct their own replication from the basic building blocks). Environmental selection is underway.

In all, the fascinating proteinoid microspheres roughly approximate simple bacterial cells, especially the most ancient cells found in the fossil record and examined closely next, in the Biological Epoch. Some of the microspheres "eat," some "grow," some "reproduce," and some "die." Can they be called life? Probably not: most researchers say almost certainly not since the microspheres lack nucleic acid or genetic coding. Yet who is to say what the first cells favored—protein metabolism or genetic reproduction? Or that protocells even remotely resembled modern cells? The distinction between matter and life is not clear-cut. And life itself, as noted earlier, is hard to define.

The great majority of biologists argue that amoebas are definitely alive but that the molecular contents of the organic soup are not. Proteinoid microspheres apparently lie somewhere in between. But if they are not at least progenitors of Earth's living systems—a kind of protolife—then Nature seems to have played a malicious joke on modern science.

The fuzzy interval between life and nonlife often troubles scientists and laypersons alike. The central idea of chemical evolution is straightforward enough: Life evolved from nonlife. But aside from biochemical intuition and laboratory simulations of some likely events on primordial Earth, do we have any direct evidence for naturally occurring complex systems within that blurred realm between living organisms and nonliving molecules? Fortunately, the answer is yes.

Virus particles are the smallest and simplest entities that sometimes appear to be alive—"sometimes" because viruses seem to display attributes of both nonliving molecules and living cells. A Latin word for "poison," viruses are of course a common cause of disease, but they may also hold clues to the origin of life. Although they come in many microscopic sizes and shapes, all viruses are smaller than a typical, modern cell; some contain only a few thousand atoms and span hardly a micron across, or a millionth of a meter. At least in terms of dimensions, viruses seem to bridge the gap between cells that are alive and molecules that are not.

Viruses contain both protein and DNA (or RNA), though not much else—no unattached amino acids or nucleotide bases by means of which living organisms normally grow and reproduce. How, then, can a virus be considered alive? When alone, it's not; a virus is absolutely lifeless when isolated from living organisms. But when inside a living system, a virus has all the properties of life. Viruses live by injecting their DNA (or RNA) into cells of healthy living organisms, after which the virus's genes seize control of the cells and establish themselves as the new master of chemical activity. Viruses then grow and reproduce copies of themselves by using the free amino acids of an invaded cell, often robbing the cell of its usual function. Some viruses multiply rapidly and wildly, spreading the disease and, if unchecked, eventually killing the invaded organism.

Biochemists are therefore unable to classify viruses as either living or nonliving. Their status depends on their environmental circumstances. Even in the modern world, life seems to shade imperceptibly into nonlife. Viruses apparently exist within that gray, uncertain realm.

One of the strongest criticisms of the laboratory simulations of life's origin is the energy needed to drive the experiments. This is especially problematic for the proteinoid microspheres, which require a great deal of heat to form, so much so that only seething volcanoes could have likely provided it. Yet who is to say that volcanoes were not helpful in exactly this way, for they surely must have been frequent and widespread during Earth's youth.

If we are willing to relax the notion that life formed on the surface of the primordial sea, the submerged tectonic cracks and oceanic ridges noted in the Planetary Epoch become potentially even better candidate sites for life's origin. For there, there's plenty of concentrated energy and not much oxygen. Recent explorations by miniature submarines have made underwater, hydrothermal vents increasingly popular places to postulate life's emergence on Earth. At places along the Mid-Atlantic

Ridge and near the Galapagos Islands, complex ecosystems harboring many diverse life-forms are known to exist, indeed to thrive, all powered by suboceanic heat engines, quite independent of the Sun.

Called "black smokers" (owing to their abundance of iron and sulfur), submarine vents are narrow crevices in the seafloor through which

Evidence of extremophilic life.

A small, two-person submarine (called Alvin, partly seen at bottom) took this picture of a hydrothermal vent, or "black smoker"—one of many along an underwater ridge in the eastern Pacific Ocean. As scalding hot water pours from the top of the vent's tube (near center), black clouds rich in sulfur billow forth, providing a strange environment for many odd life-forms that manage to thrive under totally dark and oxygen-free conditions near the vent. *Source: Woods Hole Oceanographic Institute.*

pressurized hot water (up to hundreds of degrees Celsius) squirts like a Roman fountain. The heated water, rich in metals leached from molten rocks below, interacts with the much colder seawater above, creating strong thermal gradients that enhance energy flow. The hot fluids actually rise within mineral-laden "chimneys" perched atop the suboceanic crust astride oozing magma, driving and sustaining much biological activity in, or at least near, the vents—but not conventional life-forms familiar to us elsewhere in the biosphere. Often called "extremophiles" or "thermophiles" given their surprisingly high-temperature environment, heat-loving bacteria, among other peculiar vent life, is dependent neither on oxygen nor on sunlight. Among them are the so-called archaeabacteria, a relatively newly discovered domain of life that harks back to some of the most ancient life-forms and that currently coexists within a remarkable community of two-meter-long (six-foot) worms, ten-kilogram (twenty-pound) clams, and idiosyncratic microbes thriving under what we at the surface would call decidedly uncomfortable conditions.

Undersea hydrothermal vents could well have been the natural engines that drove the early emergence of biology several billion years ago. They do have some advantages over the formation of life on or near Earth's surface, providing an environment abundant in heat, lacking free oxygen, and clearly protected from the harsh realities of incoming ultraviolet radiation and asteroid bombardment that presumably made early surface conditions a Puritanical hell on Earth. Areas near such vents could have conceivably fashioned biology out of a geological setting, even if only at a single locale that successfully bridged the Planetary and Biological Epochs.

Another dissenting viewpoint about life's origin either on or under the surface of Earth has gained strength in recent years, reinforcing the honest uncertainties plaguing the Chemical Epoch. Here the question concerns a possibly wider venue for life's origin: Did it occur terrestrially anywhere on Earth or extraterrestrially someplace beyond? Endogenesis or exogenesis? Some astrobiologists feel that none of Earth's land, sea, or air might have been particularly well suited for the initial production of organic molecules. Not even the undersea vents are viable, they say, as the heat there may have been too great for the survival of acids and bases—in effect, even harsher than early conditions on Earth's surface or in its atmosphere.

At issue once again is energy, that is, whether proper amounts of it—optimum values, neither too much nor too little—were available to power chemical reactions. Moreover, Earth's early atmosphere might not have contained enough raw material for the reactions to have become important in any case. A minority of researchers argue that much, if not all, of the organic matter that combined to form the first living cells was more likely made in interstellar space and thereafter arrived on Earth embedded in comets, meteors, or interplanetary dust, parts of which managed to avoid burning up during their violent descent through the atmosphere.

Several pieces of evidence support this idea, considered by some to be a kind of weak panspermia whereby the molecular ingredients for life were brought intact to Earth—though not necessarily already formed life itself. First, there are the interstellar molecules noted in the Stellar Epoch, many of which contain carbon and at least one of them reported (yet unconfirmed) to be the amino acid glycine. Second, laboratory experiments demonstrate that when icy mixtures of water, methane, ammonia, and carbon monoxide—precisely what's found in the near-vacuum of interstellar space—are exposed to ultraviolet radiation like that from a newborn star, the result is intriguing and perhaps more realistic than some of the earlier simulations of chemistry on the young Earth. When the irradiated ice is later placed in water, oily, hollow droplets form with cell-like dimensions and obvious membranes made of organic matter. As with the proteiniod microspheres noted earlier, these interstellar globules contain neither proteins nor DNA per se, but the results clearly show that even the alien, cold, virtual vacuum of galactic space is an apparently suitable place where simple protocellular structures can form—especially when they splash into a receptive ocean. And third, comets and meteorites are known to harbor organic matter, the comets especially, often called "dirty snowballs," being made mostly of the icy interstellar mixture just described. Since cometary impacts are thought to have provided much of Earth's water, it's perhaps only a small step to imagine that this incoming water already contained the building blocks of life.

The idea that organic matter constantly rains down on Earth from space in the form of interplanetary debris is certainly plausible. The cratering record on the Moon shows that Earth experienced a period of late bombardment a little less than four billion years ago, just a bit earlier than when the oldest life-forms appear in the fossil record. Tens of thousands of tons of extraterrestrial matter do fall to Earth annually, even now. And the notion that chemical evolution occurs in space seems certain. Analyses of comets, meteorites, and interstellar gas during the past two decades have proved beyond doubt that organic chemistry is widespread in the Universe. However, whether or not exogenesis was the primary means by which complex molecules first appeared in Earth's oceans remains unclear. The origin of life, along with the origin of galaxies, represent the two chief missing links in all of cosmic evolution.

∞

Energy is the one absolute requirement—in addition to raw materials—for any of the scenarios of life's origin. Energy, in fact, seems central to all aspects of evolution, regardless of whether that evolution involves systems that are living. Neither inanimate matter nor animate life can proceed from a simple to a complex state without energy. Complex objects have some organization, and organization of any kind requires energy—for formation, for maintenance, and for further changes. Even when fully structured and highly evolved, no advanced form of matter, whether stars or people, can sustain itself without a regular flow of energy. This energy is a fuel, a food of sorts.

In the case of the laboratory simulations just described, energy derived from the spark discharge mimics an "explosive food" used to fracture bonds of the small molecules. Part of that energy is also absorbed, enabling the molecular fragments to reunite into bigger groups of atoms. And part of it strengthens the chemical bonds needed to hold together—to reorganize—the new, more complex acids and bases. The organic scum floating on or near the surface of the primordial ocean thus became a tremendous storehouse of energy.

Repeated energizing—that is, regular feeding—was needed to construct the microspheres, globules, or whatever we wish to call those first, protocellular entities. Once formed, the organic droplets required even more energy to maintain their increasingly intricate molecular structures. They likely did so by absorbing nutritious amino acids and nucleotide bases admitted through their semipermeable membranes. The protocells then extracted energy by breaking some of the chemical bonds among the atoms forming those acids and bases. In this way, they essentially "ate" by absorbing minute amounts of energy from their surroundings.

Why did the protocells obtain energy from their immediate environment? Why didn't they continue to utilize one of the external types of energy, such as solar radiation, atmospheric lightning, or volcanic activity? The answer is that the energy that helped form the ancient protocells in the first place was often too harsh to sustain them later. As molecules become larger and more complex, they also become more fragile. They had to eat and organize themselves by absorbing energy, but that energy had to be slight and gentle. (It's a little like the difference between watering a plant and drowning it.) The small acids and bases able to pass through the minute openings in a protocell's membrane contain just the right amount of energy. They enable protocells to survive without being subjected to the harsh external energy originally needed to produce them. Although chemists have no direct evidence for the assembly of more advanced precursors of life, laboratory studies strongly support a two-step process like that outlined above: a moderate dose of energy was first needed to synthesize the precursors, after which milder energy was needed to maintain them.

A combination of circumstantial evidence and biochemical insight leads scientists to surmise that proteinoid microspheres, or something like them, were able to protect themselves from the uncontrolled energetic conditions that created them several billion years ago. This is not unreasonable, since Earth was rapidly cooling at the time, becoming less geologically active. As time passed, volcanoes, earthquakes, and atmospheric storms would have gradually subsided. The amount of solar ultraviolet radiation reaching the ground would have also diminished as terrestrial outgassing thickened the atmosphere. Many of these prebiological, microscopic clusters probably found shelter under thin layers of water, which can absorb whatever harsh solar radiation did manage to penetrate the air.

From this point on, biochemists can only presume that at least one protocell was eventually able to evolve into something everyone would agree is a genuine living cell. However, nothing yet discovered in the fossil record documents this prelife evolutionary phase. Nor have laboratory simulations of Earth's early conditions produced molecular structures more complex than the proteinoid microspheres; these organic globs possess neither the hereditary DNA molecule nor a well-defined nucleus common to most contemporary cells. Alas, researchers cannot presently explain how the first protein might have arisen from a medium containing no nucleic acids, especially when the passage of information from nucleic acid to protein is widely considered to be the central dogma of modern molecular biology.

The issue of which came first, proteins or nucleic acids—that is, "protobionts" or "naked genes"—resembles yet another chicken-or-egg paradox and clearly represents one of the biggest puzzles in all of cosmic evolution. Quite possibly, the capacities for metabolism and reproduction developed in parallel, but we don't know for sure. One way out of this dilemma notes that RNA—the single-stranded cousin of DNA—can act as both replicator *and* catalyst, in effect both the chicken and the egg. If so, then perhaps RNA, or some version of it, preceded both DNA and proteins in the primordial soup—making RNA life's chief precursor. Such "ribozymes" (analogues to protein enzyme catalysts), recently discovered in laboratory experiments among contemporary life, might have performed double duty billions of years ago by storing small amounts of information *and* catalyzing their own reproduction. Eventually, that "RNA world" must have evolved into the more complex one of today wherein DNA and proteins have separate, though complementary roles.

Other biochemists contest this idea, arguing that some type of metabolic, energy-driven chemistry must have existed even before RNA came on the scene. For them, the chicken-or-egg question will not be resolved until we understand the underlying chemical pathway that transformed raw organic matter into RNA itself. This is Darwinian evolution among prebiotic molecules—variation, competition, selection, and amplification of the fittest chemicals. Variants of simple molecules were tried and retried by Nature, acting in concert with the rules of thermodynamics and the help of energy to produce protometabolic means (called thioester bonding) to aggregate larger molecules on the road to the RNA world.

Frankly, no one knows for sure if any of these origin-of-life scenarios are correct. A notable gap plagues our *direct* knowledge of the *precise* events that occurred between the synthesis of life's precursor molecules and the appearance of the first genuine cell. The uncertainty should not be so surprising, given that these events occurred several million millennia ago.

∞

The concept of complexity keeps entering our cosmic-evolutionary story. This is especially so now that we are about to encounter animate life-forms, which everyone agrees are decidedly more complex than any inanimate systems. Soon, we shall be confronted with a central question of considerable import: How did the neural network within human brains acquire the complexity needed to build societies, weapons, cathedrals, philosophies, and the like? For what we humans culturally create now—including books like this one—is as much a part of cosmic evolution as the stars that fused the heavy elements or the planets that fostered the origin of life.

While relating this Chemical Epoch, we seem to be crossing a boundary—that between nonlife and life. Yet, in reality, there is no boundary here. To stress one of the main arguments of this book, while chronologically probing the arrow of time from the nonliving to the living—from physics and chemistry to biology, mainly—systems of greater complexity have, each in its turn, emerged. But as also noted earlier, when examined closely, living systems do not differ basically from the nonliving. Scientists have never found any evidence for a mystical *élan vital* that grants life some special, paranormal quality. Instead, all ordered systems seen in Nature differ not in kind but only in degree—namely, the degree of complexity. Just what do we mean by complexity, and how does it rise over the course of time?

Roughly halfway between the big bang and humankind is a good place to take a finer, slightly more technical look at the course of action likely embraced by ordered systems at their origins. We have already met a wide range of physical, inanimate systems, including galaxies, stars, and planets; we are about to meet a whole gamut of biological, animate systems, including plants, animals, and ourselves. How realistic is it that all these systems can be arrayed along a continuous spectrum of rising complexity with time? And what are the mechanisms—perhaps even a single set of unifying actions—that have brought forth all this impressive order and organization?

One way to address complex systems on a common basis, that is, "on the same page," is to appeal to thermodynamics. That's because, of all the known principles of Nature, thermodynamics has the most to say about the concepts of change and of energy, and especially the *changes in energy* that seem key to the origin and evolution of all ordered systems. Literally, "thermodynamics" means "movement of heat"; for our purposes here (and in keeping with the wider Greek connotation of motion as change), a more insightful translation would be "change of energy."

One of the most basic and cherished laws in all of science is the so-called second law of thermodynamics. (The first law merely states that total energy is conserved before and after any change.) The second law dictates that randomness or disorder, the technical term for which is "entropy," increases everywhere. In other words, Nature demands that a price be paid each and every time an energy transaction occurs. And that payment means that less energy is available to potentially drive change; energy is not lost, just rendered unavailable for useful work. This is true because heat (i.e., thermal energy) naturally flows from a hot source to a cold source, whether among molecules in a gas, stars in a galaxy, or flesh and blood in our bodies. The net result is diminished energy differences, causing events to run down and gradients to even out; all the while entropy inevitably increases.

Nature does abhor a vacuum; some say it abhors a gradient of any sort. Generally, matter and radiation tend to disperse (like perfume escaping from a bottle), that is, to occupy places where they had no presence—and once there, to seek a natural balance or evening out, an equilibrium. A simple example is a pendulum that eventually stops swinging once it attains its middle, lowest position; while oscillating back and forth it has uneven energy of motion, whereas at rest it's equilibrated and thus no longer working. To shorten a much longer argument by giving another pair of examples, an imperative of thermodynamics' second law is that a house of cards, once built, will tend to collapse with time; by contrast, a random collection of playing cards is not likely to assemble itself into some sort of structure. Similarly, water will of its own accord flow over a dam into a lake below, but has never been seen flowing back up to the top of the dam. These are classic examples of "isolated systems"—those closed to the environment beyond—wherein events occur in only one direction. Nature is said to be irreversible and asymmetric.

By contrast, Nature also houses "open systems" whereby energy (and sometimes matter, too) enters from the environment outside such systems. And this can make a great deal of difference. We could, for instance, by exerting some energy (and some patience, too, which also burns energy), rebuild a house of cards or reswing a pendulum clock by rewinding (i.e., energizing) it. A water pump could also be used to transport water from a low-lying lake to a place above a high-lying dam, but that also requires energy from outside the lake-dam system—again, energy obtained from the environment beyond. Usually, systems having energy flowing through them are not in equilibrium, which is why they are often also called open, nonequilibrium systems.

The infusion of energy (or matter) into any system can potentially yield organized structures. Disorder (or entropy) can actually decrease within open systems—which galaxies, stars, planets, and life-forms most assuredly are—even though that disorder increases everywhere else in the Universe. Such "islands of structure" do not violate the second law of thermodynamics because the net disorder of the *system and its environment* always increases. The energy needed to run a water pump—or any such device of our industrial civilization—is acquired at the expense of the environment, which is generally ravaged to power today's technological society. Both drilling for oil and burning it disorder (or pollute) the environment more than the order (or edifice) gained by the energy needed to light a home or drive a car. Truth be told, the second law of thermodynamics is not an environmentally friendly principle of Nature, yet it does allow organization to emerge temporarily—typically seventy years for humans, billions of years for stars and galaxies.

The energy used by humans to build anything—a house of cards, a table, chair, automobile, whatever—derives from the food we eat. We literally feed off our neighboring energy sources—locally plants and animals, and more fundamentally the Sun. The Sun, in turn, derives its energy from the Galaxy, specifically the conversion of the gravitational potential energy of its parent galactic cloud into the heat that triggered stellar nuclear fusion. And our Galaxy among all the other galaxies, in turn yet again, owe their existence to gradients established in the early Universe—and to the resultant energy flows made possible by cosmic expansion that broke the primeval symmetry between matter and radiation, thereby establishing the nonequilibrium conditions ultimately needed for the growth of ordered systems everywhere.

Recall for a moment those extremely hot and dense conditions in the Particle Epoch of the early Universe. Until neutral atoms began forming—that fundamental phase change some half-million years after the big bang—matter and radiation were intimately coupled. Equilibrium prevailed during the Radiation Era as a single temperature was enough to specify both matter and radiation—a physical state lacking order or structure, indeed one characterized by maximum entropy or minimum information content. Equilibrated systems are simple systems, requiring little information to describe them.

And that brings us to the heart of the issue regarding complexity. In some ways, the complexity of a system is a measure of the amount of information needed to describe that system. Operationally, it's also related to the amount of energy flowing through a system of given mass. *Complexity*: a state of intricacy, complication, variety, or involvement, as in the interconnected parts of a structure—a quality of having many interacting, different components.

In the early Universe, the absence of a temperature gradient between matter and radiation mandated nearly zero information. There were then no structures, no appreciable order, no complexity beyond unclustered elementary particles zipping around in a uniform field of radiation. Rather, mostly everything was part of a homogeneous, chaotic frenzy in the aftermath of the big bang. One temperature, although declining rapidly, was sufficient to model the early history of the Universe, for the overgreat density then produced so many collisions as to guarantee an equilibrium. Once matter and radiation decoupled, however, equilibrium was destroyed, symmetry broken, and the Matter Era began. Two temperatures were thereafter needed to describe the evolution of matter and radiation. Accordingly, a cosmic thermal gradient was naturally established, the essential result being a flow of energy available to work—indeed, potentially, to "build things."

The very expansion of the Universe, then, drives order from chaos; the process of cosmic evolution itself generates information. How that order became manifest as galaxies, stars, planets, and life-forms has not yet been deciphered in detail. But we can now appreciate how natural systems eventually emerged—ordered physical, biological, and cultural systems able to create and maintain information by means of localized reductions in entropy.

Furthermore, because the two temperatures portraying the Matter Era diverge—that is, their difference grows larger with time—their departure (even today) from thermodynamic equilibrium allows the cosmos to produce increasing amounts of information. That's because energy flows also increase with departures from equilibrium, and with them the potential for the growth of order. We thereby seemingly have a way to understand, at least in general terms minus the details, the observed rise in complexity throughout the eons of cosmic time—not just stars and galaxies but also structures as intricate as single cells or contracting muscles, let alone the neural architecture of human brains.

Enough about inanimate, nonliving objects: we are now well into the Chemical Epoch, on the threshold of life itself. And here we see more clearly the limited role of chance in Nature, as in all aspects of cosmic evolution. To be sure, chance cannot be the sole instrument of change. Determinism—meaning neither reductionism nor mechanism but simple obedience to precise, natural laws—must also play a part in all things that change.

Consider the precursor molecules of life's origin, as noted earlier in this epoch. Simple molecules such as ammonia, methane, water vapor, and carbon dioxide react with each other in the presence of energy to generate larger molecules. The end products are not just a random assortment of molecules; they comprise most of the two dozen amino acids and nucleotide bases common to all life on Earth. And regardless of how this chemical-evolutionary experiment is performed (provided the gases simulating our primordial planet are irradiated with realistic amounts of energy in the absence of free oxygen), the soupy organic matter trapped in the test tube always yields the same relative proportions of proteinoid compounds. The point is that if the original reactants were re-forming into larger molecules by chance alone, the products would be among billions upon billions of possibilities and would likely vary each time the experiment was run. But the results of this experiment show no such diversity. Of the myriads of basic organic groupings and compounds that could possibly result from the random combinations of all sorts of simple atoms and molecules, only about fifteen hundred are actually employed on Earth; these groups, which constitute the essence of terrestrial biology, are in turn based upon only about fifty simple organic molecules, the most important of which are the above-noted acids and bases. Some factor other than chance is necessarily involved in the prebiotic chemistry of life's origin, though one need not resort to mysticism. That other factor is the electrical bonding influence naturally at work among the microscopic molecules—forces that guide and bond small molecules into the larger clusters appropriate to life as we know it, thus granting the products some stability. A ring array, for instance (such as the benzene molecule), is a good deal more stable than a linear array of the same atoms and molecules. And it doesn't take long for reasonably complex molecules to form, not nearly as long as probability theory predicts by a chancy assembly of atoms. In short, the well-known electromagnetic force acts as a molecular sieve or probability selector, fostering only certain combinations while rejecting others and thereby guiding organization from some of the randomness.

Molecules more complex than life's simple acids and bases are even less likely to be synthesized by chance acting alone. For example, the simplest protein, insulin, comprises fifty-one amino acids linked in a specific order along a molecular chain. Probability theory tells us the chances of randomly assembling the correct number and order of acids: Given that twenty amino acids are involved, the answer is $1/20^{51}$, which equals $1/10^{66}$. This means that the twenty acids must be randomly assembled 10^{66}, or a million trillion trillion trillion trillion trillion, times for insulin to form on its own. As this is obviously a great many permutations, we could randomly arrange the twenty amino acids trillions upon trillions of times per second for the entire history of the Universe and still not achieve *by chance* the correct composition of this protein. And clearly, to assemble larger proteins and nucleic acids, let alone a human being, would be vastly less probable if it had to be done randomly, starting only with atoms or simple molecules. This is not at all an argument favoring supernaturalism; rather, it's once again the natural forces of order that tend to tame chance—much as was the case for the origin of galaxies in the earlier Galactic Epoch, when Nature was unable to form galaxies by chance and chance alone.

These are classic cases, whether among atoms in astronomy or molecules in chemistry, of Nature's twin actors—chance and necessity—jousting again with one another. And of the mechanism of selection at work as well—yet not to select "in" the "winners" as much as to select "out" the "losers." The process of selection, guided mostly by the laws of physics, serves to eliminate systems, whether molecules or galaxies, that are incompatible with their changing environments. In all such phenomena, including changes among life itself, elements of chance are often present but so are the deterministic physical laws that serve to constrain chance—to limit its effectiveness, to restrict its randomness, to ensure likely outcomes even in the presence of chance. The two operate in tandem, often triggering change in many types of systems (that's the chance part), followed by nonrandom elimination of those systems that are not optimized to their newly altered environments (that's the deterministic part). As is well known, chance, necessity, and selection are essential features of the modern Darwinian paradigm of biological evolution. But they have their roles to play in the inanimate world as well.

On the threshold of the next, Biological Epoch, how are we to analyze living systems per se, including biological structure and function, let alone attempt to define life? Surely, entropy must decrease during life's origin and evolution, for living systems are demonstrable storehouses of focused energy and much order. Once again, as earlier, thermodynamics is the key. As with other objects in the Universe, we can use the concepts of information content and energy flow to describe both the structural and functional aspects of biological organization, indeed to define life itself.

All things considered, biological systems are best depicted by their coherent behavior, for their maintenance of order requires a great number of metabolizing and synthesizing chemical reactions as well as a host of intricate mechanisms controlling the rate and timing of life's many varied actions. But this doesn't mean that life violates the second law of thermodynamics, a popular misconception. Although living organisms manage to decrease entropy locally, they do so at the expense of their environment—in short, by increasing the overall entropy of the remaining Universe.

Living things are often said to circumvent temporarily the normal entropy process by absorbing available energy from their surrounding environment. But even "circumvent" is too strong a verb, implying that life is somehow outside the usual bounds of thermodynamics. In reality, living things extend the traditional study of what is really thermostatics into the realm of genuine nonequilibrium thermo*dynamics*. They do so, during both their origin and their evolution, because of temperature gradients naturally established on Earth. What is the source of these thermal differences and ultimately of the energy utilized in the process of living? On Earth, it's our Sun. Energy flows from the hot (six-thousand-degree-Celsius) surface of the Sun to our relatively cool (twenty-five-degree-Celsius) planet. All of Earth's plants and animals depend for survival on the Sun, whose energy can be converted to useful work. Plants photosynthesize by using direct sunlight to convert water and carbon dioxide into nourishing carbohydrates; animals obtain solar energy more indirectly by eating plants and other animals.

By contrast, if left alone without any energy input, all living things, much like all else in Nature, tend toward equilibrium. While just twitching a finger (or merely thinking while reading this page), humans expend energy and eventually tire. Any action taken indefinitely, without further energizing, would drive us toward an equilibrium state of total chaos or orderlessness. That human beings manage to stay alive by steadily maintaining ourselves far from equilibrium is testimony to our evolved ability to handle optimally the flow of energy through our bodies. In point of fact, unachieved equilibrium can be taken as an essential premise, even part of an operational definition, of life, to wit:

Life: An open, coherent spacetime structure kept far from thermodynamic equilibrium by a flow of energy through it—a carbon-based system operating in a water-based medium, with higher forms metabolizing oxygen. As listed in the glossary of this book, the first part of this definition (up to the dash) is applicable to galaxies, stars, and planets, as well as to life. Only the second part of life's definition is peculiar to living systems as we know them. This lengthy definition, not entirely unreasonable given the difficulties noted earlier while portraying life, has admittedly been contrived in order to diagnose all ordered systems, once again "on that same page." Carefully crafted definitions can help to unify the sciences, admittedly a chief agenda of this book.

As humans, we maintain a reasonably comfortable steady-state posture by feeding off our surrounding energy sources, mainly plants and animals. We stress the "steady state" since, as noted for any open system, by regulating the rate of incoming energy and outgoing wastes, we can achieve a kind of stability—at least in the sense that while alive, we remain out of equilibrium by roughly a constant amount. In a paradoxical juxtaposition of terms, we might therefore describe ourselves as "dynamic steady states." Unfortunately, we waste much of the incoming energy while radiating heat into the environment; warm-blooded lifeforms are generally warmer than their surrounding air. Yet the emitted energy accords with thermodynamics' second law, as Nature has its rules. By contrast, some of the incoming energy can power useful work, thereby helping to maintain order in our lives and bodies. Once this energy flow ceases, the dynamic steady state is abandoned, and we drift toward the more common, "static" steady state known as death, where, following complete decay, our bodies reach a true equilibrium. Stated more pointedly: once we stop eating, we die.

Here's what happens in the food chain consisting of grass, grasshoppers, frogs, trout, and humans. According to the second law of thermodynamics, some available energy is converted to unavailable energy at each stage of the food chain, thus causing greater disorder in the environment. At each step of the process, when the grasshopper eats the grass, the frog eats the grasshopper, the trout eats the frog, and so on, useful energy is lost. The numbers of each species needed for the next higher species to continue decreasing entropy are staggering: to support one human being for a year requires some three hundred trout. These trout, in turn, must consume ninety thousand frogs, which yet in turn devour twenty-seven million grasshoppers, which live off some thousand tons of grass. Thus, for a single human adult to remain "ordered"—that is, to stay alive—over the course of a single year, each of us needs the energy equivalent of tens of millions of grasshoppers or a thousand tons of grass.

Humans, then, maintain order in our bodies only at the expense of an increasingly disordered environment. Every living thing, in fact, takes a toll on the environment. The only reason that the environment doesn't decay to an equilibrium state is that the Sun continues to shine daily. Our whole biosphere is a nonequilibrium system that is subject to solar heating, thus engendering much environmental energy. Earth's thin outer skin is thereby enriched, permitting us and other organisms to go about the business of living.

It's worth pursuing this point a bit further. Suppose Earth's atmosphere and outer space were to achieve thermal equilibrium. All energy flow into and out of Earth would cease, causing all thermodynamic events on our planet to decay within surprisingly short periods of time. A rough estimate shows that the reservoir of Earth atmospheric thermal energy would deplete within a few months, the latent heat bound in our planet's oceans would dissipate in a couple of weeks, and any mechanical energy (such as atmospheric circulation and weather events) would damp in a few days. So be sure to place Earth's energy budget into perspective; neither our planet's primary source of energy nor its ultimate sink are located on Earth.

Not only is life, at any given moment, a reservoir of order, but evolution itself also seems to foster the emergence of greater amounts of order from disorder. As we shall see next in the Biological Epoch, each succeeding (and successful) species becomes more complex and thus better equipped to capture and utilize available energy. The higher the species in a lineage, the greater the energy density fluxing through that species and the greater the disorder created in the Earth-Sun environment. Alas, our principal source of available energy, the Sun, is itself running down as it "pollutes" interplanetary space with increasing entropy.

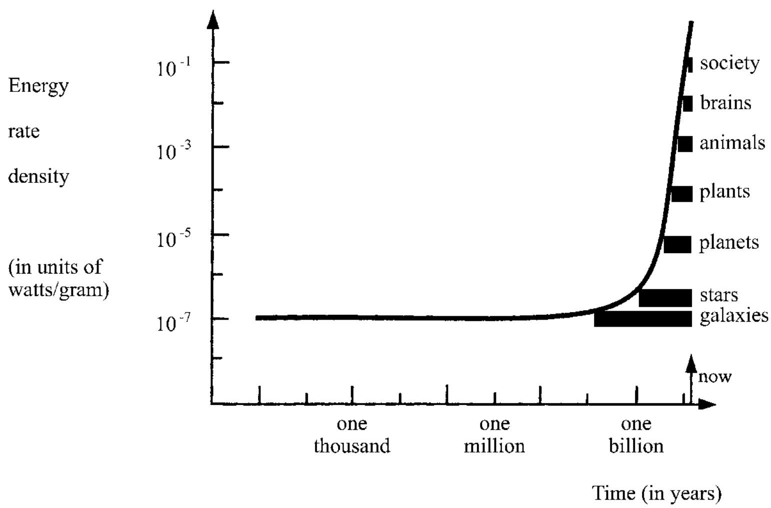
So use caution while regarding evolution as progress. Evolution fosters evermore complex islands of order at the expense of even greater seas of disorder elsewhere in the Solar System as well as in the Universe beyond.

How valid is this reasoning? Can we actually examine all ordered systems—both living and nonliving—on that single page, or level playing field, so to speak? Indeed we can, and what's more, most such systems do display increased complexity along the arrow of time. The flow of energy into and out of open, nonequilibrium systems is an integral part of the cosmic-evolutionary story, yet here we shall only sketch the main results. Suffice it to say that the concept of energy itself is a powerful unifying factor in science; energy may well be the most universal currency in all of Nature. (A more technical, detailed, and quantitative treatment of the energetics of ordered systems can be found in another of my recent books on this subject, *Cosmic Evolution: The Rise of Complexity in Nature.*)

Let's briefly evaluate the energy budgets of several ordered, structured systems experiencing physical, biological, and cultural evolution—namely, galaxies, stars, planets, plants, animals, brains, and society. For the first few of these, energy derives from the gravitational conversion of matter into heat, light, and other types of radiation, much as discussed in earlier epochs. For plants, animals, and other lifeforms—such as those in this and subsequent epochs—energy derives from our parent star, the Sun. And for social systems, too, energy flow is the key driver that powers the daily work needed to run our modern, technological civilization.

Actually, we should concern ourselves less with absolute amounts of energy than with *changes* in energy—and especially with *changes* in the *density* of energy. After all, a galaxy has much more energy than any cell, but of course galaxies also have vastly larger sizes and masses. Rather, it's the energy density that best marks the degree of order or complexity in any system, just as it was radiation-energy density and matter-energy density that described events in the earlier Universe. Of even greater import is the *rate* at which energy transits a complex system of given mass. In this way—called normalization—all systems can be compared along a fair and even spectrum. The appropriate term, "energy rate density," is familiar to astronomers as the light-to-mass ratio, to physicists as the power density, to geologists as the radiant flux, and to biologists as the metabolic rate. All scientists, each in his own specialty, label this term differently, yet all recognize its importance. Now, with today's avowed intellectual agenda to unify the natural sciences, energy rate density usefully links many disciplines, and its physical meaning is clear: the amount of energy passing through a given system per unit time per unit mass.

Take stars, for instance, in particular an average star such as the Sun. Astronomers know its luminosity and its mass, so it's easy to compute its energy rate density. Again, this is energy flowing through the star, as gravitational potential energy during the act of star formation changes into radiation released as the mature star shines. Such a star utilizes high-grade energy during nuclear fusion to produce greater organization, but only at the expense of its surrounding environment, for the star also emits low-grade light, which, by comparison, is a highly disorganized entity. However, even this is a relative statement: what is termed here "low-grade," disordering sunlight will, when reaching Earth, become a high-grade ordering form of energy compared to the even-lower-grade (infrared) energy later emitted by Earth.



. . . most such structures do display increased complexity along the arrow of time.

Furthermore, as stars evolve, their complexity grows. Cosmic expansion is not the only source of structural order in the Universe. On local scales, the evolution of gravitationally bound systems—which is what stars are—can also generate information. As described in the Stellar Epoch, stars are known to originate from dense pockets of gas and dust within chemically and thermally homogeneous galactic clouds. Initially a young star has only a relatively small temperature gradient from core to surface, and it is normally composed of a nearly uniform mixture of ninety percent hydrogen and ten percent helium, often peppered with trace amounts of heavier elements. As the star evolves, its core grows progressively hotter while adjusting its size like a cosmic thermostat, all while nuclear fusion changes its lightweight nuclei into heavier types. With time, such an object grows thermally and chemically inhomogeneous, as the heat and heavy nuclei build up near the core. The result is an aged star that has gradually become more ordered and less equilibrated—indeed one for which more information is needed to describe it, since a complete description of a thermally and chemically differentiated system requires more data than an equally complete description of its simpler, initially homogeneous state.

Planets are more complex than typical stars (or galaxies), hence they tend to have larger normalized energy flows. For example, the energy rate density that drives Earth's climasphere—the most impressively ordered inanimate system at the surface of our planet today—is roughly a hundred times that of a typical star or galaxy. (The climasphere comprises the lower atmosphere and upper ocean, which together guide meteorological climate change capable of mechanically circulating air, water, wind, and waves.)

Living systems need even larger energy densities, not surprisingly since any form of life is clearly more ordered than any nonliving system. Photosynthesizing plants use nearly a thousand times the energy of a star, and human beings consume a daily food ration some twenty times more than that—provided, we repeat, those energy flows are normalized to each system's mass. Although the total energy flowing through a star is hugely larger than that through a human body, the energy rate *density* is much larger for the latter—a fact surprising though true when comparing ourselves with stars.

In turn, our brains utilize nearly ten times more energy yet again—all told, more than a hundred thousand times the rate of a star, "pound for pound." Such a high metabolism for our human heads, mostly to maintain the electrical actions of countless neurons, testifies to the disproportionate amount of worth Nature has invested in cerebral affairs. Occupying two percent of our body's mass yet using nearly twenty percent of its energy intake, our cranium—the most exquisite clump of matter in the known Universe—is striking evidence of the superiority, in evolutionary terms, of brain over brawn.

And currently topping the complexity spectrum, civilization en masse—an open system of all humanity, forming modern society going about its daily, energy-rich business—displays nearly a million times more energy rate density than a star or galaxy. The global ensemble of more than six billion inhabitants, despite its many sociopolitical ills, works collectively to fuel and operate our modern technological culture as an open, elaborate, social system. A marvelous example of the whole equaling more than the sum of its many parts, a group of brainy organisms functioning together is more complex than all of its individuals combined.

This does not imply, by any means, that one type of ordered system evolves into another. Stars do not evolve into planets, any more than planets evolve into life, or animals per se into brains. Rather, new and more complex structures occasionally originated over time as energy flows became more prevalent yet more localized. Galaxies gave rise to the environments suited for the birth of stars, some stars spawned environments conducive to the formation of planets, and at least one planet fostered an environment ripe for the origin of life.

Thus, complexity grows as more intricately ordered systems emerge, in turn, throughout natural history, much as their surrounding environments are inevitably ravaged with rising entropy. The products of evolution are fleeting, the evolutionary process itself messy and undirectional. When energy flows cease, systems decay back to equilibrium from whence they came: when the Sun stops shining it's no longer a star; when the biosphere ends plants die; when humans starve we perish; all return their elements to Nature. Remarkably, the two—complexity locally and entropy globally—can both increase, and simultaneously so. The temporarily ordered systems that so impressively embody the diverse actors in the cosmic-evolutionary story are wholly consistent with, indeed best understood by, the most cherished of all physical laws—namely, the thermodynamic principles that undergird the arrow of time itself.

∞

Theoretical insights and experimental simulations suggest that life is a logical result of known chemical principles operating within the atomic and molecular realm. No new science is likely needed to understand life's complexity, indeed life's rise in complexity with evolution, provided we are willing to embrace the concept of nonequilibrium thermodynamics of open systems. When reduced to essentials, life is not much different, apart from its degree of complexity, than galaxies, stars, or planets. All these structured systems, including life itself, increasingly order themselves by absorbing energy from, and emitting entropy into, their surrounding environments. Time and energy are clearly key parts of Nature writ large. "Follow the energy" is as important a dictum as any in complexity science.

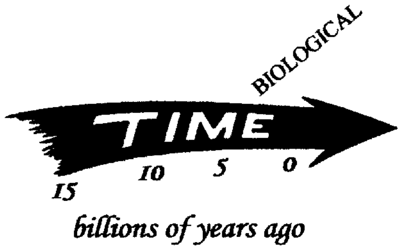
The cosmic-evolutionary story harks back not merely to the start of life but virtually the beginning of time. And it includes the scientific events that ordered the many complex systems—both living and nonliving—now acting as the main characters in this story. Heraclitus of old, who basically had the right idea in his ancient philosophy that "all things change," would nonetheless be amazed at the rich detail amassed to bolster today's new scientific philosophy.

The specifics of life's origin, however—thought to be a natural consequence of atoms and molecules interacting in an energy-rich environment—are not yet in hand, partly because the hard evidence was wiped clean billions of years ago and partly because laboratory experiments have yet to fashion chemicals more sophisticated than life's precursors. In particular, a sizeable gulf separates the early evolution of life's building blocks from the later evolution of the first living cells. This Chemical Epoch has sketched a scientific consensus concerning that blurred realm where chemical evolution ends and biological evolution begins.

What is clear is that the thermal gradients needed to foster energy flows in Earth's biosphere could not be maintained without the Sun converting gravitational and nuclear energies into radiation that emanates outward into unsaturable space. Were outer space ever to become inundated with radiation, all temperature gradients would necessarily vanish, and life, among many other ordered structures on and beyond our planet, would cease to exist. That space will never become saturated owes to the expansion of the Universe, bolstering the idea that the dynamical evolution of the cosmos is an essential prerequisite for the order and maintenance of all things—including, at least on Earth, the origin and evolution of life itself. All the more reason to include life within our cosmic-evolutionary worldview, for the observer in the small and the Universe in the large are not disconnected.

6. BIOLOGICAL EPOCH

Complexity Sustained



OUR COSMIC-EVOLUTIONARY SCENARIO is really taking shape now. From stellar atoms to planetary molecules, we have explored plausible ways that galaxies, stars, planets, and life can be surveyed, in turn, along a single range of flowing energy and rising complexity. Indeed, the origin of life seems to be a natural consequence of the evolution of matter, and, further in turn, the evolution of that life, a natural process of yet more change with time.

To grasp the entire spectacle of life—from past to present, from aardvark to zucchini—we must inquire beyond specialized, reductionistic analyses of simple matter. Single cells are sufficient to display the operational difference between life and nonlife, but these are not enough to illuminate the full expanse of all life on Earth. To appreciate life's true complexity—including its structure, function, and diversity—we need to examine whole organisms, generally and often in context of wider populations. For the same reason that no one could possibly understand the inner workings of an automobile by grinding it up and chemically measuring its basic atoms and molecules, holistic probes of living things usefully complement microscopic views of their component cells. And if microscopic studies of life-forms mainly entail the realm of chemistry, as previously in the Chemical Epoch, then their macroscopic studies fall squarely into the realm of biology, hence the arrival of the Biological Epoch, at least on Earth.

Here, we encounter the host of plants and animals on our planet. Roses and reindeer, tulips and turtles, evergreens and elephants, and uncounted more species. Where did they all come from? That they suddenly appeared intact from nothing is an interesting idea, but spontaneous or miraculous creation makes no sense scientifically. Nor is there a shred of objective evidence to support it.

Together, genes and fossils chronicle an amazing story of life on Earth. Biochemists who amass digital genomes are now pooling their talents with paleontologists who scour fossilized bones. The results provide increasingly robust details of that story, regularly revealing torn and tattered pages here, occasionally uncovering whole new chapters there. Repeatedly, throughout the millions of millennia, life-forms emerged while others perished. Some species survived for ages; others succumbed as soon as they appeared. Incredibly, more than ninety-nine percent of all life-forms that once prospered are now extinct—victims of the strides of time.

Only one factor has seemingly remained constant and unchanging throughout the eons of Earth's deep history: change itself. The phenomenon of change really does seem to have been the hallmark in the origin, development, and fate of all structures, living or nonliving.

∞

What were the first living cells like? Scientists don't know for sure, for we lack good data from the first two billion years of Earth's history—a period known as the Archaean. Most likely they were tentative entities—microbes fragile enough to be destroyed by strong bursts of energy yet sturdy enough to reproduce, thereby giving rise to generations of descendants.

One thing is certain: the first cells, often called primitive heterotrophs (for they needed outside, organic sources for nourishment), somehow had to find enough energy to continue living and organizing themselves. They presumably did so while floating on or near the ocean surface, absorbing the acid and base molecules in the rich broth of the early organic ocean. This extraction of energy via the capture and chemical breakdown of small molecules—called fermentation—is still employed on Earth today by unicellular microbes (mostly yeast) in beer casks while changing grain into alcohol, in bread dough when exposing starch to limited amounts of water, and in commercial methods to improve the flavor of tea, tobacco, and cheese. But the primitive heterotrophs could not have fed indefinitely on the organic matter from which they originated. After all, the continual passage of time brought irreversible changes in the environment.

As Earth cooled, several of the energy sources capable of producing the acids and bases began to diminish. Geologic and atmospheric activity declined, and as gases thickened the air, smaller amounts of solar ultraviolet radiation reached Earth's surface. Laboratory experiments show that these changing conditions are not conducive to the continued production of the heterotrophs' food supply, which is why we don't see a thick film of organic acids and bases floating on today's oceans and rivers.

Whereas originally Earth's waters had plenty of juicy organic molecules on which the heterotrophs could feed, the denser atmosphere and weakened tectonics meant fewer food sources over time. Consumed more rapidly than it was replenished, the organic soup gradually thinned, creating a crisis for the multiplying cells. Those primitive cells then had to compete with one another while scrounging for the decreasing supply of nourishing acids and bases. Eventually, the heterotrophs devoured every bit of organic matter floating in the ocean. The organic production of acids and bases via lightning, volcanoes, or solar radiation simply couldn't satisfy the voracious appetite of the growing population of heterotrophs.

This scarcity of molecular food was a near-fatal flaw in life's early development. Had nothing changed, Earth's simplest life-forms would have proceeded toward an evolutionary dead end—starvation. Earth would be a barren, lifeless rock, and our story aborted. Fortunately, something did change. It had to; nothing fails to change. And one change that did occur enabled the story to continue—not by some design and not solely by chance, rather more likely by the usual mixture of chance and necessity operating over long durations. At least partly, successful evolution is often a case of being at the right place and the right time.

Other cells—the forerunners of plants, called autotrophs (for they were self-nourishing)—invented a new way to get energy, thereby conceiving a unique opportunity for living. (Some researchers claim that the first cells were likely already autotrophic, acquiring energy directly from the environment and skipping altogether the heterotrophic stage.) This novel biological technique employed carbon dioxide, the major waste product of the fermentation process. While the earliest cells were busily eating organic molecules in the sea and thus polluting the atmosphere, more advanced cells were learning to use these pollutants to extract energy. In this case, the energy was not derived from the consumed gas but from another well-known source—the Sun. This newly invented process is photosynthesis, perhaps the greatest single metabolic invention in history.

The key here is the chlorophyll molecule, a green pigment having its atoms arranged so that light, when striking the surface of a plant, is captured within the molecule. Advanced cells containing chlorophyll thereby extract energy from ordinary, gentle sunlight (not harsh ultraviolet radiation) by means of a chemical reaction that exploits that sunlight to convert carbon dioxide and water into oxygen and carbohydrates. The oxygen gas escapes into the atmosphere, while the synthesized carbohydrate (sugar) is used for food. This, then, is another way a cell can "eat," or extract energy from its environment—hence its name: *photo*, meaning "light"; *synthesis*, "putting together."

How did some protoplant, microbial cells develop photosynthesis? To be sure, primitive bacteria invented photosynthesis, not plants per se, which emerged much later. But biologists are again uncertain how they actually did it, other than presuming that random events first altered the DNA molecules in some early cells, which then determinedly sucked up the needed solar energy to survive. They no longer had to compete for the organic acids and bases in the primal ocean. They were selected by Nature to endure because they adapted to the changing environment. And with photosynthesis came a big advantage since the new cells could persist on merely inorganic matter. The autotrophs were clearly more fitted for survival during what was probably the first ecological crisis on our planet.

Photosynthesis freed the early life-forms from total dependence on the diminishing supply of organic molecules in the oceanic broth. Fermentation by heterotrophs was no longer needed for survival. Early cells able to utilize sunlight overspread the watery Earth. In time—much time—the autotrophs changed into not only many types of bacteria but also all the varied types of plants now strewn across the face of our planet.

The photosynthetic process continues to this day as plants routinely use sunlight to produce carbohydrates as food (for both metabolic function, as well as cellulose structure). The plants, in turn, release oxygen gas that animals, including ourselves, breathe. Photosynthesis is, in fact, the most frequent chemical reaction on Earth. In round numbers, each day about four hundred million tons (about a trillion kilograms) of carbon dioxide mix with some two hundred million tons of water to make about three hundred million tons of organic matter and another three hundred million tons of oxygen gas. Yet despite these large numbers, it's still the small but abundant stuff that does much of it: fully half of today's global photosynthesis and oxygen production is accomplished by single-celled marine plankton living in the top oceanic layer where enough light penetrates to support their growth.

By loss of their food source, the ancient and primitive heterotrophs were naturally selected to die. The better adapted autotrophs were naturally favored to live. Life on Earth was on its way toward using a primary and plentiful source of energy—that of our parent star—in a reasonably efficient and direct manner. It all began not quite three billion years ago.

Photosynthesis over eons of time is, by the way, partly responsible for the fossil fuels. Dead, rotted plants, buried and squeezed below layers of dirt and rock, have chemically changed over megacenturies into oil, coal, and natural gas. Such fossil fuels, with their vast quantities of solar energy trapped in carbohydrates, have made industrial civilization possible. But those fuels are virtually nonrenewable, at least over time scales shorter than tens of millions of years. Billions of years of energy deposits in rotted organisms will be depleted shortly—oil and gas in the twenty-first century and coal not more than a few hundred years thereafter. Once again, things will have to change, just as they've changed in the past.

The use of sunlight by cells was a double achievement of great importance for life on Earth. Not only did the Sun provide an unlimited source of energy and assure a dependable supply of food, but it also drastically changed Earth's atmosphere by helping to generate oxygen gas. Oxygen became another pollutant of the early air, an inevitable result of autotrophs' photosynthesizing, much as the heterotrophs had soiled the primordial air even earlier with carbon dioxide gas. No anerobic organism could have escaped this "oxygen holocaust."

Atmospheric change has had an enormous influence on the abundance and diversity of life on Earth. The photosynthetic release of oxygen into an atmosphere that previously had little or none of it ensured great changes not only in the environment but also among life-forms dependent on that environment. Interacting with the Sun's ultraviolet radiation, the diatomic oxygen molecule breaks down into two oxygen atoms. Three oxygen atoms then recombine high in the atmosphere, molding large quantities of triatomic oxygen, or ozone. (Derived from the Greek and meaning "to smell," that pungent ozone gas can often be sensed near thermal copying, or Xerox, machines that use ultraviolet radiation.) Ozone now completely surrounds our planet in a thin shell at an altitude of about fifty kilometers (thirty miles), effectively shielding the surface from further exposure to harmful high-energy radiation.

As the ozone layer matured sometime in the past, survival was no longer dependent on protection by a layer of water or by some rock or other object acting as a barrier against what must have earlier been a truly hellish world. Life became possible on the surface of the water and eventually on the surface of the land. Organisms were on their way toward spreading at will, populating nearly every available nook and cranny on planet Earth. In short, life could invade areas where no life had existed before.

None of this happened overnight. The ozone layer needed time to thicken enough to screen out most of the harmful ultraviolet radiation. The process was an accelerating one: Oxygen-producing autotrophs had an increased chance for survival and therefore replication. The more offspring they produced, the more oxygen they dumped into the atmosphere. And more oxygen meant more ozone, more protection from solar radiation, and enhanced opportunities for survival. But it still took time for the protective ozone to cumulate. Perhaps as many as a couple of billion years after the onset of photosynthesis were needed, since dissolved iron in the oceans would have combined with any free oxygen, removing it from the atmosphere until waters everywhere were saturated.

Models of Earth's early atmosphere imply that the ozone layer started to form, or at least oxygen gas had initially begun to rise, somewhat more than two billion years ago. Deposits of oxidized iron (called "red-bed" sediments or banded-iron formations) in the geological record of that date, now mined for their metal to make steel, support the view that oxygen was then hardly one percent of Earth's air, well below the twenty percent we enjoy today. Some of the most ancient fossils, dating back earlier than this, as noted shortly, do show evidence for chlorophyll products, suggesting that oxygen was then being released into the atmosphere—but to what extent is unknown. Other models imply that oxygen didn't reach its current levels and nor did the ozone layer become a fully effective shield of solar radiation until about a half-billion years ago. Fossil evidence also supports that argument, as life rather suddenly became varied and widespread around six hundred million years ago, before which only primitive life-forms existed. Shortly thereafter, a rapid surge in numbers and diversity of complex living organisms came forth—a population explosion of the first magnitude.

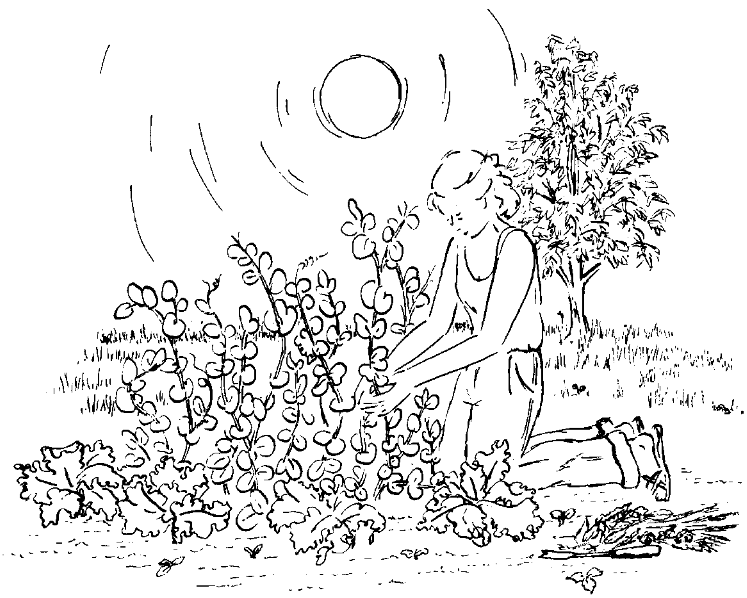
What was responsible for this burst of biological activity? The buildup of oxygen may well have been the main reason, permitting a new, more efficient way for organisms to obtain energy for living. The first, most primitive life-forms that ate via fermentation were superseded by advanced creatures that developed photosynthesis as a means of manufacturing food. Eventually, even more advanced organisms—the forerunners of the animals—began exploiting oxygen as their primary source of nourishment. By using oxygen, organisms could then obtain more energy from the same amount of food. Combined with a protective ozone shell, this global availability of oxygen meant that life was able to survive and reproduce in all sorts of new habitats.

The previously harsh conditions under which early life had struggled were gone. Earth became a relatively comfortable place in which to live. And the vanguard organisms of the time took advantage of their friendlier environment. Localized complexity was about to rise dramatically.

Since photosynthesis makes oxygen, it's likely that some other process uses it. Much of Nature is symbiotic, just as plants and animals have an interlinked relationship today. That other process is respiration, a chemical reaction whereby cells employ oxygen to release energy. Ingesting oxygen ("breathing") helps an organism to digest the carbohydrates in its body, the waste products being carbon dioxide and water. Respiration, then, is just the reverse of photosynthesis, but there is an important difference. Whereas in photosynthesis energy must be absorbed to yield the foodstuff carbohydrates, in respiration much of that energy is released as the oxygen destroys the chemical bonds of those same carbohydrates. The "burning" of food by oxidation supplies a concentrated source of energy, a beneficial trait for animals, which increasingly demanded larger flows of energy with the march of evolutionary time.

So does that make us a modern version of the ancient heterotrophs? In a way it does, yet we are much more efficient than primitive life. The rise of oxygen in Earth's atmosphere eventually permitted some life-forms to extract through respiration nearly twenty times more energy from the sugars they use as food than do the simplest life-forms via fermentation in the absence of oxygen. We humans, as all animals, are the beneficiaries of this age-old advance toward what has become the highest form of biological energy retrieval.

Today, these two actions—mainly plant photosynthesis and animal respiration—direct the flow of energy and raw materials throughout Earth's biosphere. This energy for life is unidirectional: it flows only one way. The energy originates with the Sun, is captured in photosynthesis, is released by respiration, and is consumed in the course of living. All the while, carbon dioxide, water, and oxygen are continually exchanged. These materials are used repeatedly, in a completely cyclical fashion: the plants use animal pollution, while the animals use plant pollution. Nature knows how to recycle.



. . . plant photosynthesis and animal respiration direct the flow of energy and raw materials throughout Earth's biosphere.

∞

By what means do scientists know about the previous episodes of life on Earth? Never mind the precise details: How can we sketch even a broad outline of ancient events so far back in time? Part of the answer is that we are guided by many clues preserved in the old rocks of our planet. Most of these hints are the remains of living organisms.

Life-forms usually begin to decay as soon as they die. Once a means of gathering energy has ended, disorder sets in. Entropy takes its toll again, inevitably increasing according to the laws of thermodynamics. Dead organisms decompose quickly—even their bones eventually—the former proteins becoming rancid waste within a few days. This is the way in which all life-forms, including humans, return to the planet the elements they borrowed from it.

Some special environments can limit decay, including cold polar regions, high mountain tops, and deep ocean trenches. Both low temperatures and water burial serve to retard spoilage, so living systems that perish along a stream or ocean shore might end up sandwiched under layers of sand and sediment settling down through the water. Volcanic lava is another place where life-forms can be buried yet preserved, in this case under mounds of ash. In time, the sedimentary deposits of sand or lava become hardened into rock, entombing the remains. Thus, though the fleshy parts of ancient organisms usually decay, their bony structure is sometimes preserved until later uncovered by natural causes (such as geological upheavals) or planned events (archaeological expeditions). These rare remains, or fossils, are the visible traces of dead organisms that once lived. Even boneless bacteria occasionally leave behind microfossils so small that a microscope is needed for analysis.

Exhaustive study of fossils thus far unearthed has enabled paleontologists to assemble a partial record of earlier life. These are the people who work at the interface of biology, chemistry, and geology, digging around in the dirt and rocks while seeking fossilized remnants of extinguished life. Using a variety of dating techniques, they can roughly determine when various organisms lived and sometimes how they died. More tellingly, the fossil record shows how, through eons of time, whole new life-forms appeared while others disappeared. Some types of life survived for lengthy durations; others seem to have gone extinct nearly as soon as they arose.

As a rule of thumb, we can generally say that the oldest rocks embed only simple life, whereas young rocks contain much more complex life. That's less a fanciful idea and more an empirically proven fact. The fossil record of biological specimens documents a clear and unmistakable trend over the ages. That trend, much as for galaxies, stars, and planets, is one of increasing complexity.

All the fossils, taken together, chronicle a rich and varied natural history of life on Earth. The oldest fossils usually have a spherical, cellular structure resembling that of modern blue-green algae (now called cyanobacteria)—fuzzy, bacterial scum found at the edge of today's lakes, streams, and backyard swimming pools. Not overly complex, these carbon-rich microfossils lack well-developed biological nuclei and display a plain, austere morphology only a few microns across; some of them are filamentary, as though the cells were loosely grouped on long chains. Yet, the life-forms—microbes technically termed prokaryotes—of which the fossils are the remains, are controversial. They could conceivably be minute carbon deposits formed by the action of scalding water on minerals trapped in the rock and not life at all. However, and although recent contamination cannot be completely ruled out, the consensus interpretation among experts is that these findings are indeed old fossils of true life, possibly the remains of autotrophs that reproduced asexually by splitting in two. Some of the fossils do seem to show suggestive evidence of cells caught in the act of simple, binary division.

Many of the best fossils of the most primitive, Archaean cells were discovered only within the past decade. Found embedded in South African and especially Western Australian rocks known to be three-and-a-half billion years old, these oldest cells are presumed to have lived that long ago. In truth, paleobiologists have no way of dating the fossils themselves; nor are these oldest remains fossils per se. Rather, they are fossil imprints of carbon deposits that were once bacteria, and radioactive techniques are useless for dating carbon older than forty thousand years. But it seems unlikely that latter-day algae could have gotten so deeply encased in such old rocks.

No fossils of the most ancient heterotrophs have ever been found. Tiny bits of elemental carbon (in the form of graphite specks) were recently discovered trapped in nearly four-billion-year-old Akilia rocks of western Greenland, yet these oldest of all Earth rocks have been so heavily metamorphosed at high temperatures and pressures that most original organic information is lost and the claim that life's origin dates back that far is weak and equivocal. Molecules much smaller than cells can indeed seep into all but the densest of rocks, so there's no good way of knowing if amino acids and nucleotide bases encased in rocks are genuine biosignatures as old as the rocks themselves. Even if paleontologists had methods to search for prebiotic organic matter, none would likely be found. The early heterotrophs probably devoured every bit of available organic matter, thus leaving no trace of the primordial soup anywhere on Earth; they simply ate the evidence. Consequently, scientists are unable to estimate either the amount of time needed for the autotrophs to have overwhelmed the primitive heterotrophs, or for those heterotrophs to have appeared in the first place. As best we can tell thus far, life probably originated not more than a billion years after the formation of planet Earth. It could have conceivably emerged earlier, but how much earlier is currently only a guess.

Evidence for more recent, though still ancient, life has been unearthed in numerous places on our planet. The north side of Lake Superior in Ontario, for example, is especially rich in ancient fossils, and the rocks there are radioactively dated to be some two billion years old. No reputable scientist doubts this evidence for life. What's more, these old life-forms must have photosynthesized by some means or another, since chlorophyll is often found in their immediate vicinity.

Embedded within this old Canadian limestone are whole communities of cells called stromatolites—layered colonies of bacterial microbes in columns up to a half-meter tall, created when primitive autotrophs clustered together and became trapped in sediment that later hardened into rock. (Living evidence for underwater stromatolites, closely similar in size and form to their ancient forebears and now providing the base of the ocean's food chain, can be found today in shallow seas like Exuma Sound, Bahamas, and Shark Bay, Australia, where the water is too salty for plants and animals to graze on them.) Careful study of this two-billion-year-old rock reveals at least a dozen distinctly different types of cyanobacteria, all much simpler than modern cells. Examined closely through a microscope, these ancient cells lack well-developed biological nuclei and thus were still prokaryotic. And despite their obvious clustering, each of these ancient cells functioned on its own—as surely did even less well preserved stromatolites that are long dead, buried, and crushed into flattened mats within three-billion-year-old Australian rock. All fossilized cells older than a billion years appear unicellular, as each presumably failed to collaborate with other cells nearby. The reason might have been poor nutrition, ironically brought on by rising oxygen levels in the air.

One-to-two-billion-year-old rock formations scattered across the globe often contain surprisingly well-preserved remains of autotrophs.

Evidence of ancient life.

These half-meter tall, club-shaped stromatolites are currently alive in shallow water in Shark Bay, Australia. Here, and at very few other places, the water is too salty for the animals that normally consume them. In size and form, modern stromatolites resemble their fossilized ancestors that date back at least two, and probably three, billion years. They represent the oldest known groups of prokaryotic unicells—not true multicells, but clusters of simple life that began the trek along the route to greater complexity. *Source: Australian Geological Society.*

Microfossils of at least twenty different types have been identified thus far in one Australian outcrop alone, many similar in structure to modern blue-green algae. More importantly, fossils toward the end of this period record the appearance of the first true organisms—groups of interacting cells, the ancestors of modern plants and animals. In short, sometime around a billion years ago, life had reached a whole new plateau. It was on the road toward increased organization.

Organization represents a distinct advance in complexity. By cellular organization, biologists generally mean that cells are communicating information, sharing resources, and working together as a team. Like inanimate stars and galaxies, the earlier microbial cells of stromatolites had clustered, but none of them displayed teamwork or communicated interactively. Biological organization ramps up the degree of ordered complexity considerably once collaboration begins among cells.

How did they do it? How did cells learn to communicate, presumably for the common good? Apparently, some cells managed to evolve beyond the prokaryotic stage, thereby becoming eukaryotes—sexually reproducing life-forms with genuine biological nuclei, including hereditary DNA molecules. Termed “symbiosis” to denote a mutually beneficial union of two dissimilar organisms, this process first jelled when an energy-poor, bacterial cell floating in the ocean swallowed another prokaryotic bacterium that had a talent for making the fuel-rich molecule adenosine triphosphate, or ATP for short. The cell soon realized the benefits of an in-house energy factory and kept the bug as a permanent resident.

The oldest known fossils of unicellular eukaryotes (also called protists) date back nearly two billion years, and indirect chemical evidence implies that they might have existed a billion years before that. It’s probably not a coincidence that the first eukaryote—the common ancestor of all plants and animals—appeared at the time when free oxygen was on the rise in Earth’s atmosphere. Nearly all eukaryotes need oxygen to live, whereas most bacteria find it lethal. Even so, some bacteria managed to develop strategies to survive in an oxidizing atmosphere, mostly by taking refuge as parasites.

Today, deep inside each human (eukaryotic) cell, we do see hundreds of minute structures (“mitochondria”) that are widely thought to be descendants of that early (prokaryotic) bacterium. Colonies of bacteria have become indispensable passengers within the cells of every plant and animal. Living both on and in every person’s body are more bacteria than there are human beings on Earth; the millions of microbes in our gut aid digestion and produce essential vitamins. The bacteria are metabolic wizards that energize higher forms of life, their ATP enabling a variety of crucial functions such as muscle contraction needed for motion and protein construction needed to make more cells. These bacterial groups are the mitochondria that literally power most cellular activity by burning (via respiration) the food we eat. Cells acquire their energy by using a protein to ferry the ATP through the membrane of the mitochondria into the cells’ cytoplasm, a little like the nozzle at a gas-station pump that acts as an intermediary to transport fuel from the pump to a car. Symbiosis was one of the great inventions of biological evolution—some say the most important event in the history of life (apart from its origin). At the least, symbiosis ensured bacteria’s future as vital energy brokers for nucleated cells. Yet ironically, it was also an advance that ended the microbes’ independent dominance of life on Earth.

How did sex overtake asexual reproduction some billion or more years ago, and how has it thrived ever since among many unicellular and nearly all multicellular organisms? Does sex have an advantage in the daily struggle for existence? All things being equal, asexual reproduction (involving only gene separation and division) ought to prevail on grounds of simplicity, the extra effort of having to find a mate limiting the success of sex (involving additional gene shuffling and mating)—yet sex reigns across much of the tree of life. Most researchers argue that sex greatly accelerates the rate of biological evolution by promoting genetic variety—and that alone might have been enough for Nature to embrace sex, once the unicellular eukaryotes figured out how to do it. But, with supporting evidence skimpy, perhaps it's only wishful thinking that sex is good for the birds, bees, and us. Ideas abound, most having to do with either sex assembling beneficial mutations for the common good or sex purging harmful mutations from the parental genome. The former would surely enhance the diversity of life, yet the latter would perform essential evolutionary editing lest deleterious mutations accumulate in individuals and in populations. Whichever, experiments are now underway in which biologists raise populations of organisms ranging from water fleas to mud worms, thereby testing lineages' fitness despite imposed mutations. Thus far, the results are mixed, quite possibly because both mechanisms are at work—sex collects good mutations *and* rejects bad ones—another case of "gray compromise" so often orchestrated by Nature in the wild.

By one billion years ago, unicellular life had already existed on Earth for well more than two billion years. Its basic cells had become ten times larger, vastly more sophisticated, and perhaps more diverse functionally. Furthermore, fossilized cells of this age show clear and widespread evidence for eukaryotes, which have much different structures from single-cell bacteria. More advanced life was springing from simple life, though the two were coexisting. Equally important, some of the billion-year-old organisms had discovered ways to enhance their survivability by working together as groups; they had become multicells that did collaborate with one another.

But that's all there was a billion years ago: Primitive oceanic life flourished, though not much else. The fossil record shows no evidence that plants yet adorned Earth's landscape. No animals were crawling, swimming, or flying near the surface. And certainly, by no means were men and women yet even on the evolutionary horizon.

∞

To make another astrobiological connection at the interface of astrophysics and biochemistry, the issue of life beyond Earth looms large—especially the possibility that evidence for Martian fossils might already exist. The prospects for life on other alien worlds, such as the Jovian moons Europa and Titan, are equally titillating. Biologists have recently broadened their view of life, not only in extreme Earth environments such as the hydrothermal vents noted previously in the Chemical Epoch, but also in distinctly new venues elsewhere in our Solar System.

During the past generation or so, several robots have orbited and even landed on Mars, foremost among them those of the 1976 *Viking* project—perhaps NASA's boldest scientific mission to date. Astronomers strongly suspect that Mars currently has no liquid water, diminishing the chances for life there now. But running water and a denser atmosphere (again, to "keep that lid on") in the past could have conceivably fostered conditions suitable for the emergence of life. And there is some visual evidence that water did flow on Mars long ago, including the possibility of meager oceans, presumably before the planet entered its current ice age less than a billion years ago. The *Viking* landers were programmed to perform some simple experiments designed to detect biological activity or at least organic matter, in the hope that some microbial life-forms might have survived to the present day. None was found.

However, the *Viking* experiments searched only for life now living. Perhaps ancient fossils of long-dead Martian life—simple bacterial life possibly enduring prior to the arrival of the numbing cold that likely prohibited sustained life as we know it—might show paleontological evidence for rudimentary life. If a severe ice age had locked Earth into a deep freeze a billion years ago, the only evidence that life arose on our planet would be microscopic remains of fossilized microbes—and we wouldn't be here to ponder it.

Surprisingly, one place to look for Martian fossils is right here on Earth. Planetologists agree that a small fraction of meteorites found on Earth's surface have actually come from the Moon and from Mars. These meteorites were apparently blasted off these bodies long ago during an impact of some sort, thrown into space violently enough to escape their parent bodies, and eventually captured by Earth's gravity, ultimately to fall to the ground. The most fascinating of these rocks are surely a dozen or so from the Red Planet—their trapped gases match exactly those present in Mars's atmosphere—and one of them may harbor fossil evidence for past life.

Based on estimates of the cosmic-ray exposure it received while drifting toward Earth, the meteorite catalogued as ALH84001 was ejected from the Martian surface about sixteen million years ago. The blackened rock itself, some four billion years old and about the size and weight of a grapefruit, was found in 1984 in the Allan Hills of Antarctica, a place where pristine meteorites often just sit atop the icy wastes of the barren, frozen landscape. Upon breaking it open and examining closely its cracks and crevices, scientists could see rounded, brownish "globules" of carbonate matter no larger than the period at the end of this sentence, a little like those made during origin-of-life experiments discussed in the Chemical Epoch. Because carbonates form only in the presence of water, these small globules imply carbon dioxide gas and liquid water near ground level at some time in Mars's history. This matches the inferences drawn earlier from orbital photos of valleys and tributaries apparently carved by water when the Martian climate was wetter and warmer.

Claims that the ALH84001 rock contains traces of primitive Martian life stem mainly from our knowledge of bacteria on Earth. Terrestrial bacteria do produce structures similar to the Martian globules, and they also manufacture iron-rich chemical crystals that the rock displays in tiny, teardrop-shaped crystals embedded in places where the carbonate has dissolved. Furthermore, the rock contains traces of PAHs—chemtalk for a class of messy organic molecules known as polycyclic aromatic hydrocarbons—often found among the decay products of terrestrial plants and other Earth organisms. None of these data would individually indicate life if found on Earth, but all of them collectively make the case stronger for life on Mars. Even so, as is often said in science, extraordinary claims require extraordinary evidence.

A final piece of Martian evidence is the most dramatic—and the most controversial. On very small scales seen only through a powerful microscope, elongated and egg-shaped structures are discernable inside the carbonate globules of ALH84001. And these are what some scientists have taken to be fossils of primitive organisms. Outwardly, photomicrographs reveal curved, wormlike structures clearly resembling bacteria on Earth. But scale is a crucial part of any interpretation. The minute structures are only about half a micron across, or some ten times smaller (hence a thousand times less voluminous) than ancient bacterial cells found fossilized on Earth. Many biologists argue that such minute bags of chemicals are simply too small to have functioned as life as we know it. What's more, the Martian rock contains no evidence of amino acids, cell walls, semi-permeable membranes, or any kind of internal cavities for bodily fluids—all of which properties attend even the oldest and most primitive fossils found on Earth.

Most experts are of the opinion that life has not been found on Mars—not even fossilized life. They maintain that all the meteorite data could be the result of chemical reactions not requiring any kind of biology. Carbonate compounds are common in all areas of chemistry; PAHs are found in many lifeless places (for example, glacial ice, interstellar clouds, even the exhaust fumes from automobiles); bacteria are not needed to produce crystals; and it remains unclear whether the tiny wormlike structures are animal, vegetable, or merely mineral. Contamination is also a potentially huge problem since ALH84001 apparently sat open to the elements on Earth for more than ten thousand years before being picked up by the meteorite hunters. As with most frontiers in science, early pioneering results are not as clear-cut as one might hope.

All that said, should the claim of life on Mars hold up against the weight of healthy skepticism in the scientific community, these findings will go down in history as one of the greatest discoveries of all time: we are—or at least were—not alone in the Universe! Distressingly, if life conceivably did originate on Mars and later arrived on Earth, then we might all be Martians!

When considering the presence of life under adversity, we shouldn't be too quick to rule out environments based solely on extreme properties. The underwater hydrothermal vents noted in the Chemical Epoch are very hostile places as judged by life on or near Earth's surface, yet life manages to thrive around them under conditions quite unlike anything at the surface. Undersea volcanic activity spills forth scalding-hot water rich in sulfur and poor in oxygen that manages to feed "extremophilic" life by a process known as chemosynthesis—an analog of photosynthesis, yet one that operates in total darkness. Here, teeming colonies of organisms survive and prosper at temperatures close to, and sometimes even exceeding, the usual boiling point of water on a diet of hydrogen, carbon dioxide, and elemental sulfur, while exhaling toxic (to surface creatures) hydrogen sulfide. A variety of deep-sea animals resembling clams and worms form symbiotic partnerships with bacteria that get their energy from sulfides rather than light. Despite the decidedly odd conditions and even odder metabolisms, all known extremophiles inhabiting vent environments are still based on the element carbon, just like the rest of us living in the more traditional biosphere in or around Earth's surface.

The volcanically heated springs of Yellowstone National Park are another good example of an exotic site where a wealth of life flourishes under extreme conditions inhospitable by human standards. There, the rich microbial diversity hardly includes garden-variety types, yet, surprisingly, many of the microorganisms' genes approximate many of ours. At the molecular level, these hot little creatures resemble eukaryotes more than bacteria; in fact they differ from conventional bacteria more than do humans from a crab. Though in the public eye microbes are often seen in the context of disease and rot, these heat-loving bugs might be telling us something important about how the earliest life-forms employed inorganic nutrition (in the absence of carbon) and geothermal heat (without the Sun's radiation). Carbon-centered metabolism and solar-driven photosynthesis arose comparatively later.

A decidedly unorthodox category of life—indeed, a newly proposed primary lineage—is the archaeabacteria that are often found in extreme environments once thought devoid of any life. These microorganisms populate only oxygen-free ecosystems, such as on the seafloor, in sewage, or in the hot springs seeping through Earth's crust. The archaea (as they are called for short) stay alive by converting carbon dioxide and hydrogen into methane, which is the main chemical of natural gas (explaining why some of them are called "methanogens"). Such scalding, oxygen-free conditions resemble those thought to have existed on our planet during its first billion years or so, implying that the archaea hark back to times prior to the formation of conventional bacteria. Accordingly, today's archaea might be relatively unchanged descendants of the most ancient class of life on Earth—and as such the best link to that ultrasimple, original, "last common ancestor" from which all forms of life have subsequently evolved.

What we don't know is whether this anerobic, submarine life was the origin of more familiar life that eventually made its way up to the surface or, by contrast, whether the archaea are merely the result of primitive life once on the surface that managed to survive only by diving (and adapting) deep underwater to avoid poisonous oxygen, to develop its own odd metabolism and to seek alternative energy sources.

Underground hot springs and the extremophilic life thriving near them raise the possibility of life-forms with even greater diversity amid even wilder conditions than those known to us on Earth. This is especially true if we broaden our perspective yet more to consider life at the other extreme—cold. Household refrigerators (or at least their freezers) surely retard the growth of bacteria—that's the job of those machines—but life can sometimes still eke out a living there. In fact, the bottoms of perennially frozen lakes in Antarctica harbor entire communities of microbes, despite temperatures nearly equaling the freezing point of water. These are not merely bacteria but also microscopic plants and animals. In a few places, microbial life holds on and avoids death even within the thick, hardened ice itself, surviving in a kind of suspended animation, apparently indefinitely. The frigid, dry, Antarctic climate resembles that of Mars today, bolstering the idea that frozen tundra on the red planet could support life under Spartan conditions. Other inhospitable places on Earth where simple, yet active life has been found include subterranean rock, salt deposits, and even oil fields, all more than a kilometer below Earth's surface.

Discoveries during the past decade have brought much wider appreciation for life on our own planet, revealing bacteria in places where biologists once thought nothing could possibly live. Adaptation is the key, as is so often the case, and the simplest forms of life seem to be surprisingly adaptive to all sorts of environmental extremes. Marine microbes alone, living in the unpromising milieu of seafloor sediment, are now thought to account for nearly one-third of all living organisms on Earth, yet very little is known about them. Perhaps fully half of Earth's total biomass is made of microbes, many of them extending as much as a kilometer into the crust. This is a whole new "deep biosphere" that geologists are only now beginning to explore by underwater drilling. Even in the more accessible (upper) parts of the oceans, bacteria are known to be much more numerous and diverse than previously thought—roughly several billion microbes infuse every teaspoon of seawater. All told, the oceans are brimming with an estimated billion billion billion bacteria, or roughly a million times more cells in the sea than stars in the visible Universe. And if many of these microbial life-forms within and under the sea are sucking up carbon, they may collectively beget a huge sink that absorbs carbon pollution produced by today's civilization and thus mediates climate warming on a global scale. How many more species are yet to be found in the depths of this, the largest habitat on Earth? And if life is so robust in unlikely places on Earth, to what extent does that raise the prospects for extraterrestrial life, even if it's only simple, creepy-crawly life?

Consider two candidates for alien life: Jupiter's moon Europa has a metallic core, rocky mantle, and probably more water locked in and beneath ice near its surface than in all the seas on Earth. Though the evidence for water is only conjectural, its likelihood opens up many interesting avenues for speculation about life. The *Galileo* mission to Jupiter recently returned direct imagery showing Europa to be totally icebound, yet those pictures also show a smooth yet tangled surface resembling the huge ice flows that cover Earth's polar regions. Something, most probably the tidal effects of Jupiter, is causing this moon (which is comparable in size to our own Moon) to be active and thus to allow water to be energized independent of the Sun. But a caveat is in order: water does not necessarily mean there's life.

Likewise, Saturn's big moon Titan is a place where odd life-forms, or at least the prebiological ingredients that embody life, could be present. Titan has twice the mass of our Moon and an atmosphere thicker than Earth's. Ninety percent of Titan's gas is nitrogen, much like Earth's air, laced with hydrocarbons (which are molecules made solely of hydrogen and carbon). Titan's environment likely resembles a gigantic biochemical factory powered by the energy of sunlight—and where there's energy and organic matter, well, who knows? It's mighty chilly there, though; direct measurements prove that Titan's surface temperature is a frigid two hundred degrees below zero Celsius—so cold that ordinary ice evokes steel. If life-forms do exist, or have existed, on alien worlds, they will probably be quite unlike those populating the sea ice on Earth today.

Life thus far has been couched with the qualifier "as we know it." That's carbon-based life, operating in a water-based medium, with higher forms metabolizing oxygen. All forms of life on Earth—from slimy bacteria to sentient humans—share this same basic biochemistry. And on the basis of what we now know about the various chemical elements, carbon would seem to be the atom best suited to form the long-chain molecules needed for life. But are we being chauvinistic? How do we know that other biochemistries are not possible?

Conceivably, other kinds of biology might be so different from life on Earth that we know neither how to study them nor how to test for them. For example, the abundant element silicon has chemical properties similar to those of carbon and thus might be an alternative to carbon as a basis for living organisms. Such weird biochemistries might have real advantages, implying that silicon-based life might be selected for survival in odd nooks and crannies on our planet, or especially in alien environments on extraterrestrial bodies. Heat comes to mind as one such property that would perhaps favor silicon chemistry over carbon chemistry. Silicon-oxygen bonds can withstand heat to more than three hundred Celsius and silicon-aluminum bonds to nearly six hundred Celsius. By contrast, carbon bonding, which molds the backbone of life as we know it, breaks apart at such high temperatures, well above the boiling point of water. This heat-resistant property of silicon is the main reason that silicone compounds are often used as industrial lubricants; even hot machinery runs smoothly with silicon-based grease.

Why, then, are there no silicon-based life-forms on Earth, especially given that silicon is more than a hundred times more abundant than carbon on our planet? The answer is that although silicon would seem to have an advantage in intense heat, carbon chemistry prevails within typical environments at or near Earth's surface. That is, at so-called room temperature, carbon bonds to other atoms more strongly, and especially so to other carbon atoms; furthermore, carbon bonds are unaffected by water, whereas silicon bonds break apart in most liquids, especially water. Yet another reason favoring carbon concerns the abundant atom oxygen. When carbon chemically reacts with oxygen, the result is carbon dioxide (CO_2), a gas that easily combines with other compounds; in our case, humans exhale carbon dioxide after inhaled oxygen has reacted with the carbon in our bodies. When silicon reacts with oxygen, however, the result is quartz (SiO_2), which is a solid. Can you imagine living creatures exhaling quartz bricks each time they took a breath?

The medium in which life operates can be likewise challenged. Must it be liquid? Admittedly, a solid is a poor interaction medium, unless perhaps the solid is pulverized as powder; atoms and molecules within hard solids have little mobility unless they are on the verge of liquefying. Gases are also poor substitutes for liquids; a gas doesn't easily stay put unless restrained by gravity or in a container of some sort. This type of loose reasoning leaves the liquid phase as the most reasonable interaction medium. But do liquids seem best only because we ourselves are partly made of liquids? Again, is our conclusion chauvinistic?

Several arguments do favor water as the most likely liquid medium for life, the best one being that the water molecule is made of two of the most abundant atoms—hydrogen and oxygen. Another reason favoring water as a preferred medium for life is its widely separated freezing and boiling points; for the conditions typical of Earth's surface, that range is one hundred degrees, allowing vital biochemical reactions to proceed anywhere between zero and a hundred Celsius. Yet another unique property of water is its reversal in density just before freezing; ice is less dense than liquid water, a statement that is untrue for any other known substance—hence the reason ice floats. If ice were denser than liquid water, it would sink and water would solidify from the bottom up, as do all other chemicals. Collecting at the bottom of a lake or ocean, ice would hardly have a chance to melt. It wouldn't be long before entire bodies of water, including whole oceans, became solid blocks of ice. Fortunately, water's peculiar density property prohibits this, ensuring lots of terrestrial liquid in which molecules can freely interact.

Ammonia (which is made of the common elements hydrogen and nitrogen) is sometimes proposed as a possible liquid medium in which life might develop, at least on a planet cold enough for ammonia to exist in the liquid state. To remain fluid, pure ammonia must be several tens of degrees colder than water, yet such low temperatures would inevitably cause metabolisms to slow, as less energy would be available to drive biological reactions. Although admittedly billions of years are available, the molecular interactions in liquid ammonia needed to produce more complex life-forms would be a good deal more sluggish than on watery Earth. Furthermore, ammonia doesn't have the peculiar density reversal near its freezing point, as for water, so large bodies of ammonia would likely freeze solid.

Together or separately, these and other alternatives to life as we know it would surely give rise to organisms with radically different biochemistries from that encountered on Earth. Still, scientists have little empirical data about noncarbon, nonwater biochemistries, for the very good reason that we have no examples of them to study. Nor is there much incentive to theorize about non-carbon-based, non-water-based life when today's biochemists themselves are clearly made of eighty percent water laced liberally with carbon. For health and medical reasons alone, we tend to study ourselves. Speculation will continue to run amuck about alien life beyond Earth, but currently it remains a subject for which there are no data.

∞

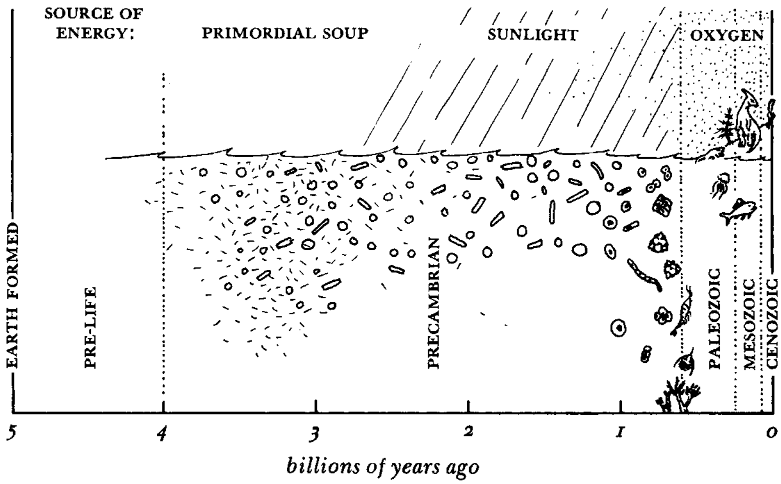
Biologists know of nearly two million different kinds of plants and animals, excluding microorganisms, now making a living on planet Earth.

This number of species includes a broad range of current, nonextinct life all around us, from tiny insects and pesky weeds to giant whales and redwood trees. And new life-forms are constantly being discovered, the total number of living species, including microorganisms, perhaps topping ten million. Yet even with this enormous variety of extant life, as noted earlier, more than ninety-nine percent of all species that have ever resided on Earth are now extinct—a number still rising as biodiversity falls under the onslaught of twenty-first-century society.

The fossil record narrates the following history of relatively recent life on our planet: Less than a billion years ago, multicellular organisms, having learned to utilize oxygen, quickly evolved into highly specialized creatures. These oxygen-breathing animals—the remotest ancestors of humans—swarmed in the sea, feeding on plants and on one another. Some could only float on the water, others anchored to undersea slopes, while still others had some mobility within the water and on the sandy bottom. Almost all the beings alive between one-half and one billion years ago had soft bodies. Hence fossil findings of the earliest respiratory organisms are understandably sketchy, for without bones or shells, few of them have remained intact to this day.

As the years wore on, lifestyles multiplied. Each type of organism responded to changes in the oceanic, continental, and atmospheric environment. Each attempted to adapt for better survivability. By some six hundred million years ago, tubular, soft-bodied fauna, such as the humble, long-extinct Ediacaran worms (whose fossils now notably dot Australia and Newfoundland), ruled the world.

Suddenly and dramatically, yet for reasons only partly known, the fossil record of a half-billion years ago erupts in numbers and diversity of species. Perhaps the reason was the establishment of a thick enough ozone layer or the efficient operation of the photosynthesis-respiration cycle, or perhaps skeletonization began leaving behind hard evidence of soft-bodied animal predecessors. The cause is uncertain, but the effect is clear: whereas previously only rather simple life-forms prevailed, in the relatively short period of a few tens of millions years—a mere geological moment—a bewildering array of new and more complex forms are seen among the fossils. Accordingly, paleontologists mark this "Cambrian explosion" as the start of a whole new age in the history of life on our planet—"biology's big-bang," best revealed perhaps by the plethora of novel fossils of the Burgess Shale in the Canadian Rockies and of the Chengjiang fauna in southern China's Yunnan province.



As the years wore on, lifestyles multiplied.

For ease in reference, all Earth times older than about a half-billion years are called simply the Precambrian, while all times since are subdivided into more recent ages: the Paleozoic (Greek for "ancient life") began about five-hundred-forty million years ago; the Mesozoic ("middle life"), some two-hundred-fifty million years ago; and the Cenozoic ("recent life"), sixty-five million years ago. These relatively recent time intervals delineate three great waves of animals in the evolutionary history of life on Earth.

The Paleozoic fossil record shows that the oldest known fishes originated more than five hundred million years ago. At first small and jawless yet with bodies completely covered in bony plates of armor, they were among the forerunners of true vertebrates. They probably dined on seafloor invertebrates while using those external skeletons to defend against predators. Soon thereafter, minute spores of land plants appeared nearly five hundred million years ago; megafossils of land plants themselves enter the record roughly fifty million years later as algae adapted to drier environments. These were followed by the first insects and amphibians some four hundred million years ago, and the first forests and reptiles about fifty million years later. Over the course of two hundred million years or so, not only had life greatly multiplied and diversified in the sea—tens of thousands of marine species are documented—but it had also spread from its oceanic nursery. Life had come ashore and colonized the land.

Naturally, as with all frontiers in science, some uncertainty (and often controversy) prevails—the sign of a robust, active subject, frankly. In this case, the most elementary, microscopic forms of life may have actually landed much earlier than a half-billion years ago. Unusual carbon-rich clays, but no fossils, have been found in ancient South African rock layers that are surely one and possibly even more than two billion years old. Just as carbon deposits at a few Greenland sites dating back nearly four billion years imply a possible origin for life earlier than the oldest fossils of three-and-a-half-billion-year age, bacterial landlubbers may have been reluctantly stranded on ancient lake shores or tidal flats earlier than at the start of the Cambrian, several hundred million years ago.

Sea plants, such as unicellular algae and then mosses, were probably the first life-forms to migrate from the water and along previously barren, rock-strewn coasts—an act of heady exploration of an alien landscape not too unlike humans later exploring the Moon. Animals that depended on this vegetation then apparently followed their source of food, among the first of the air breathers being the insect scorpions. Some early fish may have crept ashore intentionally, whereas others likely washed up onto beaches during storms or were left high and dry during droughts. Those species of fish able to use their fins to successfully negotiate the land became four-legged, air-breathing amphibians; those that tried and failed became extinct. Others became reptiles, which prospered among a bewildering variety of flying bugs that had a field day exploring new spaces. The result was an explosive invasion of the land well prior to the breakup of the ancestral supercontinent of Pangaea. Details are sketchy because geologists have no clear record of where shorelines were so long ago, what the climate was then like, and just how fast the environment was changing at the time.

The fossil record does provide incontrovertible evidence for an evolutionary trend: Descendants of those first shore plants became the world's first forests, and certain descendants of the amphibians eventually became the animals that lived in those forests. Some of that fossil record is richly detailed beyond our imagination. Who would have predicted that nearly half of all fossils would be trilobites, lobsterlike creatures of which the horseshoe crab is the closest living relative? Some trilobite species had heads, some apparently not; others had a dozen eyes, still others none at all. Most were quite small, measuring a few centimeters in length, yet some stretched half a meter from head to tail. Though all trilobites have been extinct for the past two-hundred-fifty million years, paleobiologists are reasonably sure that some version of them gave rise to many of today's animals.

Fossils also record when, and sometimes how, many of the features regarded as important evolutionary advances arose, including true bone, paired appendages, and cartilaginous jaws. Bone itself was a new material, needed later to support body weight when animals left the water. Made of cartilage mineralized largely with calcium, bone is nearly as solid, yet more flexible, than cast iron. And not only had spinal cords come forth, but gills too, rudimentary lungs that filter water to remove dissolved oxygen to power these aquatic newcomers. Not least, jaws served as the impetus for many of the sophisticated vertebrate qualities that followed. Whereas invertebrate fauna, such as worms and snails, ingest organic particles in mud for food, small fishlike creatures that

Evidence of early invertebrates.

Trilobites were enormously widespread organisms some five hundred million years ago. They appear in the fossil record suddenly and with great diversity at the time of the Cambrian "explosion" yet show no trace of earlier, ancestral forms. Most fossils found today are trilobites, of which this is a typical example, measuring about a third of a meter (or a foot) in length. All trilobites are now extinct, though surely modern lobsters and horseshoe crabs must be among their descendants. *Source: Smithsonian.*

evolved jaws had a distinct advantage beginning not quite five hundred million years ago: their jaws allowed a greater intake of energy by biting off larger pieces of other organisms. And their offspring eventually led to most vertebrates on Earth today, including human beings.

Once again, much as for many advances along the arrow of time from the early Universe to humankind itself, a general trend is evident whereby increased energy use (per unit mass) parallels the rise of complexity. Evolutionary biologists don't normally think in terms of energy expenditures, but when attempting to unify the natural sciences—a central feature of this book—energy is a powerful concept, as noted toward the end of the Chemical Epoch.

By the late stages of the Paleozoic Age, life was firmly implanted at sea, on land, and in the air. Some two-hundred-fifty million years ago, with Pangaea still intact, there existed a broad opportunity for living. The land in particular, with its green expanses and virgin forests, enabled animal life to proliferate with astonishing diversity. Species multiplied rapidly, so much so that the fossil record documents, just to give one homey example, nearly a thousand different kinds of roaches coexisting at the time. The household version—the common cockroach—is a direct, and very durable, survivor of the late Paleozoic.

All life on Earth had become dominated by the cold-blooded reptiles, a whole new life-form that had, over millions of years, evolved from vertebrate amphibians. The conquest of the land was complete, as reptiles spread out to fill every conceivable niche on the planet. As ancestors of nearly every animal now on Earth, the reptiles of two hundred million years ago had developed supple backbones, mobile legs, and keener brains than any other creature inhabiting Earth until that time.

The Mesozoic fossil record shows that many forms of life not only thrived but also evolved toward greater complexity. Plant life flourished, diversifying wildly while taking additional steps toward the current half-million green species. Simple angiosperms appeared in dazzling colors and rich scents (though true flowers came much later), all for the purpose of attracting pollinating insects. And the first birds took flight, most as small as today's sparrows.

A—perhaps *the*—highlight of the Mesozoic Age was the first appearance of the mammals, warm-blooded animals able to derive body heat from digested food and thus stay comfortable in cold environments. Fossil evidence reveals that three types of mammals originated midway through this nearly two-hundred-million-year period, partly because of widespread environmental change caused by the breakup of the giant Pangaean landmass that predated the modern continents as we know them. The earliest mammals, having the size and weight of a paperclip, were probably ancestors of the present-day anteater and aardvark—primitive creatures that had fur and nursed their young with milk, but, like reptiles, laid eggs instead of bearing live young. Another, more advanced group of mammals probably bore their young live like their descendants, the modern kangaroo and the koala bear. These young were so small and immature, however, that they had to be incubated in a fur-lined pouch under their mother's belly after the live births. Toward the end of the Mesozoic, true mammals appeared, laying no eggs at birth and needing no pouch for their young. Though the last sixty-five million years is often popularly called "the age of mammals," their Lilliputian ancestors were clearly occupying many ecological niches well into the Mesozoic, possibly as long ago as two hundred million years.

This outline aside, details of the mammals' line of ascent during the Mesozoic Age are somewhat obscure, as they were completely overshadowed by the mightiest reptiles of all time—the dinosaurs. Nothing at all like the snakes, lizards, or crocodiles of present times, in their prime roughly a hundred million years ago the dinosaurs roamed the Earth with skill and power, overrunning the air, land, and sea until they completely and devastatingly dominated our planet. Taking their name from the Greek words *deinos* (terrible) and *sauros* (lizard), these monstrous beasts were mostly department-store-sized land creatures often weighing as much as twenty-five tons. Their relatives included awesome seagoing reptiles capable of swallowing today's great white shark in one gulp and fearsome airborne brutes having wingspans comparable to those of modern fighter aircraft. Their fossils have been uncovered on all the world's continents, except at the poles.

The popular, stereotyped dinosaur was downright dumb—cold-blooded and small-brained. In chilly climates, or even at night, the metabolisms of such huge reptiles would have become sluggish, making it hard for them to move around, secure food, and thus survive. However, a new and controversial view has recently been embraced by several paleontologists. Studies of dinosaur fossils suggest that many of these monsters might have had large, four-chambered hearts, like those of mammals and birds. Such a heart could have pumped blood through organs, enabling the dinosaurs to sustain a high level of physical activity. If these revised interpretations are correct, some dinosaurs were probably warm-blooded and thus relatively fast-moving creatures. Also, though the dinosaurs clearly had small brains compared to those of today's mammals, they were still smart for their time. Indeed, no species able to rule Earth for more than a hundred million years could have been too dumb. By comparison, humans have thus far dominated for hardly a few million years.

All the flying, swimming, or landlocked prehistoric predators disappeared with bewildering abruptness near the end of the Mesozoic Age. The presence of the dinosaurs simply vanishes from the fossil record. No one is certain why, nor does anyone know how the mousy mammals were able to survive throughout the hundred-fifty-million-year reign of terror when these beasts were overwhelming and the mammals were scared. Whatever the cause of their demise, it affected not only the dinosaurs but also many other forms of life. Fossil records demonstrate that about sixty-five million years ago, nearly half of all plants ceased to exist, including more than eighty percent of the plant species in present-day North America. Most mammals, reptiles, and birds also perished; fossils of all animals larger than twenty kilograms (about fifty pounds) are absent from later geological layers.

Explanations abound for the dinosaurs' complete and total extinction. Devastating microbial plagues, magnetic-field reversals, sea-level changes, as well as deep volcanic eruptions and severe climatic shifts: any and all of these, perhaps triggered by asteroid impacts or supernova explosions, have been proposed. Each of these ideas has some merit, though none is entirely convincing. Out of seeming desperation, some researchers even joke that the dinosaurs died of constipation, since many oily plants on which they probably feasted also became extinct at about this time.

Currently, the most popular idea is that the dinosaurs rapidly expired because a huge extraterrestrial object collided with Earth some sixty-five million years ago—another obvious and dramatic interaction of astronomy with biology. Studies of impact cratering on the Moon do imply that about sixty meteorites at least several kilometers across have probably hit the Earth since the start of the Cambrian nearly six hundred million years ago. Most of the direct evidence has eroded away on Earth, but it's there for all to see on the weatherless Moon, for if the Moon were belted then the more massive Earth must have been bombarded at least as much. Even the smallest of those collisions would have released a blast of kinetic energy equal to that of more than a million nuclear bombs. So the notion that asteroids have belted our planet—and could have caused biological upheaval—is not as far-fetched as widely thought a generation ago. In an about-face in scientific circles, we now realize that Earth is under a regular barrage by the cosmos. It's the natural scheme of things for a Solar System still littered with debris from its formative stages.

In the specific case of the dinosaurs, a large, ten-kilometer-wide asteroid or comet almost surely struck the Earth, causing great quantities of dust (mostly its pulverized self) to become airborne. The dust, in turn, reached the altitude of the jet stream, encircled the planet for several years, darkened much of the atmosphere, shut down photosynthesis, and disrupted the base of the food chain by killing off many plants. A thin horizontal band of reddish clay enriched with the rare element iridium found sandwiched between layers of sixty-five-million-year-old ancient limestone (one from the Cretaceous and the other from the Tertiary geologic periods) is the main piece of evidence supporting the asteroid-impact idea. Although rare on Earth's surface (since most of it long ago sunk to the planet's interior), the iridium in the clay is hundreds of times more abundant than in native crustal rock yet matches levels found in meteorites. This idea is not without problems, however: If the iridium then "rained down" out of the atmosphere, why is its abundance so highly varied from place to place on Earth's surface?—unless the highest iridium content is pointing toward a probable impact zone, seemingly somewhere in the Americas. Where is the crater left by the impact?—though a good candidate is the so-called Chicxulub crater buried under a kilometer of partly submerged sediment on the north coast of what is now Mexico's Yucatan Peninsula. And if instead the asteroid landed in the ocean—the most likely scenario given that three-quarters of Earth's surface is water—then how did it uplift dust and debris high into the air? Perhaps, argue some opponents, the iridium-enriched clay was laid down by volcanoes and had nothing to do with an extraterrestrial impact. Or, perhaps the blast of an impact also prompted intense volcanism, in which case they both wreaked havoc.

For whatever reason the dinosaurs perished, environmental change of some sort, and probably dramatically so, was responsible. It would be useful to continue to seek the cause of their extinction, for there's no telling if sudden, global change might strike again. As the dominant species on Earth, we are the ones who perhaps now stand to lose the most.

Despite our knowledge of dinosaurs, no human ever saw one alive—unless today's birds are an evolutionary offspring of the dinosaurs, as some paleontologists now think. Those movies showing hulking dinosaurs terrorizing cave dwellers (let alone modern cities) are simply wrong. Dinosaur remains lay hidden for about sixty-five million years before *Homo sapiens* discovered their prior existence less than two centuries ago. Our great-great-grandparents and all those who preceded them knew nothing of the dinosaurs. Ours is the first generation to realize that we probably wouldn't be here had these great hulks not gone extinct. Only when the dinosaurs disappeared did the spectacular rise of the mammals—including human beings—begin.

The onset of the Cenozoic Age, some sixty-five million years ago, saw the appearance of an almost entirely new cast of characters. The huge landmass of Pangaea had fully broken up into continents nearly familiar to us now. The dinosaurs were completely gone, along with nearly three-fourths of all life on Earth to that time. The earlier reptilian dominance over the mammals had been totally reversed. Flowering plants bloomed widely. Fruits first grew upon the landscape. And the planet had returned to its pre-dinosaur, Paleozoic tranquility. Clearly, the mammals had taken over the world, although they had apparently done so by default. In a certain sense, the meek had indeed inherited the Earth.

Fossils as recent as fifty million years old show that most mammals had small brains, large jaws, and clumsy and inefficient feet and teeth. None was larger than an opossum, and most had evolved eye sockets, adapted to night vision. Life wasn't too tough, though, as those fossils show that they freely multiplied, swelling in numbers and diversity. As always, change was rampant. Ice ages came and went; continents split and drifted. In generation after generation, life-forms constantly fine-tuned their daily routines for better survivability. Accordingly, many of these early mammals passed into extinction, to be replaced by better-adapted stocks.

In a relatively short time, the mammals had evolved into an amazing assortment of creatures. Some forty million years ago, the ancestors of such modern mammals as the horse, the camel, the elephant, the whale, and the rhinoceros, among others, gradually ventured forth, though often in shapes and sizes nearly unrecognizable compared to their descendants of today. Most of these life-forms improved their overall performance between forty and twenty million years ago, converging by means of seemingly endless changes toward the multitudinous, yet threatened, biodiversity of flora and fauna seen around us today in the twenty-first century.

∞

Myriad life-forms have come and gone in a broad ecological panorama stretching across both space and time on planet Earth. Some were weak and trifling organisms, while others ruled the land, sea, and air. For hundreds of millions of years, a steady parade of new creatures has marched forth, many of whom led in turn. Yet only the latest of these dominant life-forms—the men and women of today—know about all those that went before. Modern humans alone have been able to unearth and chronicle the absolutely amazing story of the now-extinct and bizarre life-forms once prevalent on the third planet out from the Sun.

What sense can be made of the copious array of past and present life on Earth? Does any logic link it all—anything to unify it into a coherent package of understanding? Classification is the first step in any attempt to discover underlying causes of the abundance and diversity of life on our planet. And some general trends are indeed apparent, though surprisingly biologists are still debating life's basic categories.

All current life-forms as well as fossilized remains of ancient life are often broadly classified as bacteria, plant, or animal. These classes, in turn, are further divided into different species, a subclassification generally used to denote not only structural similarity but also ability to mate and produce fertile offspring. But that was twentieth-century biology, and this is now.

Recent discoveries of extremeophilic life as well as renewed awareness of the astounding biodiversity on Earth have caused biologists to revise the "universal tree (or bush) of life"—in other words, to rewrite a main and important part of all biology textbooks. A new consensus has emerged that molecular studies of extant life—especially the rates of change among genetic markers (mainly nucleotide-base sequences of RNA) shared by different groups of organisms—probably grant a more accurate picture of evolutionary relationships among all known life on Earth. Such "molecular clocks" stipulate that the three main branches, or domains, of life are now bacteria, eukarya, and archaea—the last of these comprising mostly the newly found thermophilic (heat-loving) life-forms residing in extreme environments. Plants and animals of the old classification scheme are now both grouped into eukarya, all of whose members are the eukaryotes noted a few pages earlier. And then there are the bacteria, whose microbial members probably constitute well more than half of Earth's entire biomass. It is the past overlap of the bacteria and the archaea branches of prokaryotic life, way back in deep time, that will someday likely reveal the last common ancestor that sprang forth from life's origin nearly four billion years ago.

Not all life-forms, however, fit cleanly into neat categories. Life isn't so simple, and exceptions abound in Nature. The unicellular Euglena is one such good example. Considered an animal by zoologists (animal biologists) since it can move rapidly like an animal, the Euglena is also claimed to be a plant by botanists (plant biologists) since it consumes energy much like a plant. Further research might resolve to which category it really belongs, as with fungi (including mushrooms, molds, and mildew) that are nearly always treated as plants yet don't photosynthesize and have much in common structurally with animals. Diatoms, too—those simple, single-celled microbes that resemble microscopic, aqueous snowflakes—lie somewhere between plants and animals. Exceptions to rules are common in biology because either the data are incomplete or the subject is complex—often both. The biological sciences have few hard-and-fast rules, quite unlike the physical sciences.

A thorough understanding of life goes beyond its cataloging; this isn't mere stamp collecting. Real creatures don't always match what's expected of a species. Rather, individual species often show small, though noticeable variations from their "ideal" categories—slight deviations from some standard specimen to which each individual organism may be compared. This is true of all species, whether they are now living, dead, or fossilized.

As with many types of matter in the Universe, change is key to their being, and mentally modeling that change is key to our understanding. Here too, as with other aspects of the cosmic-evolutionary scenario, the study of change contains insights needed to fathom how life has evolved throughout the course of time. We have entered the realm of biological evolution—the changes experienced by life-forms, from generation to generation, throughout the history of life on Earth—perhaps the most intellectually powerful core of this, the Biological Epoch.

The theory of biological evolution, independently conceived by the mid-nineteenth-century British naturalists Charles Darwin and Alfred Wallace, can account for two outstanding features of the fossil record: First, living systems have generally become more complex with time. Second, variations among members of all species are more the rule than the exception. (Darwin gets most of the credit since his thinking predated that of Wallace and he provided many more data and examples in support of the theory. However, the very idea of evolution had been "in the air" for at least fifty years before they made it famous.)

These two facts clash head-on with the age-old assumption that Nature is immutable. Like Copernicus and Galileo a few centuries earlier and Heraclitus twenty centuries before that, Darwin and his colleagues faced the same kind of opposition made popular by Aristotle, who refused to concede that species change. But given these undeniable facts of Nature, a static theory of life is simply untenable. Everything changes with time, life included. The only feasible explanation is a dynamic, evolutionary one.

The central tenet of biological evolution maintains that living things change, some for the better, others for the worse. Some species thrive, others go extinct, and yet others arise anew. Those that survive for lengthy periods of time are often drastically modified, sometimes becoming whole new species. Amidst all this change, organisms with similar structures share similar ancestry and are closely related. Those with very different structures have accumulated those differences over long durations and are therefore now only distantly related.

Biological evolution is not faith. It's as much a fact as Earth's orbiting the Sun. The fossil record no longer leaves room for reasonable doubt that evolution does happen. That "what" aspect of evolution is backed by data. The "how" aspect is less clear, which is why biological evolution is rightfully called a theory.

Although evolution itself is factual, the mechanisms that cause evolution remain theoretical. What we have here is a mental model, yet one solidly based on the scientific method: observations were made of the fossilized remains of life; an idea was proposed to explain those facts; and subsequent experiments have served to strengthen and revise the intricacies of the theory during the past century and a half.

In particular, observations show that although all species reproduce, few of them display huge increases in population. The total number of any one species remains fairly constant, with no dramatic rise in offspring from generation to generation. Furthermore, the process of reproduction is almost never perfect; offspring in each generation are hardly ever exact copies of their parents. The implication is that not all offspring survive to reproduce. Life must struggle and compete in order to endure. (Humankind is an exception whose population is exponentially rising, but that's because we are now affected more by cultural evolution, whereby not only do the fittest survive and reproduce but nearly everyone else does too.)

What is the primary agent of biological evolution? How does it work? The chief instigator is the environment, the physical conditions surrounding all living things. Temperature, density, foodstuffs, air composition and quality, in addition to natural barriers such as rivers, lakes, oceans, and mountains, are among a whole host of influential environmental factors. Other factors are more subtle, such as personality clashes (or individual love), neighborhood squabbles (or group harmony), among scores of exceedingly complex sociological pressures (some beneficial, some not). Further complications arise given that environmental conditions frequently change, albeit often slowly. Biological evolution asserts that all life-forms respond to their changing environments, inhibiting some traits while promoting others, thereby yielding an immense diversity of species throughout the course of time. Changes—in the environment and in life, and to repeat for emphasis—occur as a rule, not as an exception.

Natural selection, an expression coined by Darwin himself, is the mechanism that guides much of life's evolution along time's arrow. Recognizing that most members of a species exhibit some variation from their ideal standard, Darwin argued that organisms having a variation particularly suited to their environment would most likely survive. They are quite naturally selected to live. By contrast, those organisms having unfavorable variations would most likely perish. They are naturally selected to die. In short, only those life-forms able to adapt to a changing environment tend to survive long enough to reproduce, thereby passing their favorable variations or traits on to their descendants.

A central feature of biological evolution is modification followed by adaptation—the positive response to a changing environment by an organism having some variation or trait that improves the organism's chance for survival and reproduction. Pedagogically, though, it's often more instructive to regard natural selection as a negative influence on organisms within a population. An even a better word for "selection" would be "elimination," for it is by means of nonrandom elimination that biological evolution really operates in Nature. Nature eliminates more than it selects, and it does so deterministically in response to chancy events. That said, evolution rarely throws out a good scheme but instead modifies and embellishes on whatever already exists.

In successive generations, advantageous traits become more pronounced in each individual; they accumulate. Not only that, but the numbers of individuals possessing favorable traits also increase. Favored individuals generally produce larger families, as they and their offspring have greater opportunities for survival. Their favored descendants multiply more rapidly than those of their less advantaged neighbors, and over many generations their progeny replace the heirs of individuals lacking the desirable trait.

Natural selection truly does smack of the well-known phrase "survival of the fittest." It literally molds life-forms. With the passage of sufficient time, the action of natural selection can greatly alter the shape, disposition, ability, character, and even the existence of individuals. Old species can disappear in response to changing conditions, while entirely novel ones can arise anew. Yet because the element of chance is indeed a factor, the outcomes are not predictable.

Note once more the twin roles played by chance and necessity, as elsewhere in cosmic evolution. The mechanism of natural selection is not an active force, like those often guiding strictly physical change. Instead, natural selection in biology acts as a sifting process—an "editor" of sorts—permitting some species to thrive quite naturally while others become extinct. Chance events admittedly often trigger evolutionary change, but natural selection has a decidedly deterministic bent that directs that change—though mostly by a process of elimination. Contrary to popular opinion, Darwin never said that the order so prevalent in our biosphere arises from chance alone. Yet even the limited role of chance in modern neo-Darwinism, when coupled with the deterministic part of natural selection, is capable of generating highly improbable results. Chance and necessity, mutation and selection: it is the synthesis of randomness and determinism that ultimately gives rise to the spectacular novelty and creativity seen among the wonderfully diverse and talented life-forms adorning planet Earth today.

Natural selection cannot be easily observed at work. Passages of time usually far longer than a human lifespan are needed to witness large-scale evolutionary change in any population of a species. Some experimental success has been achieved in laboratory settings that mimic those of Nature. Like the origin-of-life experiments discussed in the Chemical Epoch, these simulations study the adaptation of life to a changing environment. In all cases, the results support the theory of biological evolution via natural selection.

Here is one such experiment, conducted under carefully controlled conditions: Two groups of field mice, one group with dark fur and the other with light, were let loose in a small barn with an owl. The straw and ground cover were chosen to match closely the dark coloring of one set of mice. This camouflage, then, gave the darker-colored mice an environmental advantage to hide; the lighter-colored mice were clearly at a disadvantage. At the end of this well-documented experiment, the owl had captured many more of the lighter-colored mice. When the ground cover was lightened—corresponding to an environmental change granting the lighter-colored mice a greater opportunity to survive—the results were reversed; the owl readily captured the darker-colored mice. This is an example of how small variations in one species can grant a competitive advantage. As might be expected in the real world outside the laboratory, the natural habitat of the light mice is cornfields, of the dark mice, forests. In each case, the mice best adapted to their environment were naturally selected to live and thus to reproduce their kind.

Recently, the tools of molecular biology have allowed researchers to track minute changes among life in test tubes. With computers that can store lots of numbers and analyze them quickly, evolution has become visible as whole populations move from generation to generation. For example, in one such laboratory experiment that ran for many years, a sparse sugar broth was laced with a quick-replicating bacteria and each day siphoned off into a fresh flask of food. Over the course of several thousand generations of the bug—known as *Escherichia coli*—biologists were able to study evolutionary dynamics in action. The result of the experiment—with food availability acting as the selector—suggests that rare, beneficial changes gradually led to increased cell size, much as expected if natural selection was truly at work.

Examples of natural selection have also been observed in Nature's outdoor setting whenever environmental factors change exceptionally fast. Consider a classic study done over the course of the past century. Normally, in rural areas, the bark of many trees in the English countryside is abundant with light-colored lichens growing freely on their dark trunks, enabling the famous "peppered moths" to blend nicely with their environment. In their struggle to survive, the moths prospered while resting on the trees against whose bark they were nearly invisible. By contrast, their darker-colored moth relatives lacked this competitive advantage because they stood out clearly against the lichen-rich bark, allowing birds to snare an easy meal. About a hundred years ago at the height of the Industrial Revolution, however, many trees near manufacturing cities had become heavily soiled since lichens are highly susceptible to airborne pollutants. This environmental change—rapid by Nature's standards—had killed the lichens and thus removed the benefit previously enjoyed by the lighter-colored peppered moths. The result was that few pale moths prevailed, at least near industrialized areas; instead, the darker moths possessed the advantage of camouflage, enabling them to avoid the birds, mate in peace, and freely reproduce. Today, the situation has once again reversed: with the reduction of industrial pollution in recent years, the number of peppered moths has rebounded along with the growth of lichens, each inversely correlated with the lower levels of soot and sulfur dioxide in the cleaner air. This is an example of how simple variations—in this case color, as with the mice in the barn—serve to guide natural selection within a changing environment. For some moths, in fact, a small change became an issue of life and death.

The common housefly presents yet another natural example of how some members of a single species can adapt to a changing environment, granting them an enhanced chance to be naturally selected for survival—or, better and boldly stated, to avoid being eliminated. Originally, the pesticide DDT was successful in killing houseflies. During the first several years of its use, DDT killed almost all the flies; few of them could adapt to this sudden environmental change caused by the chemical DDT in the air. A small minority of flies, however, managed to survive because they possessed a chance variation or trait that made them resistant to this chemical. These oddball survivors reproduced freely and thus passed the advantageous trait onto their descendants. Within a decade, the offspring of the survivors outnumbered the original majority type of fly. Accordingly, DDT has grown less effective over the years. Now most houseflies have inherited a resistance to DDT, and the pesticide is useless against these flies. The chemical DDT did not give this resistance to the flies; rather, it provided an environmental change enabling natural selection to go to work. To survive as a species, the housefly had to adapt to the changing environment. Those that managed, survived. Those that were unable, are long gone.

Evidence of natural selection.

In a classic example of Darwinian selection, moths having small variations enabling them to adapt to changing environments are clearly favored in the struggle for survival. Here, on the dirty bark of a tree, a darker-colored moth blends with its environment better than does a lighter-colored moth. Hungry birds clearly and preferentially pick off such light moths, allowing the darker ones to prosper and reproduce. Literally hundreds of similar cases of natural selection have been reported in the medical and agricultural literature. *Source: Harvard.*

These last two cases are examples of evolutionary response to environmental change induced by humans—a whole new aspect of evolution whereby technologically equipped beings play the role of Nature, and one best left for the next, Cultural Epoch.

Over lengthy durations, chance variations in living things can accumulate. Hair color, eye color, size, shape, appearance, and a host of other attributes all change as Nature naturally selects for survival those life-forms best adapted to the environment at any given time—or again to stress the real action in the wild, eliminates those life-forms that are poorly adapted. Eventually, some life-forms come to differ noticeably from members of their original species. In this way, the environment helps new species to evolve from old ones.

For example, some members of a single species might become isolated by some physical change in the environment, such as a river that reroutes (due to flooding or plate tectonics, for instance) through an area inhabited by a population of butterflies. Should the river act as a physical barrier too wide to be crossed, butterflies on one side would be unable to mate with those on the other side. The two populations of butterflies, then completely separated, will begin to differ as minute changes accumulate in each of them over long periods of time. In some cases, one group of butterflies—often called a "founder population"—will eventually differ greatly, even in outward appearance. Should the geographical barrier later be removed—if the river dries up, for instance—then the two populations would be able to intermingle once again. Provided they were separated long enough to allow real changes in form, they will be unable to interbreed. Two new species of butterflies will then exist where previously there was only one. Furthermore, each new species will stake out its own claim or fill a separate niche, thereby coexisting within the new environment.

Environmental disruptions of this sort often guide the transformation of a single species into two or more species. Known as speciation, or "disruptive selection," this is the mechanism behind the diversification of all life. It might take centuries, or even millions of years. The rate of evolution depends on a whole array of factors including the amount of initial environmental change and the extent of resulting adaptation.

Environmental disruptions . . . often guide the change of a single species into two or more species.

An actual case study entails recently upthrust mountains and eroded ravines in the Grand Canyon—dramatic environmental change over geological timescales. Two distinctly different populations of squirrels live on the north and south rims of the canyon. The Kaibib squirrels of the north rim have black bellies and white tails, whereas the Abert squirrels of the south rim have white underparts and gray tails. Both feed on pine-tree bark growing only on the kilometer-high plateaus. The two populations are now separated, and presumably have been for thousands of years, by the intensely hot and dry conditions in the canyon. But they have so many similarities that it seems safe to assume that their ancestors were once members of the same species.

Other examples abound of many slightly different species, coexistent but isolated and apparently sharing a common ancestral heritage. Scientists find more of them every day, including members of species that are not even separated by a physical barrier but that for one reason or another do not interbreed. The factors are so numerous and the timescales so long that it's often impossible to reconstruct the myriad changes that led to the current state of biological affairs.

One final clarification about the distributions of traits: When biologists refer to populations of species, they are really thinking in terms of gene pools—the whole spread of variations within a given population of a species. It's the genes, or DNA fragments, of different populations that are isolated from one another and that gradually develop variations. Therefore, we need a better understanding of changes among the microscopic genes if we are to fully appreciate modern biological evolution. This we shall take up in a page or two, but before leaving our discussion of Darwinism, let's consider one other, currently popular yet controversial, alternative to classical Darwinian evolution.

The fossil record of the history of life on Earth clearly documents many episodes of mass extinction—times when at least half of all life perished, after which life's diversity lay dormant for up to several million years. In addition to the "great dying" some sixty-five million years ago that ended the reign of the dinosaurs (as well as about two-thirds of all other living things), several other dark times devastated biodiversity. To cite two of the worst episodes, about four-hundred-and-forty million years ago vast numbers of animals then living in the sea quickly vanished, and some two-hundred-and-fifty million years ago as much as ninety percent of all marine species and seventy percent of all land species suddenly became extinct. The last of the trilobites disappeared during this quarter-billion-year-old event, as did all of Earth's ancient corals, most of its amphibian families, and nearly all of its reptiles. Life itself was nearly extinguished on our planet. As for the dinosaurs later in time, a cosmic killer might have been the trigger, namely, an asteroid that impacted southern Pangaea and likely caused huge volcanic eruptions and massive lava outflows, supplemented by dramatic fluctuations in climate and sea level. Ironically, this potential impact might have opened the door to the later emergence of the dinosaurs, though not all experts look to the sky for causes of dramatic changes on Earth.

Some paleontologists have proposed that the most notable mass extinctions have been periodic, occurring roughly every thirty million years. One possible reason for this (hotly debated) periodicity in the fossil record conjures up another astrobiological connection: the remote Oort Cloud of comets, some fifty thousand times the distance from Earth to the Sun away, is disturbed each time the Solar System oscillates above and below the galactic plane of our Milky Way, potentially causing numerous comets to be ejected from their regular orbits and toward the Sun. Some of these comets would rain down on Earth, disrupting the climate, reversing the poles, or otherwise upsetting our planet's environment, thus causing mass extinction of life on Earth. Another, astounding proposal is that our Sun might have a companion star which, in a highly elliptical orbit, would periodically pass through the Oort Cloud every thirty million years or so, thereby creating a gravitational uproar that sends comets sunward. Observational efforts using infrared equipment have searched for this dim, dwarf, and decidedly hypothetical star—already named Nemesis, after the Greek goddess who relentlessly persecutes the excessively rich, proud, and powerful—but thus far to no avail. It probably doesn't exist.

The last known sizeable asteroid to strike North America occurred some thirty-five million years ago at the base of the present-day Chesapeake Bay. A submerged and mostly buried crater nearly a hundred kilometers across (found by energy companies while prospecting for oil) implies that an impacting rock several kilometers in diameter flung out a hail of white-hot debris and a surging tsunami that must have turned the eastern seaboard into a wasteland. The fossil record shows a moderate extinction of numerous sea creatures about a million years later. Though not as large as the asteroid that likely felled the dinosaurs, the resulting damage must have been awesome nonetheless—and if it happened today would likely wipe out most East Coast cities, threatening millions of people.

The most recent significant hit was in 1908, when an asteroid (or comet) exploded over Tunguska, in Siberia. That object was probably only about a hundred meters across (or the size of an apartment building), yet was still able to flatten trees over thousands of square kilometers with the force of a multimegaton nuclear blast. And in early 2002, an even larger asteroid whizzed past Earth at about double the distance to the Moon—a relatively close call. Stunned astronomers didn't see it coming until it was about a week away.

Regardless of whether comets and asteroids have regularly belted Earth sufficiently to cause mass extinctions, biologists now realize that the rate and tempo of evolution have apparently not been steady throughout the history of life on Earth. By contrast, Darwin himself argued that natural selection operates *gradually*, with slow, steady, uniform changes occurring through time. He envisioned the development of species much as he saw environmental change on Earth—as a process of smooth, gradual change proceeding at a uniform rate. But paleontologists are now gathering increasing evidence that biological evolution may well have operated more erratically, with rapid, even catastrophic changes occurring every so often.

The fossil record is spotty and not always seemingly in accord with the gradualism posited by Darwin. Few fossils show clean, continuous transitions from one form to another with myriad small, intermediate steps linking one species to another. Of course, one possible explanation of these so-called gaps is that the fossil record is incomplete—which it surely is to some extent. But perhaps Nature is also not so continuous; perhaps it changes abruptly owing to irregular upheavals. Today's fossil record implies that many species have remained more or less the same for long stretches of time often measured in millions of generations, after which they rather suddenly underwent bursts of evolution in perhaps less than a thousand generations—too short to have left clear, transitional forms in the rock layers.

To more closely match the geological record, a contrasting idea, called punctuated equilibrium, has attracted some adherents during the past few decades. According to this theory, life stays pretty much unchanged until something drastic happens and then changes fast—sometimes for the better and species diversification, at other times for the worse and species extinction. Life's "equilibrium," or "stasis," is said to be "punctured" by rapid environmental change. And when speciation rates—the pace at which new organisms emerge—cannot keep up with extinction rates—the pace at which they vanish—the net result is often long-lasting depletion of biodiversity.

The theory of punctuated equilibrium is a slight variation on classical Darwinism. It's not at all a violation of the basic means of biological evolution; natural selection remains the principal way that life changes from species to species. Punctuated equilibrium merely claims that the *rate* of evolutionary change is not so gradual. Instead, the "motor of evolution" occasionally accelerates during periods of dramatic environmental change, such as asteroid impacts, magnetic reversals, volcanic eruptions, and the like. We might say that evolution is imperceptibly gradual most of the time and shockingly sudden some of the time.

Not all scientists have yet accepted the notion of punctuated equilibrium. Perhaps the rate of evolution is largely a matter of perspective: those who examine the fossil record in fine detail, across short durations, will occasionally see evidence for periods of rapid speciation; yet those who step back and take a broader view, over very long periods of time, will see more gradual change. This debate is still underway—a perfectly reasonable debate about the tempo and not the mode of change—and is not likely to be resolved until the fossil record is more complete or the genetic record better understood.

What is it that alters living systems to make members of a single species sometimes unable to interbreed? Basically, the minuscule gene is the culprit, for it's the genetic code that dictates if and how life-forms reproduce. Pioneered more than a century ago by the Austrian monk Gregor Mendel, the subject of genetics has become a good deal more complex than he could have ever imagined. Darwin himself would probably be surprised at the microscopic roots of biological evolution as we know it today—a modern synthesis of Darwinian and Mendelian ideas, often referred to as neo-Darwinism.

Still, what causes the genetic alterations? What factors contribute to the similarities and differences in organisms? In short, what is the origin of all the myriad variations seen throughout the living world?

Hereditary error is a major factor promoting the evolution of living systems; it is in fact a prerequisite for evolutionary change. Note the phrase hereditary *error*, not heredity itself, which is an agent of continuity, not change. Heredity, by definition, is the transmission of genetic traits from parents to offspring, thus ensuring the preservation of certain characteristics among future generations of a species. Otherwise, the basic life functions and body organs of each and every organism entering the world would have to be created from first principles. Normally, chemically coded instructions of the DNA molecules enable cells to duplicate themselves flawlessly millions of times. But occasionally, mistakes do occur at the microscopic level. Not even genes are immutable. Everything changes.

For reasons only partly understood, a DNA molecule can sometimes drop one of its nucleotide bases during replication. Or it may pick up an extra one. Further, a single base can suddenly change into another type of base. Even such slight errors in the DNA molecule's copying mechanism mean that the genetic message carried in the DNA molecule for that particular cell is changed. The change does not have to be large; even a change in a single nucleotide base among millions strung along the DNA molecule can produce a distinct difference in the genetic code. This, in turn, causes a slightly modified protein to be synthesized in the cell. Furthermore, the error is perpetuated, spreading to all subsequent generations of cells containing that DNA.

Microscopic changes in the genetic message—mutations—can affect offspring in various ways. Sometimes the effect is small, and newly born organisms seem hardly any different. At other times, mutations can alter a more important part of a DNA molecule, inducing marked change in the makeup of an organism. And, at still other times, a single mutation can rupture a DNA molecule severely enough to cause the death of individual cells and even whole organisms. An example of the last of these is cancer.

Mutations are responsible for differences in hair and eye color, body height and finger length, skin texture, internal organs, individual talents, and numerous other traits among a population of life-forms of any given species. Virtually any aspect of the life of an organism can be modified by genetic mutations. Such mutations provide a never-ending variety of new kinds of DNA molecules.

Not all mutations are detrimental. Most of them do indeed create traits inferior to the previous generation's, especially in many of today's highly evolved and exquisitely adapted organisms. But some mutations are favorable and serve to better the life of an individual. These can then be passed on to succeeding generations, making life more bearable for members of that species. Nature provides a fine balance between error and accuracy in replication: too many mutations and an organism can't function; too few and it loses adaptability. Beneficial mutations aid the motor of evolution, steering life-forms toward increasing opportunities to adapt further to ever-changing environments.

What causes genes to mutate? Why do some DNA molecules occasionally replicate differently, though they may have copied themselves exactly for millions or even billions of previous cell divisions? Whatever the reasons, the outcome is unpredictable, for chance, indiscriminate events are at work, at least in part. Biologists have performed laboratory experiments with cells and have managed to increase the number of mutations by artificial means, thereby helping to unravel some of the causes of genetic change. The results so far show that the easiest way to enhance gene mutations is to treat reproductive cells with external agents.

Three of the most important mutation-inducing agents revealed in the last few decades are temperature, chemicals, and radiation. When cells are heated or treated with industrially generated nerve gas or chemical drugs, mutations are clearly enhanced. In addition, ultraviolet and X-ray radiation seem to be an especially striking source of genetic mutations. Radiation, in one form or another, has been present on Earth since the beginning of geologic time. Radioactive elements embedded in rocks, cosmic rays bombarding Earth from outer space, and even small amounts of solar ultraviolet energy reaching the ground all serve to prove that life originated and evolved in a radiation-filled environment.

Generally, there is nothing wrong with immersion in radiation. We and other life-forms probably wouldn't be here without it. In the absence of radiation, life itself might not have progressed beyond the primitive, unconscious, unicellular organisms drifting in the oceanic slime. Of some valid concern now, however, is the fact that human inventions such as atomic bombs, nuclear reactors, and some medical devices also release radiation. Intense doses of radiation, let alone nerve-gas agents, can kill directly, though more subtle doses cause changes in the reproductive cycle that are then passed on to future generations. It's unclear that these human-induced mutations are in all cases harmful, but, in the absence of evidence to the contrary, a healthy degree of skepticism is surely warranted.

We must not risk damaging the refined work of several billion years of organic evolution, for the long sequence of changes that evolution engenders cannot likely ever be repeated.

∞

Hardly a century ago, the concept of biological evolution was intellectually and morally shocking. Few people embraced it; even many scientists of the late nineteenth century rejected it. The problem was not really the idea of evolution. Surely evolution occurs; that much was known well more than a hundred years ago. Fossils were already abundant then, and agriculturists had for centuries bred crops and livestock in a successful effort to develop healthy, disease-resistant foodstuffs.

The real problem was that many people were disturbed to hear that humans have anything in common with a bunch of hairy apes. When ideas involve ourselves, vanity seems to surface like an irrepressible force. Largely owing to this conceit—often a dogmatic desire by some to put humankind on an anthropocentric pedestal—minor segments of twenty-first-century society still refuse to accept the reality of biological evolution.

Bioscientists now combine fossil discoveries, genetic assays, and behavioral studies to virtually prove that of all nonextinct species of life now on Earth, the chimpanzee and gorilla are our closest relatives. Humans have not descended from these apes, a common misunderstanding. Rather, modern science demonstrates that apes and humans share so many attributes that they likely have a common ancestor. We should not be able to identify that ancestor from among any of the presently living creatures on Earth, for genes and environments change over the course of millions of years. But such an ancestor should be part of the fossil record. Our common ancestor more likely resides in a museum than in a zoo.

To discern our most recent ancestors and thereby trace the ways and means of relatively recent biological evolution, paleontologists rely heavily on the fossil record. Fossils of recent times are generally well preserved, enabling researchers to document evolution with reasonable accuracy. Not surprisingly, older fossils are in poorer condition, often in pieces and hardly recognizable. Much as with stars and stellar remnants of different ages and states, reassembling the pieces of decayed and broken fossils is very much like solving a jigsaw puzzle. Trying to understand where and when the reconstructed organisms fit into the evolutionary line of descent is often an even trickier task.

Teeth and skull bones account for the majority of fossils found since people began digging around for artifacts in Earth's rubble a few centuries ago. Teeth are the most enduring part of any life-form because of their very hard enamel. Skulls are the most recognizable part, largely because they are more noticeable than arm bones or leg joints among the sticks and stones and other ground litter. Careful study of these and other bone fragments has enabled researchers to arrive at a consensus for the lines of descent eventuating in humanlike creatures. In brief, the chronicle goes like this:

Early in the Cenozoic Age, some sixty million years ago, squirrel-like mammals were open to ways to increase their chances of survival within a rather meager environment. These were insect-eating creatures living mainly off the land that had recently undergone severe change, the probable result of an asteroid impact that nearly extinguished life itself. The dinosaurs had vanished by that time, but life at ground level was still a challenge for these ancient mammals. The fossil record implies the existence of somewhat larger reptiles that no doubt survived at the expense of the smaller mammals. Fortunately, sporadic mutations and changing conditions granted some of the mammals opportunities to alter their lifestyles.

At about this time, many mammalian species invaded the trees. We know this again because of dated fossils found buried in the dirt. Small, furry, with large eyes and grasping hands, they were undoubtedly searching for more food (especially fruit) while trying to escape fierce competition prevalent on the ground. Some of those species found the trees even rougher going and thus became extinct. Others discovered the trees to be to their liking, surviving famously. A few, like the tree shrews of Southeast Asia, still thrive today, their traits just-in-time well adapted to life aloft. In fact, the trees became a whole new niche, helping to transform these creatures from ground-dwelling, insect-eating mammals to tree-dwelling, banana-eating prosimians. These protomonkeys are the least advanced members of the order of primates, a zoological category to which both apes and humans belong.

Fossils reveal untold refinements in life-forms, lifestyles, and life-and-death issues, as many successful prosimians adjusted to the best available niches within constantly changing environments. Generation after generation of natural selection gradually changed paws into hands. Stubby claws eventually became flexible fingers. And the opposable thumb took shape as a superior tool for maneuvering among the branches. These were not anatomical changes that occurred as individuals grew during single lifetimes. Rather, they were genetic changes endured over the course of millions of years as the environment promoted some traits, rejected many others, and generally made life miserable for most. Favorable mutations gave those prosimians with good balance, keen eyesight, and dexterous hands and fingers a naturally increased opportunity for survival within their newly found tree-based venue. Some creatures more than others excelled in jumping, leaping, swinging, clinging, grasping, and scavenging. Those better adapted to the lofty environment not only reproduced more efficiently but also passed their favorable traits on to future generations of offspring. The result, documented by the fossils, was widespread speciation causing the creation of legions of novel tree-dwelling life-forms.

The evolution of accurate sight was a particularly important development. Trees are, after all, three-dimensional, unlike the flat two-dimensional ground. The advantageous trait of smelling on the ground gave way to that of seeing in the trees. Fossils show how, during millions of years and tens of thousands of generations, mutations gradually brought the eyes of some of these tree-dwellers around toward the front of the head, thereby gaining binocular vision and some real value-added benefit. With eyes at the sides of the head, two independent fields of view result, much like the flat perception when placing our nose and forehead against the edge of an open door. Better yet, try catching a baseball or hammering a nail with one eye closed; it's not so easy without a way to gauge depth.

The gradual shortening of the snout and the slow displacement of the eyes toward the front granted some early prosimians an overlapping field of view—called "orbital convergence" of the eye sockets—and thus more sophisticated, stereoscopic vision. Depth perception, in particular, enabled them to estimate distances more accurately among the branches. Clearly, these versatile ancestors of about fifty million years ago had acquired distinct advantages in the struggle for survival. They had become monkeylike creatures—small, four-legged, yet highly mobile. A major new evolutionary path had originated.

Fossils also disclose that some species of monkeys gradually became larger. Again, a single generation of a given species did not suddenly balloon in size because food was plentiful and life in the trees sedentary. Rather, sporadic mutations in their DNA molecules, spread over scores of generations, gave larger monkeys advantages in the competition for survival—not the least being that larger, more aggressive males have clear superiority over smaller ones in the sexual competition for females. Also, bulky bodies usually provide additional protection from predators. On the other hand, large size is not a wholesale advantage. Sheer mass brings some problems too. Bigger monkeys, for instance, find it harder to hide, and they also need more food to survive. Both advantages and disadvantages accompany most genetic changes. Only when the advantages outweigh the disadvantages are there enhanced opportunities for living.

Those better adapted to the lofty environment not only reproduced more efficiently but also passed their favorable traits on to future generations of offspring.

The ability to grasp a branch securely while simultaneously extending an arm to secure food also provided an important advantage at the time. Manipulative fingers and opposable thumbs were already favored by natural selection for some tree dwellers. In fact, fossil evidence suggests that advances in grasping ability preceded the evolution of binocular eyesight (much as we shall note later in the Cultural Epoch that manual dexterity among early hominids preceded the enlargement of the human brain). Those species whose members were unable to cling well enough to hold on, plunged, died, and became extinct. Those without long enough arms to reach the food starved, died, and also became extinct. The obvious advantage was had by those prosimians able to coordinate clinging and grasping *simultaneously.* Of course, being smart enough to repel attacks from an array of enemies didn't hurt—the paramount issue of rising intelligence to be examined shortly.

The elevated tree environment of forty million years ago thus became a fairly comfortable niche for some species of monkeys. These well-adapted creatures could have probably remained on high indefinitely if a problem had not arisen. Fortunately for us, change stirred yet again. Otherwise, we wouldn't be here.

Leisurely life in the trees thenceforth became troublesome. The ancestral monkeys of forty million years ago were so cozily accustomed to their tree-dwelling environment that they multiplied faster than many other species stuck in harsher environments. Time not used trying to survive can be most agreeably spent trying to reproduce. The sexual urge seems an innate biological tendency dating back literally hundreds of millions of years. The result was probably a population explosion, the type of crisis often inevitably followed by a food shortage. Consequently, the prosimians likely survived only by using their limited ingenuity to find new sources of food. And for some, that meant leaving the trees and returning once again to the ground.

Some monkey species elected to stay in the trees; their genes changed unappreciably. Most of those species eventually became extinct, though some still survive in altered form today. Baboons, gibbons, orangutans, and many other modern tree-dwelling creatures are descendants of the well-adapted monkeys that remained in the trees. By contrast, those prosimians that successfully abandoned the trees embarked on a whole new evolutionary path—a path leading to the progressive primates, including humankind itself.

Aegyptopithecus, a species whose fossils were first discovered decades ago near Cairo, is widely considered to be a good candidate for the ancestral creature common to the Old World monkeys and the lineage that led to both apes and humans. The now-extinct aegyptopithecines of thirty million years ago were as big as large cats or small dogs, with long tails and moderate snouts, yet they stood apart from their predecessors by virtue of their notably developed cerebral volume, equaling several percent of our present human brain. Resembling today's lemurs, they dwelt in forests, not venturing too far from the protection afforded by the trees, yet they were also likely venturesome, exploratory individuals. They foraged for food mainly on the ground, all the while gradually evolving larger brains and rudimentary bipedalism, as well as the roots of social and communicative skills needed for the onset of culture nearly thirty thousand millennia later.

At first notice, it seems foolish for some of the mammals to have taken up residence in the trees about sixty million years ago if some of their descendants were just going to have to come down out of those trees a few tens of millions of years later. But a critically important change occurred while they were aloft: evolution played out in that much more challenging three-dimensional environment. When the tree-dwellers returned to live at ground level, they were equipped with qualities that probably would not have been naturally selected had they originally stayed grounded. With their manual dexterity and binocular vision, among other assets, they were far more advanced than any other type of life then on the ground. In short, had our prosimian ancestors not taken refuge in the trees, they might never have developed several exquisite qualities, many of which we now use, for example, while writing and reading books like this one. The thirty-million-year detour into the trees was well worth the time and effort—so say we humans who have roundly benefited from it.

By twenty million years ago, the ground-based prosimians had become the dominant life on the planet. Our ancestors of the relatively recent past had become more agile, versatile, and smart. They faced few roadblocks in their rush along evolutionary paths that eventually led to a variety of peculiar animals now inhabiting our planet, not least streetwise human beings.

Fossils, in principle, ought to tell what and how evolution happened among all life-forms. Additional fossil finds will eventually fill the transitional gaps that now hamper a full understanding, all while the fossil record grows more comprehensive. But, in practice and at present, that record is surely incomplete and will likely remain so indefinitely. For how would we know if or when our knowledge of the fossils is whole? More key fossils might always be found yet currently lurk in the rubble awaiting discovery.

Genes should also be able to determine the ways and means of evolution. In fact, most geneticists feel so confident about molecular research now underway in biotech laboratories that they someday expect to put paleontologists out of business. Those bioscientists who work with atoms, molecules, and genes often claim a decided advantage over those who work with bones, stones, and whole organisms. Some geneticists even regard as irrelevant the many past decades of paleontological studies of the bumps, grooves, and uncertainties of ancient fossils. This ongoing feud between evolutionary and molecular biologists has raged for half a century since DNA's double-helical code was broken and much of the mystery of life solved.

Just as DNA samples have become the evidence of choice in criminal courts—even more so today than fingerprint identification—genetic studies, it would seem, promise to bring clarity and objectivity to a field often mired in subjective interpretation of withering bones and ancient artifacts. With genetic mapping, there's no digging for fossils, no reconstruction of skulls or skeletons, just the need to collect small samples of blood. Since variations among DNA nucleotide sequences in a modern population is a summary of events of the remote past, comparing those sequences should enable biologists to construct an evolutionary tree of sorts. In other words, since genes change naturally with time, researchers can use genetic differences among species to test how long ago those species' ancestors split apart—provided that genes mutate at a relatively constant rate, which is tantamount to saying that we know how fast the "molecular clock" runs. Metaphorically, each mutation represents a "tick" of the molecular clock, and therefore the greater the difference between two species, the more time has passed since they diverged from a common ancestor. Although genetically mapping current diversity among living things is straightforward, extracting information about long-dead life-forms is more challenging—and this alone will likely keep paleontologists in business for the foreseeable future. In fact, most paleontologists counter the geneticists, claiming that molecular clocks are inaccurate since some of them tick with different rates in different lineages and at different times.

Huge databanks have now been created as part of the Human Genome Project, an international collaboration to decipher the number and sequence of nucleotides among all of humankind's thirty thousand or so genes—in all, some three billion nucleotide bases, or enough A, T, C, and G's to fill several hundred telephone books. (By contrast, typical bacteria each have about a thousand genes and typically a few million bases.) For some scientists, this project—the last major science goal of the twentieth century—was as spectacular a triumph as the landing of men on the Moon. Its size and scale (and expense) were unprecedented for biology, indeed its achievement changed much of the way biologists now ply their trade at the start of the new millennium. We have likely exited the "century of physics" and entered the "century of biology," with major implications for medicine and society. Alas, though an average human genome might be a thorough description of our DNA, it's by no means a complete explanation of what makes us human. Furthermore, there's no "the" human genome, as everyone's genome differs slightly lest we all be identical. The book of life needs to be read, interpreted, and comprehended for a full understanding of who we really are, individually.

Genomes have been constructed for dozens of other living species as well, with thousands more to come in the next few years, altogether granting enormous promise for appreciating the coevolution of life and Earth. Preliminary studies already suggest that gene sequences shared by different groups of organisms can be used to infer evolutionary relationships—hence the new interdisciplinary field of phylogenetics, which during the past decade has begun to challenge widely held notions about the history and evolution of life on Earth. As noted earlier in this Biological Epoch, these are the sequence-based methods that have caused biologists to revamp recently the way all of life is classified—namely, the three domains of bacteria, archaea, and eukarya, the first two of these being prokaryotic and the last eukaryotic, thus including all known animals. DNA sequence data are also questioning the conventional wisdom of evolutionary branching, especially those times when speciation occurred in the ancestral lineages of many groups of higher organisms. Whether attempting to determine the past time of common ancestry of hippopotamuses and whales (about sixty million years ago), or perhaps that of giraffes and antelopes (about thirty million years ago), or even that of humans and great apes (about ten million years ago), geneticists are questing to better decipher the mode and especially the tempo of evolution. However, their molecular techniques are new and their results often at odds with the traditional methods of paleontology, hence more discord between what should be complementary avenues of historical enquiry.

Take chimps, for example. Studies of selected genes from our closest relative have consistently found that about ninety-eight percent of human DNA is identical to that of chimpanzees (and more than 99.9 percent of yours is identical to the person next to you). That means there are fewer genetic differences between chimps and humans than between horses and zebras or between dolphins and porpoises. Only a small part of the human genome is responsible for the traits that make us human—including our ability to walk, talk, write, build complex things, and enact moral imperatives. That said, chimps look and act like us only to an extent; their anatomy and behavior do distinctly differ from ours. (Admittedly, a two percent difference among an estimated billion base pairs does still allow for millions of variations among strings of nucleotides that govern protein manufacture.) So, what constitutes those two-percent biochemical differences, and can we trace them back to their genetic origins—looking back in time to infer evolutionary insight? If recent studies are correct, it's not only the number of gene differences that's telling but also the relative activity (or "expression" by which they produce proteins) of certain genes. Evidently, gene expression in human brains differs greatly from that in chimps, implying faster rates of neural evolution for our ancestors while on the road to humanity.

Now take living apes and monkeys. Differences in DNA sequences between these two contemporary life-forms can help locate when in time they last shared a common ancestor. First, the gene differences themselves indicate how closely or distantly the primates' lineages are likely to be on the evolutionary tree. Second, using a molecular clock that specifies the rate of genetic mutations, the past time when those lineages split apart can be estimated. The answer is about twenty-five million years ago, in this case in good agreement with the fossils of archaic apes and monkeys. By contrast, other phylogenetic studies disagree with those of paleontology, some greatly. Most notably, for the Cambrian explosion when life greatly diversified among many major animal types, the fossils suggest that animals appeared abruptly about five hundred million years ago, whereas the genes imply origins roughly twice that old and a good deal more gradually. Could tiny and squishy animalistic creatures have existed for hundreds of millions of years before leaving any hard evidence, or is the fossil record merely incomplete? Generally, molecular data imply older ages than fossil data, including a potential origin of life itself dating back six or seven billion years—which is either nonsense given Earth's younger age, or highly significant if life did arrive from space already intact. Such molecular clocks, admittedly subject to assumptions and uncertainties regarding mutation rates—assumptions, ironically, that rely on the fossil record for calibration—nonetheless promise to provide evolutionary information where the fossil record is fragmentary or missing.

In the end, both genetics and paleontology will surely be needed to build an intricate evolutionary tree, or bush, detailing all the many varied paths from life's origin to the present. Much as astronomers often find useful both optical and radio observations of, say, galactic nebulae or newborn stars, both these subjects—genetics dissecting biology reductionistically from the bottom up and paleontology treating it more holistically from the top down—will together yield insights into the ways and means that Nature mixed chance and necessity to give rise to novel, complex life-forms. Only then shall we know our deep roots in deep time, based on a deep understanding provided by both microscopic and macroscopic studies of our ancestral origins.

∞

One of the most remarkable aspects of life is its awareness of its surroundings. Unlike nonliving matter, life can monitor impressions from, and respond to, the outside world. Through life's various senses—hearing, seeing, smelling, touching, tasting—all organisms acquire and file vast amounts of information. The extent to which beings are successful in doing so depends largely on their complexity. Organisms manifest this complexity best by means of one exquisite piece of matter—the brain. The brain is the central clearinghouse of all animated acts.

As these words are first written and then read, matter within our skulls is full of electrical impulses. Silently and efficiently, millions of nerve cells pass messages back and forth within our brains. These microscopic neurons guide our eyes along this printed line, quickly scanning the shapes of the letters. By matching them against memory, we recognize clusters of letters as words and often know their meaning.

Different sensory organs transmit signals that stimulate the brain, which then reacts by sending, in turn, instructions to the muscles. Nerve cells constantly interchange these signals in our heads, ordering our hearts to beat, our lungs to pump, and our hands to get ready to turn the pages of this book. The body's nervous system, of which the brain is the paramount part, controls all mental and physical activity. In fact, every thought, feeling, or action begins in the brain. All human behavior is controlled by it.

Most amazing of all, these silent, unfelt activities inside our heads make us aware that we are now thinking about them: the human brain can contemplate and explore the human brain. That alone makes our brains the most complex clumps of matter found anywhere thus far. The brain is Nature's most tantalizing, talented, and versatile creation—the ultimate example of the extent to which matter has evolved in the *known* Universe.

Brains are made of cells just like any other part of the human body. Each adult human brain has nearly a billion neurons per cubic centimeter, or a total of a few hundred billion neurons in a typical cranium—roughly the same number as stars in our Galaxy. Though neurons come in assorted sizes and shapes, their bulk properties are nonetheless similar. In addition to the main cell body that contains the biological nucleus and manufactures protein in the usual way, neurons also have numerous long and wiry extensions resembling roots of trees. The main, thick extrusion on one side of a neuron is the axon, which acts as a transmitter of information, carrying signals away from the cell body. The network of thin extrusions on the other side of a neuron are collectively the dendrites, which act as microscopic antennae, picking up signals sent by other neurons and carrying those signals to the cell body.

Axons and dendrites enable neurons to "talk" to one another in order to monitor and control the many diverse functions of an intelligent being. On average, each neuron communicates directly with a thousand others. Together, neurons form an intricate network of trillions upon trillions of intercommunicators, each performing a function either assigned by heredity or learned by experience. That network of neurons in a single human brain, when examined through a microscope, strikingly resembles the wisps and filaments, when seen through a telescope, of the largest-scale structures in the Universe.

How does this communication system work, and how quickly? Much like a series of electrical circuits, each neuron passes along a signal from one place to another. When a neuron is stimulated by some external effect—a touch, sight, sound, smell, or taste—the charges on some of the atoms and ions in that neuron changes. This rearrangement of charges can quickly alter the voltage of a neuron, thereby launching an electrical impulse. Neurons, in effect, act like chemical batteries, discharging rapidly in a burst of electricity. They can then recharge themselves in a fraction of a second. All this electrical activity requires energy—energy that is derived by absorbing oxygen during respiration.

Modern research has proved that electrical signals travel swiftly through neurons (in mammals) with a velocity close to a tenth of one kilometer per second, or about two hundred miles per hour. That's reasonably fast compared to everyday speeds, but much slower than an electric current traveling along a metal wire and slower still than the speed of light. A good analogy might imagine information traveling across a neuron, and from one neuron to another, like a rapidly lighted fuse.

Why so fast? Speed is essential to get information to the brain and then back to the appropriate muscles in order to respond to incoming signals. Since the speed of information depends mostly on the diameter of the neuron, some life-forms requiring extremely quick responses in order to escape their predators, and hence survive, have developed surprisingly thick neurons. Sea squids, for example, have neurons with diameters about a hundred times those in humans, enabling them to coordinate movement away from a site of danger or toward a source of food with a system resembling jet propulsion.

Not all neurons in our skulls are physically, or "hard," wired together; in fact, none of them are. Instead, a small gap, called a synapse, separates an axon of one neuron from a dendrite of another. Such synapses need to be magnified many times to be seen, as the gaps themselves measure no more than about a tenth of a micron, or nearly a thousand times thinner than the width of a human hair.

For two neurons to communicate, information must jump the synaptic gap between the axon transmitters and the dendrite receivers. However, this information is not passed along by emitting electrical impulses across a synapse. Instead, an electrical impulse traveling the length of an axon induces the axon to excrete chemicals, known as neurotransmitters. These chemicals then spread across the synapse and cause a new nerve impulse to begin in the next neuron.

About a dozen neurotransmitting chemicals have been identified by modern medicine. Each can, under certain circumstances, inhibit or enhance the voltage on a nearby dendrite. Consequently, this type of unconnected wiring scheme engenders enormous complexity—much, much more than would be possible if the neurons were physically wired to one another. Each neuron can have as many as two hundred thousand synapses, and each of these might or might not trigger an electrical impulse in any given circumstance. And since nearly a quadrillion synapses inhabit a typical human brain, the number of possible routes any electrical impulse can take is mind boggling—no pun intended!

To sharpen the explanation and yet make this topic even more snarled, experiments have demonstrated that these electrical impulses actually travel along a thin covering outside each neuron. Made of a fatty white substance called the myelin sheath, this covering apparently serves as insulation, much like the rubber wrapping around ordinary electrical wires, preventing the network of neurons from short-circuiting. Unfortunately, the myelin sheath erodes in some people; multiple sclerosis attacks myelin, causing some circuits to misfire, which, in turn, produces jerky movements due to uncoordinated timing. It is the complete lack of myelin in newborn infants that prevents their neurons from working in coordinated ways. The result is a child's gradual ability to crawl and then walk, while the myelin grows during the first year or so after birth.

A final note of caution: The excretion of neurotransmitting chemicals largely dictates human behavior. These natural chemicals in our brains can be affected by what we breathe and eat. Poisons and drugs in particular—strychnine, tranquillizers, LSD, amphetamines, marijuana, and many others—change the brain's firing mechanisms, thus changing human behavior. Even caffeine, in coffee-cup doses, lowers our synaptic thresholds so that a tired nervous system, although mostly depleted of transmitters, can keep us alert just a little longer. The chemistry of synapses might underlie many societal ills, although today's neurobiologists are really only beginning to explore how chemicals specify a person's response to circumstantial change—whether genetic or environmental.

With that briefest of introductions to brain structure and function, we return to explore the ways that simpler, ancestral life-forms might have evolved the complexity now resident inside our human heads. Speculation about the paths along which intelligence originated and developed relies mainly on the fossil record—that remarkable evidence written in the stones.

The one-celled amoeba is the most primitive eukaryotic form of life known in the contemporary world—perhaps the most primal of all save the virus, which sometimes acts alive, as noted previously in the Chemical Epoch. Roughly halfway in size between an atom and a human, the amoeba has poor awareness and coordination. It generally responds only at the point stimulated, communicating the information sluggishly through the rest of its body. Although amoebas have developed a crude nervous system, living things that aspire to be more agile—and smarter—surely need quicker internal reactions.

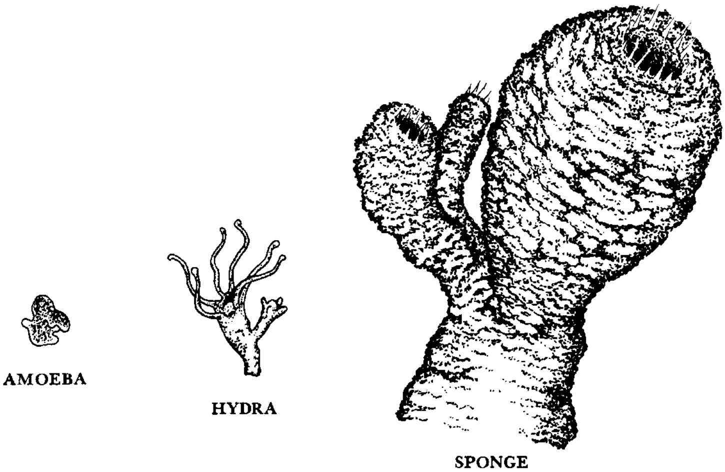
Other single-celled creatures have managed to develop primitive intercom systems. The microscopic paramecium, for instance, has an array of oarlike hairs enabling it to move rapidly through water. The "oars" must act in a coordinated manner, for if they functioned independently the paramecium would make little progress. The hairs are regulated by minute nerves that respond to chemicals emitted within the cell. In this way, messages can be transmitted swiftly and precisely from one part of the cell to another.

Paramecia clearly have more "intelligence" than amoebas. An amoeba searches for food essentially by drifting into water-plant algae. Finding none, it often repeatedly gropes toward the same alga, even though the alga offers no satisfactory food. The amoeba has no memory. A paramecium, on the other hand, has better coordination and a memory of sorts. Having found no food near one alga, it will back off and seek resources in another direction. Paramecia momentarily retain traces of experience.

Compared to the amoeba, then, the paramecium is a genius. But it's a genius operating in a watery world less than a few millimeters across. Even paramecia are unaware of anything beyond this range. No unicellular creature can be much smarter, for it can develop no further.

Despite having complexity well beyond that of any inanimate object, a single cell can boast only the simplest intelligence. To become smarter—that is, to evolve an intricate nervous system—a single cell would need elaborate sense organs to inform it, as well as developed muscles to implement its instructions. Why can't there exist, then, larger cells incorporating these added features, perhaps equipped with miniature hands, eyes, and brain? The answer is that single cells cannot become much larger than 0.001-centimeter creatures. Should they try to do so, their surface areas would increase with the square of their size (1, 4, 9, 16, . . .), whereas their masses, which must be fed through the cells' membranes, would increase as the cube of their size (1, 8, 27, 64, . . .). So cells cannot become too large, lest they starve. The basic smarts of unicellular life-forms are therefore limited; their physical size prevents them from developing the many and more complex organs needed for higher intelligence. Mutations have undoubtedly helped them try every conceivable means to do so during the past three billion years, but they have failed.

The road to greater intelligence required many cells. But quality counts here too, for quantity is not the only issue. A haphazard accumulation of many independent cells will not do; clusters of hundreds of self-sufficient cells are hardly more intelligent than one cell. Consider a sponge, for instance, much like those harvested for use in our bathtubs. Though a sponge is multicellular, most of its millions of cells act independently. A sponge has no central nervous system, thus its "intelligence" is not much more than that of an amoeba. For some reason, sponges failed to profit by their multicellularity. As a result, they have produced no higher forms of life. Sponges are examples of life-forms that long ago reached an evolutionary dead end.



The road to greater intelligence required many cells.

What was needed was a favorable mutation allowing an accumulation of many cells to work together as a community. Interactive, multicellular organisms do have some clear advantages, not least of which is that they avoid the surface-volume problem just noted. More importantly, groups of cells within a multicellular organism can embrace particular functions. This division of labor was one of Nature's greatest inventions. One group of cells might be highly sensitive to foods; others, more efficient in carrying oxygen; still others, tough muscular entities or protective skin casings. The net result was that each group of cells within a multicellular organism became more skilled in one capacity and less so in the rest. Specialization emerged. Accordingly, the total intelligence of such an organism greatly increased as cells, working as a team, became better able to protect themselves from predators and to obtain the food needed for survival. These were the first steps toward a symbiotic society.

The hydra is a good example of a multicellular system that did evolve some intelligence. No larger than a toothpick, the modern hydra resembles a stalk of celery, closed at the lower end and raveled into writhing appendages at the upper end. In contrast to any sponge, the hydra can move its entire body in coordinated fashion to, again and among other things, avoid danger and seek food. In short, the cells within a hydra can communicate. And communication is the essence of organized intelligence.

Cells able to communicate—nerve cells—probably formed originally near the surface of multicellular life-forms such as hydra or, more realistically, hydralike progenitors. Being exposed, these cells had the greatest opportunities to sample their environment. But being near the surface also made them more vulnerable. So, mutations and natural selection likely favored those hydralike ancestors with deeply rooted nerve cells. Over the course of generations, these cells gradually retreated inside the organisms yet kept their link to the environment by sprouting expendable tentacles that reached the surface of the organisms and sometimes beyond. These miniature octopuslike tentacles became the dendrites of modern neurons, the specialized cells that communicate information in more intelligent beings.

As evolution advanced, the bulk of the neurons withdrew ever deeper within multicellular organisms. Eventually, the buried neurons merged, forming clumps of interacting nerve cells—the first and most important step in the building of a central nervous system. This clustering of neurons was one of the greatest of all evolutionary breakthroughs. Once that barrier was crossed, roughly a billion years ago, our hydralike ancestors, as well as other sophisticated organisms like them, were on their way toward generating all of Earth's brainy animal life-forms, including humans.

What does the fossil record say about the evolution of the brain? In the main, it shows a clear ripening of the central nervous system, with organisms branching out in many directions while trending toward greater complexity. Most of these branches, or evolutionary paths, however, represent organisms that either became extinct long ago or survived only as dead ends. The extinct ones are obvious, for their presence simply terminates in the fossil record. The dead-ended ones are just as clear, yet far more interesting. Apparently, at some point in their evolution, insuperable biological obstacles meant that some organisms, such as the amoeba, paramecium, sponge, hydra, as well as worms of all sorts, made no further advance yet survived. These are the invertebrates, or backboneless organisms, many of them skilled and crafty in their own domain. Spiders, for instance, are marvelously accomplished performers within their particular environment; their nervous systems are clever and effective in their limited worlds, their sense organs even more varied and subtle. Bees, wasps, ants, and moths also have highly refined bodies for dealing with their specialized needs. Some—especially bees and ants—even have impressive social organizations that rely on symbolic communication.

Virtually all these invertebrate animals have reached evolutionary dead ends. They are trapped in endless cycles of perfected daily routines. Fossilized spiders of a hundred million years ago show little variation from their modern descendants. Bees in the bush, spiders in the shed are, in a sense, living fossils.

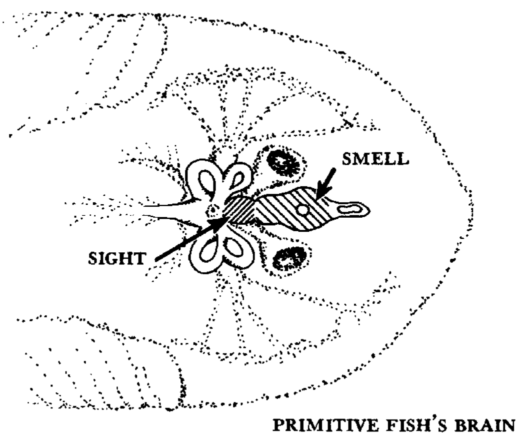
Invertebrates are successes and failures at the same time. On the one hand, they are fabulously talented within their own restricted environments, such as the deerfly that outpaces the fastest animal, the flea that jumps a hundred times it own height, and the octopus whose eye is exceptional among the invertebrates. Successes certainly, for the invertebrates dominated Earth for nearly a half-billion years. But failures, too, because they neglected to develop the vertebral column of bones so conspicuous in fish as well as humans—bones that form the spinal column and protective skull of more complex species.

As humans, we take brains for granted. But the vast majority of animals are invertebrates and have no true brain, no *centralized* nervous system. Most invertebrates' neurons are diffusely spread in a network of fibers throughout their bodies, reminiscent of the distinction made earlier between simple colonies of unicells and more complex multicells. As such, they cannot be creative, adventurous, or visionary—at least not as we have come to know these qualities.

Humans and our fellow vertebrates (backboned fish, reptiles, and mammals) are anomalies to the great invertebrate failure. Vertebrates that did evolve skeletonized parts are but a minor offshoot from the vast, teeming world of the invertebrates. It seems that brains are the exception, not the rule.

The vertebrates' foremost property, aside from their telltale backbone, is their central nervous system. Even so, and like the invertebrates, many vertebrates were apparently unable to utilize their sensory and motor organs to full capacity. A vast array of fish, amphibians, and reptiles, including modern versions of many birds, lizards, snakes, crocodiles, turtles, and many other vertebrates, dead-ended long ago. A good number became extinct, and even the survivors seem to have been unable to decide on a division of authority between the "sight" and "smell" neurons.

Skulls of primitive fish have been reconstructed in some detail from the fossil record. These fish lived several hundred million years ago and are among the simplest known true vertebrates. Though crude, their brains nonetheless contained all the essentials present in modern fish as well as humans. The grouping of small organs caused a bulge toward the snout and was the precursor of the much larger cerebral hemispheres in humans. Likewise, their eyes caused another bulge farther back, the forerunner of our occipital lobe, on which we "see" images projected at the rear of our brain. Lateral-line organs also branched out to the side, antecedent to our cerebellum where our body movements are refined and coordinated. These ancient sense organs, though not in themselves rivaling those of some modern invertebrates, were employed more effectively because of their connection to a unified central nervous system.



Skulls of primitive fish have been reconstructed in some detail from the fossil record.

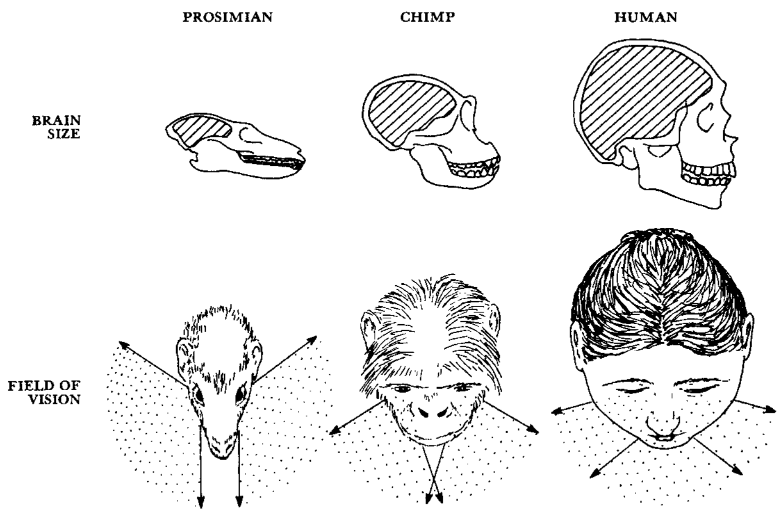
The development of specialized sense organs, and especially their integration with a centralized brain, aided the intellectual dominance of the vertebrates over the invertebrates. Complexity—as often exemplified by the eye, which originated as a photosynthetic organ whose initial purpose was to use light as a source of energy but that eventually evolved into photoreceptors to use light as a source of information—rose ever more.

Sight certainly did play a major role in the advancement of these early vertebrates, as the fossil record documents the ripening over time of a relatively large visual brain. Mutations doubtlessly gave an advantage to certain species of fish, enabling them to utilize improved eyesight to move, survive, and reproduce better in the water. The sense of sight did not rule unchallenged, however. The sense of smell remained a keen rival in the ever-refining evolution of Earth's life several hundred million years ago.

Competition between sight and smell continued with time's passage. When the amphibians transferred from the sea to the land, the flood of two-dimensional sight data probably overwhelmed even the crafty brains of these gifted vertebrates. Smell input, on the other hand, more one-dimensional by comparison, was still within the grasp of such a brain. Accordingly, the first amphibians likely found smell to be of more practical use than sight. The fossil record does show how the occipital lobe shrank while the cerebral hemispheres expanded over many generations. Gradually, the sense of sight regained greater usefulness as the brain of the mammals grew larger through continued mutations and natural selection. The multitude of out-of-water images no longer saturated once-oceanic eyes; in fact it was the eyes that caused the brain to ramp up in size, speed, and sophistication in order to process the incoming information. The larger brains of the mammals were then able to cope with the full world of sight as well as sound. Those creatures having more intricate, complex brains were better suited to survive in a changing terrestrial environment.

The fossils also depict this decline of the importance of smell. Although the sense of smell was of greatest value to the lower vertebrates, as the brain increased, other senses, such as seeing and hearing, became equally advantageous and eventually more so. The eye, in particular, seems to have played an essential role in the maturation of intelligence. Our much larger human cerebral hemispheres are indeed derived from the ancient smell brain, but the preeminence of this sense was long ago surpassed by sight, sound, and other general sensations.

The most recent step in the evolution of the brain occurred in the mammals. Once again, the search for energy efficiency was at the core, as the most successful mammals developed multicomponent hearts that permit more complete oxygenation of the blood, warm-bloodedness that allows sustained activity in cold environments, and external fur that effectively conserves energy. Basically, every living tissue needs a minimum amount of energy to function, and the brain is the most energy-demanding tissue of all—which probably explains why true intelligence is found only in warm-blooded animals, since brains are metabolically intensive organs and do have high energy needs (per unit mass).



. . . more accurate eyesight gave our ancestors distinct advantages along the evolutionary path toward greater complexity.

Also once again, many evolutionary dead ends are evident, the result of mutations that simply didn't offer much advantage to some species. However, other mutations altered for the better the traits of selected organisms, granting them clear advantage in the overall struggle for survival. Parts of this brain development can be watched as the head of a human fetus rapidly retraces much of evolution among the vertebrates: Two weeks after conception, the embryo's (reptilian) brain resembles that of a frog with the olfactory (smell) bulb protruding ahead of a very small brain. Several weeks later, the olfactory has shrunk, while the occipital (sight) lobe has swollen. By many weeks beyond that, several additional neural layers have grown, indicative of evolution having superposed more layers of neurons needed for advanced mammalian functions.

More recent neural advances were driven, over millions of years and generation upon generation, by favorable mutations, which advantaged those mammals with longer arms and gripping paws for leaping, swinging, and reaching food. Other mutations also gradually shifted the eyes of the early prosimians from the side to the front of the head, thereby producing the earlier-noted binocular, three-dimensional sight. In turn, the gradual refinement of dexterous arms and manipulative hands combined with more accurate eyesight to give those proto-monkey ancestors distinct advantages along the evolutionary path toward even greater complexity.

The eye-hand-brain combination had a powerful evolutionary effect, not only for enhancing survival in the near term but also for developing intelligence in the long term—one upshot being the seemingly limitless manual skills and boundless curiosity of *Homo sapiens*, a decidedly intriguing creature to be scrutinized next in the Cultural Epoch.

∞

Natural selection is central to all biological evolution. Darwin's idea was novel and innovative—quite possibly the most significant advance in all of post-Renaissance science. Yet aspects of selection are seen throughout much of Nature—during physical evolution in simpler systems and during cultural evolution in some of the most complex systems. Not that genes are involved in physical or cultural change; not that inheritance and reproduction are prominent for any but biological evolution; not that *natural*, or Darwinian, selection works among inanimate objects. Yet, the process of selection, generally construed, is present all through cosmic evolution, operating in realms well beyond biology. To be sure, selection functions more robustly for living systems than for those nonliving—and probably even more actively, or at least faster, for cultural systems. But selection, in the main, is a common feature at work throughout Nature writ large.

Both to clarify and to stress a point made earlier, the term "selection" is actually a bit of a misnomer, for there is no known agent in Nature that deliberately selects. Selection itself is not an active force or promoter of evolution as much as a passive pruning device to weed out the unfit. As such, selected objects are simply those that remain after all the poorly adapted or less fortunate ones have been removed from a population. The better term again might be "nonrandom elimination," since what we really mean to describe is the aggregate of adverse circumstances responsible for the deletion of some members of a group. Accordingly, selection can be broadly taken to mean preferential interaction of any system—living or nonliving—with its surrounding environment.

Selection works alongside the flow of energy into and out of all open systems, not just life-forms, often providing a formative step in the production of order. In short, all ordered systems are selected by their ability to command energy resources—not too much energy as to be destructive and not so little as to be ineffective. Sometimes, when the energy flow exceeds a critical threshold, thereby driving a system well beyond equilibrium, selection aids in generating newly ordered forms much as described toward the end of the Chemical Epoch.

Nature displays numerous examples of the process of selection operating among inanimate systems, but always in ways simpler than among living systems and always, it seems, in the presence of energy. We noted earlier how prebiological molecules bathed in energy were "selected" in a soupy sea to become the building blocks of life. Certain bondings of amino acids were advantaged while others were excluded, implying that the chemical-evolutionary steps toward life yielded new states more thermodynamically stable than their precursor molecules; all the while entropy increased in their watery surroundings. Selection—call it chemical selection—was clearly working to help tame chance—albeit not the more subtle yet powerful Darwinian, biological selection involving species modification, inheritance, and adaptation.

Crystal growth also demonstrates aspects of selection—call it physical selection—helping to order nonliving substances in ways much simpler than biological selection. To grow an ice crystal, water molecules must collide so that they stick and are not rejected. The initial molecular collisions are entirely random, but once they occur the migrating molecules are then guided by well-known electromagnetic forces into favorable positions on the surface. If the incoming molecule lands at a surface position physically conducive to the growth of ice-crystal structure, it's "selected" to stay and contribute to the crystal; otherwise it's expelled. Its arrival is random, but the result is not.

An atmospheric storm is another example of a physical system undergoing selection while complexifying. Kilometer-sized vortices come and go at random in Earth's atmosphere, mostly the result of swirling updrafts and turbulent eddies caused by winds hitting high-relief surface areas, such as mountains or islands. Patterns of cumulous clouds develop as rising currents form small competing cumuli that draw on the solar energy stored as latent heat in water vapor molecules. Those cumuli able to attract more air flow and thus more energy are “selected,” leaving the others to dissipate. By the end of a typical hot, sunny summer day, selection has fostered the growth of a few large-scale thunderstorms—and sometimes, with adequate moisture and energy, occasional eddies mature into full-scale hurricanes hundreds of kilometers across.

Stars, too, are subject to selection. Our Sun, for example, in about five billion years will become a red-giant star, increasing its gradient of temperature and chemical composition from core to surface. But, as noted earlier in the Stellar Epoch, the Sun will never fuse carbon into heavier elements, never become more complex than an old red giant, and never detonate as a supernova. In short, the Sun will not be selected to evolve much further, since its energy flow will likely never reach those critical values needed for the natural emergence of greater complexity. Although our Sun is not alive by any stretch of the imagination, it will have been nonrandomly eliminated from further stellar evolution.

Many other examples of nonbiological selection pervade the physical world, affecting both matter and radiation, including, for example, galaxies that are “selected” to form in the earlier Universe by means not too different from that described above for hurricanes, and certain modes of radiation able to handle energy coherently that are “selected” for laser propagation in the laboratory. Likewise, as will be noted in the next, Cultural Epoch, selection—call it cultural selection—was just as surely at work among our ancestors. The ability to start a fire, for example, would have been a major selective asset for those hominids who possessed it, as are, in more recent times, dealer competition and customer demand when combining to create selective pressure for better automobiles in the social marketplace.

Selection does operate among inanimate, nonbiological systems, even if not as robustly as for animate, biological systems. Physical and chemical selection obeys well-understood, if statistical laws of physics, whereas biological (Darwinian) selection is, appropriately, richer and more multifaceted, drawing on genetic exchange and vast information storage. Even so, all these selective mechanisms, including accelerated cultural selection in today’s world, help to build order and complexity in basically the same way: They all mix a random initiator with a deterministic response in the presence of energy, a theme integral to the onset of structure throughout all of Nature. Provided we think broadly enough, there is indeed unity among the natural sciences.

∞

This Biological Epoch has outlined the salient features of life and its erratic drift through ages so long as to be measured in millions of millennia. For the first few billion years it remained starkly unicellular, resembling the blue-green algae found today in backyard swimming pools. Eventually, around a billion years ago, cells clustered into groups, coordinated their activities, and became multicellular organisms. Not long thereafter, the fossil record documents what must have been a population explosion in the number and diversity of species.

Change was rampant as life-forms multiplied and speciated rapidly. Legions of fish swam the seas little more than half a billion years ago. Plants came ashore some four hundred million years ago, and amphibians quickly followed, doubtless in search of food. Animals mastered the land about two hundred million years ago, while birds, mammals, and flowers flourished for not quite half that time. By contrast, hominids—the subject of the Cultural Epoch—have endured for only the past few million years, a span so brief that if we were to imagine all of cosmic history compressed into a single year, then earthly humans would have existed for only the past hour or so. In this analogy, our specific species, *Homo sapiens*, would not have emerged until some ten minutes ago.

Darwin was basically right. The fossil record leaves little doubt that biological evolution by natural selection has occurred and is continuing. The rate at which evolution works, however, remains unresolved, as does the possibility that other mechanisms of change are also operative in Nature. Cosmic rays, chemical drugs, intense radiation, or just plain DNA-copying errors enable mutations to accelerate the motor of evolution, altering life's genetic structure and causing some organisms to adapt to new niches in ever-changing environments. In thermodynamic terms, microbes, plants, and animals evolve far-from-equilibrium systems with combinations of properties that are unpredictable in detail. But the outcome isn't all chance, for evolution does have its deterministic component; selection prunes, edits, and decides who is optimally fit for a given set of environmental conditions. Changes, and adaptations to those changes, are the keys to the genesis—and destiny—of all living things.

Admittedly, gaps in the fossil record hinder our complete understanding of life's history, just as missing links hamper our current knowledge of galaxies, stars, and planets. It's not easy to squeeze evidence from stones, and, in any case, genome mapping will now begin to fill in those gaps. To be sure, each day brings new ideas, new tests, new discoveries, and further refinement of our modern conception of biological evolution. And with these advances come greater objectivity, and progress too, while searching the sands of time to decipher erstwhile reality.

The Biological Epoch has sketched the traditional view of evolution, namely, Darwinian evolution via natural selection. Yet life is but a small, albeit important, part of the grander cosmic evolutionary worldview. This book suggests that evolution, universally considered, pertains to much more than mere life on Earth. The word "evolution" has been intentionally used in a broad, provocative way, attempting to capture the process of change on all spatial and temporal scales by means surely including, but not restricted to, biological Darwinism. Within the expansive, all-inclusive scenario of cosmic evolution, general trends are identifiable among Nature's myriad, persistent changes over the course of an impressively long span of natural history, from the origin of time to intelligent life on Earth. Next, in the Cultural Epoch, we shall extend the narrative to include our technological selves, for humankind, too, is most assuredly part of this unfolding and unifying epic-class story.

7. CULTURAL EPOCH

Intelligence to Technology

RISING COMPLEXITY IS AN INTEGRAL feature of cosmic evolution, an outstanding example of which is humankind itself. By no means an anthropocentric statement, our human complexity is clear and demonstrable. Large amounts of information are needed to describe ordered structures like ourselves in a Universe that is otherwise growing increasingly chaotic. We may not be the most well-adapted species on the planet (the microbes probably are), nor those with the greatest potential for long-term survival (the microbes again?), but we are currently the most complex clump of matter known anywhere. There's no denying it.

While approaching the here and now along the arrow of time, we naturally wonder about the evolutionary route that led from our ancient ancestors to modern humans. Not that it was a straight and narrow path, rather more likely branched and convoluted, which is the way evolution works—by fits and starts, with lots of blind avenues and dead ends, among wonderfully adapted new species. Questions flood the mind: Where did we come from? How did humankind emerge from all that went before? What were the circumstances that led to our decidedly odd body shape, our strong behavioral attitude, our desire to know who we are? Specifically, what factors caused the development of our fabulous attributes of thinking with our brain, seeing with our eyes, talking with our mouth, walking on our legs, constructing with our hands, and wondering about ourselves?

Having gained an appreciation for the origins of matter and life, we naturally confront other trenchant questions close to home in time and space: What are *we*? Not what is our Sun, our planet, or life itself, but what and who are these twenty-first-century human beings inhabiting Earth? Everyone ponders these queries at one time or another. They are among the most profound and interesting issues of all—not least because we are asking them.

The seventh, Cultural Epoch doesn't concern human names, social security numbers, or political affiliations, though admittedly these and other vital statistics do tell others a little about ourselves. Instead, we seek a more general understanding of the origin of the human species, which, in a nutshell, seems to be as follows: each of us is the product of many ancestral life-forms—a cluster of genes inherited from all of them and shaped partly by environments that are partly ours, partly our parents', partly our parents' parents', and so on, far back through time.

Tracing back a thousand years, each of us would have had more than a million ancestors, all alive simultaneously. They were likely spread across much of the world, living in a range of environments, most struggling to survive and few better off than any others. Going back another couple thousand years, some of our ancestors could well have been leaders of ancient Greece or Rome; yet another few thousand years, of the ruling class of old Egypt or Babylonia. But the bulk of our forebears were probably slaves or peasants, likely able to neither read nor write. They were probably ignorant, superstitious, and cruel—primitive farmers at best. Few of them would have touched metal or spun a wheel. By modern standards, most of our ancestors of several millennia ago were savages. They survived largely by hunting and gathering, living only within their immediate environment. It's hard to relate to them, but modern science suggests that we must. Evolution stipulates that we carry in our bodies some of their genes, and perhaps in our minds some of their temperament toward the world around them. Part of our anatomy, abilities, attitudes, and desires, as well as our outlook on life and way of thinking, all derive to some extent from the genes of our ancestors, molded partly by the environments in which they lived.

Answers to the fundamental questions about ourselves, then, are still evolutionary ones. These are issues that permit us to relate our individual selves to all of humankind, indeed to all living things. If we can find rational answers, then perhaps we can learn who we really are, as well as how it is that we can think and experiment about ourselves and our Universe.

Long ago, our distant ancestors possessed none of these traits. They were not human. They made no tools, had no intellectual vision, probably displayed little curiosity. They were small-brained beings populating forests, largely intent on surviving and reproducing. Somehow they gave rise to humans. Somewhere in our ancestral past, evolutionary links join creatures that clearly were human with creatures that clearly were not.

∞

Fossils allow scientists to sketch the divergent paths of evolution and to understand the fine lines of distinction between various closely related species. Not surprisingly, fossils of recent times are usually well preserved, enabling researchers to document evolution with some confidence. The older fossils are in poorer condition, often in pieces and sometimes hardly recognizable. Reassembling the pieces is much like working another of those "jigsaw puzzles" alluded to earlier. Trying to decipher where and when the reconstructed fossils fit into the evolutionary line of descent is another kind of puzzle.

The entire effort of unraveling human genealogy resembles the restoration of a gigantic mural painted over the course of millions of years. To the right, where the scene has been rendered recently by modern humans, the message is reasonably clear. In the middle, the mural is soiled, peeling, and generally deteriorating. Painted by our earlier ancestors, most of the central mural cannot be easily cleaned, nor can it be easily repaired to reveal the message once the dirt and grime have been removed. Toward the left, the oldest part of the painting is usually torn and tattered with perhaps pieces missing since it was sketched so long ago. Like any restoration, the discovery process is slow and painstaking, done very deliberately to avoid destroying the mural and thus the message.

Anthropologists and archaeologists examining old fossils and artifacts need the virtue of patience in addition to an inquiring, unbiased mind. Very much like detective work, their research strives to tell a story based on scattered, uncertain data—not unlike our earlier efforts to understand stars and galaxies based on a few hints and clues about their origin and evolution. The story here, however, in this Cultural Epoch, means more than closing another crime case. It means learning about the origin and evolution of humanity itself, a goal that some find more satisfying, indeed more relevant, than deciphering the details of distant celestial objects.

Knowledge of social and cultural advances has accumulated impressively during the past century—an exciting time for science, indeed an era full of startling revelations about our planet and ourselves. Field excursions and site excavations grew in popularity in the early twentieth century when physical scientists began realizing that the ground beneath our feet held clues to the nature of Earth, and especially the nature of former life on Earth. Biological scientists ramped up their discoveries of that life, including a rich fossil record of many ancient creatures that long preceded our existence. And social scientists began uncovering old stone ax heads among rough cutting and scraping implements along rivers and inside caves of western Europe. The latter were crude tools, but tools nonetheless—double-edged, teardrop-shaped rocks symmetrically crafted according to a preconceived plan and now termed Acheulean technology, after St. Acheul, a suburb of Amiens, in northern France, where the first examples were found. Subsequent radioactive dating showed many of them to be nearly a hundred thousand years old. The question arose: Who were the makers of those ancient tools?

Many of the first great fossil finds of prehumans occurred in the mid-nineteenth century. (By prehumans, we mean ancestors of the hominid line, namely, humans and their extinct close relatives.) Around the time that Darwin published his seminal treatise on biological evolution, a primitive-looking skull was found in a cave in the Neander Valley in Germany. This stocky, flattened skull has a low, sloping forehead, a receding chin, and thick ridges over the eye sockets, though it still displays an overall "manlike" appearance. Given that the German word for valley is "*tal*," Neanderthal man became the generic name attached to the original owner of this skull. Though a bit odd compared to today's human skulls, there's little doubt of its human origin. However, with only one such skull fossil then known, it was easy to classify it as a deformed specimen of modern man, which was exactly the approach taken at the time. Even as recently as a hundred years ago, many biologists who embraced the evolution of plants and simpler animals were still unwilling to accept this skull's evolutionary relevance to humans.

Throughout the twentieth century, many more Neanderthal-type fossils were found at dozens of sites in Europe and western Asia. In addition, less primitive humanlike skulls were excavated from many places scattered across Eurasia, initially near the French village of Cro-Magnon. These newer skulls are among those belonging to Cro-Magnon man, an entire subspecies of human ancestors. Regardless of the designation, the important point is that many of these odd-looking, though clearly human-related, skulls were unearthed alongside the ancient Acheulean tools. Their proximity clearly implies that toolmaking humans of some sort resided in Europe a hundred thousand years or so ago. Since the Cro-Magnon skulls are more modern and their skeletons more slender than those of the muscular Neanderthal variety, Neanderthal man is assumed, in some quarters, to be the ancestor of Cro-Magnon man; the two may have interbred and coevolved into modern humans. Other anthropologists demur, claiming that the Neanderthals represent a divergent branch of hominids that died out (or were killed off) about thirty thousand years ago; this majority view is supported by recent genetic and anatomical studies. Whichever it was, and for whatever reason, the Cro-Magnons replaced the Neanderthals some three hundred centuries ago. A pivotal question then becomes: Who were the ancestors of Neanderthal man, whose fossils date back some three hundred thousand years?

Rich fossil troves trace our roots much farther back in time. Humanlike skulls and teeth were eventually discovered far from Europe, in an arid riverbed in Java, a large island in Indonesia. These remains—initially one skull cap, a thighbone, and a single molar—date back nearly a million years and seem even more primitive than those of either the Neanderthal or Cro-Magnon people. Yet the size, shape, and overall features of the Java man bone fragments still resemble those of today's humans. Moreover, the hole at the base of the skull, through which the spinal cord passes, is positioned in such as way that those creatures must have stood erect.

Astonishing findings hardly more than a half-century ago, these old humanlike fossils predictably drew a great deal of skepticism. It's understandably hard for us to imagine that erect, humanesque creatures could have lived anywhere on Earth as long ago as a million years before the present. That's a terribly long time by human standards, equivalent to some forty thousand generations of human life. In fact, a million years is more than a hundred times longer than all of recorded history. Put another way, more than ninety-nine percent of humankind's history is told almost exclusively by its fossils.

Confirmation of these startling results followed when many similar fossils were exhumed at widespread sites throughout the temperate zones of our planet. Diggers have now uncovered numerous Java man skulls, as well as bones of Heidelberg man in Germany, Peking man in China, and a variety of other ancient though clearly humanlike fossils in Hungary, France, Spain, and Africa. Most of these fossils are on the order of a half-million years old, though some are closer to a full million years and a few might even be older—such as a primitive jaw and partial skull found recently in the former Soviet republic of Georgia and estimated to be not quite two million years old. Significantly, these are not skulls of apes. Nor are they skulls of ape-men. They are skulls of humans—erect men and women who lived an awfully long time ago.

Since humanlike fossils dated to be less than a couple of million years old have skulls and teeth closely resembling those of modern humans, all of them are granted the designation *Homo*, a Latin word meaning "man." And to distinguish these older fossils from contemporary bones, a suffix is often added. For example, Neanderthals, Cro-Magnons, and fossils of other humanlike creatures dated to have lived less than a few hundred thousand years ago are collectively grouped by the name *Homo sapiens*, meaning "wise man." This is the same biological species as modern men and women, though some researchers prefer to endow the most recent humans of recorded history (including ourselves) with yet a special appellation—*Homo sapiens sapiens.* No doubt another expression of human vanity, "very wise man" is a highly debatable label, especially given the plethora of global predicaments we've created for ourselves on twenty-first-century Earth.

In contrast, Java man and other humanlike fossils between a few hundred thousand and a million (maybe twice that) years old are collectively referred to by the species name *Homo erectus*, meaning "erect man." Definitely of human stock, these closely related creatures walked erect and displayed surprising manual dexterity, but their brain volume was not as large and their tool use not as advanced as those of *Homo sapiens.* Several of these subspecies may have coexisted and even possibly interacted and competed. Apparently many different kinds of hominids lived simultaneously in Africa between one and two million years ago.

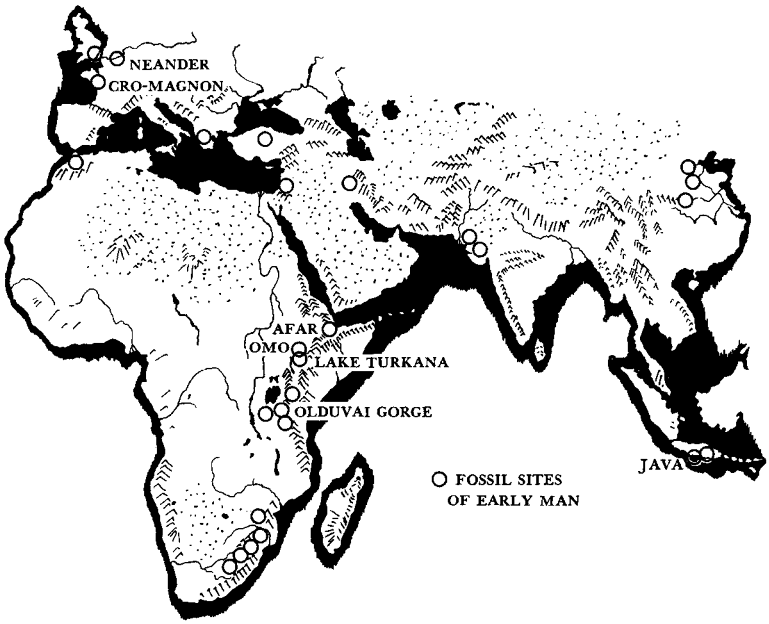
The treasury of humanlike skulls at least as old as a million years doesn't really solve the central issue of human origins. It merely pushes back in time the key question: Who were the ancestors of *Homo erectus*?

Not until the last quarter of the twentieth century—a generation of scientists still working today—has a reasonably clear line of descent emerged. Earlier anthropological expeditions, as far back as the 1920s, provided some of the first clues, mainly rare fossilized skulls having simultaneously human and ape characteristics. More recently, many more hybrid skulls have been found throughout warm climate regions, notably on the African continent. After each skull was carefully dug up, dusted off, and reassembled from pieces, analysis showed them to have the following curious blend of ape and human traits: an interior skull volume (brain capacity) larger than an ape's, though smaller than a human's; a jaw larger than a human's, though smaller than an ape's; a forehead resembling an ape's more than a human's; canine teeth more like those of a human than those of an ape; a skull aperture through which the upper spinal cord passed, implying that this creature had walked upright, or nearly so.

Such a mixed bag of bone qualities strongly implies that this creature belongs, in both place and time, near the threshold of humanity. Fossils of this hybrid ape/human kind have subsequently been given the jawbreaking Latin name of *Australopithecus,* meaning "southern ape." Unfortunately, the earliest findings in the sandy soil of southern Africa could not be dated; sand isn't radioactive and tends to shift with time. But newer discoveries, with firmer dates have focused modern paleoanthropological research on the African continent, where it has been ever since.

Excavations during the past few decades have revealed many additional australopithecine skull and tooth fragments. Some of these findings have been made in the same southern African area where the mobile soil hampers dating. Numerous similar fossils were also gathered all along the East African Rift Valley, a giant crack produced by the disjointed drift of that large continent, and there the ordered layers of volcanic rock can be accurately dated. For example, an *Australopithecus* fossil was noticed protruding from volcanic ash along a dried riverbed at Olduvai Gorge, Tanzania, thus the australopithecine fossils could be set in time. That date is approximately two million years ago, an age estimate since verified by more recent findings. Clearly these protohumans, alias ape-men, inhabited our planet a good long time ago.

The official naming of the two-million-year-old skull remains found at Olduvai Gorge is not without controversy. The discovering Leakey family of Britain and Kenya argue that these are the skulls of a species related to but distinct from the australopithecines. In particular, the codiscovery of primitive stone tools prompted them to propose a new



Numerous fossils have been gathered all along the East African Rift Valley . . .

species designation, *Homo habilis*, or "handy man," and they claimed the utensils these creatures used as products of the earliest known technology, namely, the pre–Stone Age Oldowan period (after the gorge where they were found). However, the rock chips and bone flakes that the Leakeys consider tools are very primitive indeed—much simpler than the crude Acheulean implements noted a few pages ago—making it hard to assess just how handy these creatures really were. Perhaps the *H. habilis* fossils are merely those of advanced australopithecines and not ones deserving of the humanesque status of the genus *Homo*.

This and other controversies have fueled competing theories for the origin of modern humans. One, the multiregional hypothesis, based mostly on fossils, holds that humans arose in several parts of the world as long ago as two million years, thereafter spreading, evolving, and culturally exchanging as a single species. When descendants of *H. erectus* later left Africa, they interbred with hominids already and locally sapienized in Europe and Asia, including Neanderthals, giving rise to the ethnic and racial diversity seen in today's populations. By contrast, the out-of-Africa (or uniregional) hypothesis, based mostly on genetic analyses, posits a much more recent African origin for modern humans, perhaps hardly more than a hundred-fifty thousand years ago—a new species, distinct from Neanderthals and other ancient humans, whom they later replaced without any interbreeding. In support of the latter idea, DNA samples recovered thus far from Neanderthal fossils have failed to provide any evidence that modern Europeans carry Neanderthal genes. Therefore, the molecular clock once again clashes with the dates of many fossilized bones. Like other "either-or" issues in our cosmic-evolutionary story, the truth lay most probably somewhere in between—in this case several migrations across Eurasia, each perhaps interbreeding with or otherwise displacing those prehumans and humans who went before.

The two-million-year-old ancestors were no larger than a hundred-fifty centimeters tall, weighing about fifty kilograms, or roughly five feet and a hundred pounds. Although they were surely smarter than any other life-forms with which they shared the open plains away from the forests, their brains were probably not large enough to have managed speech; rather, they more likely communicated using a repertoire of grunts, groans, arm gestures, and other body movements. The more talented members surely possessed dexterous hands, nimble fingers, and keen eyesight—not as good as ours, but good enough to fashion simple stone tools. The eye-hand-brain combination was again at work, surely to their advantage. Whatever their full attributes, these creatures seem to have adapted well to their changing environments, for adaptation above all else is a key to survival.

More recent fieldwork reveals that at least two, and maybe four or five, types of hominid creatures may have coexisted throughout Africa several million years ago. Hundreds of australopithecine fossils have now been classified into at least two distinct species of prehumans. One of these is characterized by a huge jaw and large grinding teeth, suggestive of a species that enjoyed a diet of mostly coarse vegetation, much like that eaten by modern gorillas. This more robust type is often called *Australopithecus boisei* or *Paranthropus boisei*, or even *A. boisei* or *A. robustus* for short. Paleoanthropologists are notorious for inventing tongue-twisting names, and whether these are different species, subspecies, or variations within a given species, no one yet knows. The truth still lies well-interred as buried bones, for the fossils are too old for genes to help much here. The other type, *Australopithecus africanus*, or *A. africanus* or *A. gracile*, is of the originally discovered southern African variety. This species is typified by a more slender jaw and smaller molars, implying a gentler anatomy and thus a class of prehumans that probably feasted on meat. Those implications are just that—suggestions and not conclusions—but they do represent the prevailing view among anthropologists today.

Given the duality of these findings, we might naturally wonder if the observed differences in the australopithecine fossils could simply be variations of the same species. After all, today's humans display slight, yet obvious, differences; contrast thin hurdlers against husky weight lifters. This interpretation doesn't seem tenable, however, since all the two-million-year-old fossils of prehuman creatures fall into two distinct categories: either skulls and teeth are clearly big and oversized, or they are small and graceful. How about male and female? Could these two types of australopithecine fossils correspond to sexuality? Again, this interpretation seems improbable because the two classes of fossils are hardly ever found at the same place within sedimentary rock; *boisei* is linked to East Africa, *gracile* to southern Africa. Unless highly peculiar behavior among prehuman cultures kept tribes of males separated from females, it would seem impossible that these classes correspond to sexual differences. Besides and obviously, males and females could hardly reproduce their species while segregated.

Thus, at least two species of protohumans, and quite possibly more, apparently shared the same niche on Earth a few million years ago. Presumably, only one of these species is our true ancestor. Further field work tentatively shows which one.

Expeditions throughout the East African Rift Valley have revealed much new information during the past few decades. Besides the rich lode at Olduvai Gorge in Tanzania, several groups are continuing to trace the thread of our origins by examining fossils found along the shores of Lake Turkana (formerly Lake Rudolf) in Kenya. And, before the (human) "guerrilla" war in the 1980s prevented further digging, particularly well-preserved fossils were found in easily datable volcanic rock at Omo, Ethiopia.

Among the recent discoveries at several of these sites, the most interesting is perhaps what's missing: no *A. boisei* fossils are less than a million years old. This more muscular prehuman species rather abruptly disappears from the fossil record, implying rapid and unorthodox extinction. The most popular explanation contends that competition between *A. boisei* and *A. africanus* was inevitable. Each biological niche can ultimately be filled by only one species, yet here were two prehuman species trying to make a go of it simultaneously. Surprising at first thought, it then becomes clear why the bigger, more robust species was the loser.

Despite their larger physique, *A. boisei* found vegetation plentiful, yielding a rather comfortable way of life. Such easy living, however, is not necessarily conducive to rapid evolution toward something more complex—such as an intelligent technological society. The smaller species was almost surely more versatile, quicker, and perhaps smarter. Only basic intelligence could help *A. africanus* capture the less abundant meat needed for survival. As a result, natural selection worked to help generations of *A. africanus* expand their brains, their capabilities, and their niche, all the while apparently crowding *A. boisei* right off the face of the Earth. This idea is supported not only by the documented demise of *A. boisei* but also by the finding that alongside *A. africanus* are often found stone tools, however primitive. Whether these tools are a measure of *A. africanus*'s proclivity for manual dexterity and gradual brain development, or whether they might have been used more directly as weapons to accelerate *A.boisei*'s extinction, is unknown.

As for many evolutionary scenarios, the details have yet to be worked out. Many of those details haven't yet even been excavated, and in any case new fossils still likely lurk below the surface that will push dates further back. Accordingly, several alternative views invite revisions in the evolutionary picture sketched here. For example, there is the issue of whether the tool-using creatures of two million years ago should be labeled *H. habilis* or *A. africanus.* Some workers assert that these oldest "tools" are not tools at all. The controversies mostly amount to semantics yet typify the wide range of interpretation among experts. Other researchers contend that *Homo* dates back not much more than a million years, while a few argue that some species of *Homo* existed at least two, and perhaps three, million years ago.

Recent fossils do seem to push our lineage further back, as skull, tooth, and bone fragments discovered in the Afar lowlands of Ethiopia argue for humanlike creatures nearly four million years ago. Similar skulls with smaller brains and larger canine teeth than ours were found alongside footprints preserved in the hardened radioactive ash near Laetoli, Tanzania, implying that these awfully ancient creatures—the most famous of which is "Lucy," a partially complete skeleton of a teenaged female—must have stood erect. Yet she was also an apelike tree climber, based on study of her elbow and shoulder joints. Accordingly, these fossils probably provide the best evidence for the remote ancestor midway between apes and humans—a missing link of sorts. A whole new species, *A. afarensis*, has been proposed as this common ancestor of *H. sapiens* and the extinct *A. boisei*. Opponents argue that these Ethiopian fossils simply belong to the *A. africanus* species but acknowledge that that already distant ancestor must be pushed even farther back in time. Still others claim these fossils imply a more primitive version of *H. habilis*. Whichever evolutionary viewpoint is correct, these ancient humanlike fossils virtually prove that our ancestors walked erect before their brain enlarged appreciably.

So many different evolutionary paths are consistent with all the fossil data that a cynic might remark that there seem nearly as many possible paths as there are paleoanthropologists. More than a dozen different, yet often coexisting, hominid ancestors have been proffered by researchers, many of whom tend to assign new fossil finds to novel rather than established species. Such was the case recently when archaeologists an-

Evidence of ancestral links.

The jaw structure of the *Austalopithecus afarensis* species (center) has qualities midway between apes (left) and humans (right). The jaw's protrusion beyond the face plane, the extent of curvature of the tooth pattern, and the canine teeth themselves are all intermediate for the *A. afarensis* fossil. These findings suggest a link, or common ancestry, between the great apes and erect prehumans dating back at least four million years. *Source: Smithsonian.*

nounced the discovery of skeletal remains of extinct, diminutive humans, termed *Homo floresiensis*, who lived on the tropical Indonesian island of Flores only eighteen thousand years ago. Each new hominid bone often seems more like a monkey wrench upsetting experts' cherished ideas of our human lineage, and the resulting subjectivity helps to reinforce this social science's reputation as a "soft science." Pirating of data, poaching of fossils, acrimonious debate, even lawsuits have plagued work in this field as fiercely contentious investigators vie to decipher one of the biggest puzzles in all of science. The real problem here is that the current picture of human evolution is based on hardly more than a roomful of partially crushed skulls and broken teeth, most of them found scattered across East Africa, Asia, and central Europe. The whole lot thus far unearthed doesn't contain enough parts to reconstruct a single skeleton of an australopithecine; the oldest complete hominid skeleton is that of a sixty-thousand-year-old Neanderthal.

We need not overly confuse the issue here. Most anthropologists do agree upon a general evolutionary trend: *Australopithecus* → *Homo*, or near man → true man. What's more, nearly everyone concurs that hominids arose in Africa, stayed there for a few million years, and then began colonizing the globe—migrating first into Asia, then Europe, and thereafter the Americas. The controversies, frequently shedding more heat than light, essentially concern details—specific dates, emergence of new species, coexistence of many species, invention of tools, cooperation in hunting, and development of language, among many other humanist issues. The emotion is genuine, for at stake are our own origins. But until more fossils are dug up and genes begin to weigh more heavily in the diagnosis, multiple viewpoints will continue to flourish. All of which is fine and even useful, since each viewpoint is a slightly different idea to be tested experimentally with fossils or genes. This is the way the scientific method really works—warts and all—indeed the way science progresses, somewhat subjectively in the short term and often more objectively in the long term.

Ample evidence does exist that environmental change has made its mark—right up to the present day and continuing. As with all changes, whether physical, biological or cultural, evolution among our ancestral species is an accumulation of adaptive responses to changing environments: Nearby supernovae can trigger the infall of galactic clouds, causing some of their contents to form stars while hopelessly scattering others. Flash floods or geological faulting can speciate organisms caught in the midst of change so rapid they can no longer interbreed, allowing some to adapt and survive while others go extinct. Climate change can create long-term hardships for previously thriving hominids, forcing some of them to come to grips with new venues, whereas others aren't so lucky—or resourceful.

In Africa, the birthplace of humanity by virtually all accounts, the Rift Valley cuts across the eastern part of the continent from north to south, causing climate and vegetation to vary dramatically on either side. Today, we find wet western woods giving way to dry eastern grasslands, the whole valley displaying a rapidly changing ecology. As tectonic events and widespread droughts began affecting the valley landscape some eight million years ago, environmental change naturally separated our distant ancestors into two groups. Those in the western, tropically forested part of the valley became our closest cousins, the chimpanzee. Those in the eastern, drier, open savanna evolved differently, ultimately becoming human. To be sure, today's chimps live only in the wet and woody western part of the valley, whereas hominid fossils are found mostly to its east.

Closer in time, some three million years ago, changing environments once again speeded evolution. Climatologists know that global climate shifts increased in frequency and that the whole Earth was then cooler as ice sheets advanced over parts of North America and northern Europe. Eastern Africa, in particular, became yet drier, changing the vegetation from plants adapted to humidity to those capable of thriving in more arid lands. And it was in those open plains that the earliest hominids had to become more mobile, adept at long-distance hunting and skilled at opportunistic scavenging, for it was meat they were then after—and with it presumably came the rudiments of bipedalism. Our remote ancestors, also subject to those climatic fluctuations and more or less in turn, had evolved: *afarensis* → *africanus* → *habilis* → *erectus* → *sapiens.* And despite the diversity of fossil finds suggestive of an evolutionary tree or bush, some researchers contend that they probably came forth in a reasonably linear parade within a single lineage of hominids over the past several million years.

Naturally, not all workers in the field agree. Instead of a steady procession of hominids, others do prefer a bushy tree with different hominids hanging from different branches at the same time, making it difficult to draw a clear line of descent. More than most sciences, paleoanthropology has a greater share of controversy among its "splitters" and "lumpers." The splitters tend to interpret skeletons with pronounced shape differences as belonging to separate species; the lumpers often regard such disparities as anatomical variations within a single species. Usually, it's a matter of perspective, with subjectivity challenging objectivity in the absence of much hard data.

The prevailing view that the *A. africanus* and *afarensis* species were on the evolutionary path linking modern humans and whatever ancestry we share with the apes is certainly instructive but, even if valid in every respect, it simply pushes the basic question at hand still farther back in time: Who were the ancestors of the australopithecines? And here the answer becomes much more vague because the fossils are older, scarcer, and less well preserved.

Few discoveries have been made of hominid fossils predating the four-million-year-old specimens of the Afar lowlands. Just recently, however, an international team of paleoanthropologists working in Ethiopia has apparently pushed back our ancient kin yet again in time. A handful of bones, a partial jaw, and several teeth from at least a half-dozen individuals of the species *Ardipithecus* (a name meaning "root ape" in local dialect) have been dated to a little more than five million years old, and a hominid cranium (called *Sahelanthropus*, or Toumai, after a region in the West African Sahara) from Chad is perhaps as old as six million years. Though these bones are comparable in size to those of modern chimpanzees, their dental features resemble other hominids more than either fossilized or living apes. These fossil finds now overlap in time when DNA studies suggest that a common ancestor of humans and chimps lived in Africa from five to seven million years ago—a common ancestor who was chimplike, forest-dwelling, knuckle-walking, and mainly arboreal and fruit-eating.

Creatures having some human qualities, then, possibly resided on Earth even more than five million years back, but it's frustrating that so few prehuman fossils have yet been found at the base of the hominid tree for the period between five and ten million years ago. Plenty of fossils from that time span—known to some researchers as "the gap," to others as "anthropology's black hole"—lie scattered about in Earth's soil, but these are usually the remains of animals unrelated to humans. It was during this period that East African habitats changed dramatically, causing both widespread extinctions and the rapid rise of ancestors of animals such as giraffes, rhinoceroses, and antelopes—but not leaving much by way of early hominid remains. Some additional exceptions are fossils of an arm bone and some jaw fragments found near Lake Turkana and embedded in rock dating back about five million years. Most workers concur that these remains belong to an australopithecine, or whatever preceded that species, though no one can be sure on the basis of one arm bone and a partially crushed jaw. And a six-million-year-old, thickly enameled molar tooth has been uncovered at a nearby location. While a single tooth can hardly be used to trace hominid ancestry with any degree of assurance, it's enough to know that hominids were there then.

Creatures having human qualities possibly resided on Earth more than five million years back . . .

Tooth and jaw fragments of the oldest known creatures having any resemblance to humans or prehumans were discovered in India and later at several other places in Africa, Asia, and Europe. Radioactive dating of the rocky dirt in which they were buried implies that these fossils are eight to twelve million years old. Despite this old age, the jaws in particular still seem to have a mix of apelike and humanlike qualities. The creature's brain capacity and anatomical posture are unknown, however, since a complete skull has never been found. Only a few good fossils exist, none of them well preserved. Some anthropologists, examining mainly bone shapes, contend that this fox-sized fossil ape, originally called *Ramapithecus* in honor of the Indian god Ram, is possibly the ancestor of the australopithecines—a sort of protoaustralopithecine. Again, this is only conjecture based on the meager data currently available, but if correct this creature must have ventured in and out of jungles of the time, living partly in the forests and partly on the plains. Other researchers—mostly biologists examining molecules trapped in the bones—insist that *Ramapithecus* (now called *Sivapithecus*) is more likely the direct ancestor of our great-ape cousin, the orangutan, and not part of any direct lineage toward humankind. Instead, they nominate yet another fossil primate, *Dryopithecus*, found mostly in Europe and having powerful grasping capabilities for hanging and swinging below branches, as the more likely forebear of humans. Much more fieldwork and critical analysis are needed to sketch a reliable portrait of our distant relatives who roamed Earth some ten or so million years ago.

Earlier than that, and to connect with the prosimians of the Biological Epoch who came down from the trees some thirty million years ago, anthropologists have found a key transition species between the primate apes, which at the time resembled monkeys, and the living great apes of the present—the gibbons and orangutans of Asia and the chimpanzees and gorillas of Africa. *Kenyapithecus*, now extinct and dating back some fourteen million years, is taken by many (though not all) workers to be the ancestor of today's apes. At the least, all agree that this species was part of a wholesale migration of apes out of Africa and toward Europe and Asia. However, this interpretation hangs on hardly more than a handful of teeth found in western Kenya several decades ago. Other extinct species, such as the prehistoric ape *Proconsul* found in Africa and dated to eighteen million years old or a partially complete thirteen-million-year-old fossil skeleton named *Pierolapithecus* discovered recently in Spain, have been proposed as more reasonable candidates for the last common ancestor of both modern men and great apes. The fossil record is sparse, the evolutionary picture murky. Some doubt whether the fossilized bones of extinct apes will ever offer enough clues to fill in the branches of the ape family tree reliably.

Precisely when and where one species changed into another cannot be pinned down much better than described here, for one life-form slowly transforms into another over the course of history. The fossil record will never document an apelike mother giving birth to a distinctly human infant, or *A. africanus* parents raising an *H. erectus* baby. Evolution just doesn't operate that way. Changes of this sort are usually so gradual as to be imperceptible, occurring over long, long periods of time.

Paleoanthropologists aren't the only scientists tracing the evolution of hominids. As noted in the Biological Epoch, geneticists can scan the DNA of *living* primates and tally the number of mutations that have occurred over comparable durations of time. Most studies to date contend that humans and our closest relatives, the chimps, last shared a common ancestor some five to seven million years ago—in quite close agreement with recent studies of fossilized bones. However, such a molecular clock tells us little about the specific lines of descent among the more recent hominids and their ancestors during the past few million years—namely, what separates modern humans from our extinct forebears. And it doesn't resolve the heated controversy regarding the number of coexisting hominids in Africa and elsewhere, or the times of hominid migration and species divergence in Europe and Asia, or the likely reasons for the extinction of all but one of them—us. Even so, the history of the human species must be faithfully recorded in the genes of people still alive today—and therein lies the origin of humankind, one of the great prizes of the Cultural Epoch.

Part of the problem in untangling the genetic routes that led to humanity is that not all molecular clocks are calibrated in the same way. And some of them may tick at different rates, if, owing to variable lifespans and generational turnover, mutations occur less frequently in humans than in other primates and mammals. Discord regarding timescales often arises between bone experts and molecular biologists, thereby prompting us to ask: Do genes or fossils give the best results when mapping evolutionary lineages? One might think that, in principle, careful laboratory counting and sequencing of the nucleotide bases within genes ought to deliver accurate, objective answers, especially when the unearthed bones are often in such terrible shape, their interpretations laden with subjective opinion. But geneticists must still rely on the fossil record to calibrate their molecular clocks, that is, to calculate the number of nucleotide changes that have likely occurred per million years of evolutionary time. The calibration point currently used by most researchers for primate studies is pegged to the twenty-million-year-old split between apes and monkeys. But if this pivotal date is wrong—and it's uncertain by at least twenty percent—then the clock's calibration is faulty and so are the results derived from it.

Of course, if biologists could extract DNA samples from the fossil bones themselves, then most of the uncertainty would cease, thereby allowing a precise tracking of the evolutionary paths among the hominids. For example, how many genetic changes were needed for the onset of *Homo sapiens sapiens*, say, compared to the Neanderthals? Were they our immediate ancestors or an evolutionary dead end? Usable DNA is easily extracted even from single hairs of people deceased for a few hundred years, yet this formidable task has been successfully performed on only a few Neanderthal skeletons dating back several tens of thousands of years. (DNA begins to degrade from the moment of death as water, oxygen, and microbes attack it; all claims of DNA extracted from dinosaur bones or insects stuck in amber have been discredited.) The tentative conclusion is that Neanderthal DNA was sufficiently different from ours and thus their species was a side branch of the human family tree that diverged from ours about a half-million years ago, ending in extinction about thirty thousand years ago. However, the genetic trail quickly fades away farther back in time. Fossils older than a hundred thousand years do not yield measurable DNA samples and thus do not currently bring order to the confusion of bones and stones among the hominids dating back millions of years. Though genes and molecular clocks hold great promise to grant coherence to a very contentious subject—ultimately identifying the specific changes that made us human—they have thus far contributed marginally to this great origins enterprise.

∞

Some day, scientists might well have enough fossil and genetic data to prove the exact paths evolution took. Bones and molecules will ultimately show *what* evolved. But they're not as useful regarding *how* evolution occurred. To understand the reasons behind evolution in recent times, anthropologists are now studying the behavioral patterns of our closest *living* relative—the chimpanzee.

How do we know that chimps are so closely related to humans? Because the fossil record and the genetic markers say so—pretty clearly.

Laboratory studies of protein molecules are now routinely used to measure differences in the kind and sequence of amino acids among many animals. Comparisons of human proteins with those of, for example, a horse, a rat, or a frog show large differences in the number and order of their amino acids. But comparisons of human and chimp proteins show very few differences: The average human protein is more than ninety-eight percent identical to that of chimpanzees, making humans as close to chimps as, say, a fox is to a dog. Some proteins, such as hemoglobin (blood), have exactly the same numbering and ordering of amino acids in humans and chimps. Thus of all the members of the ape family, chimpanzees have a genetic makeup closest to that of humans. Bonobos (sometimes called pygmy chimps), relatively unknown members of the ape family and one of the last large mammals to be found on Earth less than a century ago, may have a genetic profile even closer to ours—and a fraternal behavior that rivals ours as well.

Chimpanzees also have the lifestyle closest to our own. (Gorillas outwardly resemble humans more than do chimps, but their genetic structure is a little different and their daily habits much different.) More than any other animal, chimps apparently resemble the ancestor from which other apes as well as humans descended. By studying modern chimps, then, behaviorists discern a little bit of what life was like for our ancestors several million years ago. Present attributes, adopted environment, and the social conduct of chimps might all tell us something about the evolutionary events that led to the divergence of humans from our common ancestor, but of course chimps too have since evolved.

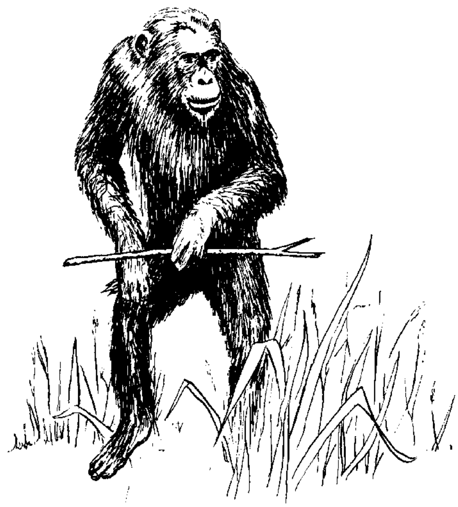
Many attempts have been made to study the lifestyles of caged chimps in zoos. But each time it soon became clear that the intricacies of ape society can likely be unraveled only in the wild. And "the wild" means just that; chimps and their ape relatives live in remote places, for their niche must differ from that of humans, lest we be unable to coexist. Most chimps are shy and unaccustomed to being watched by intruding humans. Many inhabit nearly inaccessible mountain retreats, while others stay up in the treetops of thick jungles. Reaching the appropriate places often proves tricky for anthropologists, as do the problems of which chimp properties to study and how to interpret the data once collected.

Organized fieldwork has now shown that chimps and other semi-bipedal (two-legged) apes are clearly more intelligent than monkeys and other quadrupedal animals. Bipedalism permits erect posture, thereby freeing the hands, and the resulting manual dexterity in turn provides a wealth of new opportunities for living. To give but one example, the uncanny handiwork of modern chimps is evident when they routinely fashion an implement by stripping leaves from a tree branch; they then insert it into a hole in a termite mound, remove it carefully, and systematically lick off the termites clinging to the branch.

Which was the cause and which the effect—the ability to stand on hind legs or the capacity to work with the hands—is currently an issue of dispute. Instead of the explanation just given, namely, that erect posture promoted toolmaking, the cause and effect may have been reversed. The need to manipulate food may have helped human ancestors to become permanently upright over the course of millions of years. Perhaps these two critically important evolutionary traits are so hopelessly intertwined as to preclude knowing which was the original instigator. The truth may be even more complicated, as each trait might have contributed to the other in an intricate way: Primitive bipedalism may have led to a small amount of manual dexterity, after which increased hand use accelerated the change toward more erect posture, which in turn fostered the development of even more complex tools, and so on. This is an example of positive feedback, whereby the advancement of one attribute stimulates another, after which the second reflects back on the first, and so on back and forth, thereby causing further and faster mutual development of both. Such feedback reinforcement was probably a key mechanism in the optimization of many traits throughout prehuman evolution.

The physiology and habits of modern chimps imply many things about our australopithecine forebears. Chimps are small enough to get around in the trees, yet large enough to ward off most predators while on the ground. They can be especially formidable when traveling as part of a large group, as they often do. Chimps' favorite food is fruit, especially ripe figs, though they also eat meat and birds' eggs, as well as small snakes, lizards, and insects. They are known to experiment occasionally with different foods, revealing some innate curiosity. To be sure, basic chimp intelligence is evident in many ways, such as when they use twigs as tools, bang rocks to smash objects, wave branches overhead to intimidate enemies, and employ grass as a sponge to hold water. Chimps have an open, free society, enabling them to try many new things.

Perhaps even more interesting than their expressions of curiosity are reports that chimps may show some degree of self-awareness. For ex-



Bipedalism permits erect posture, thereby freeing the hands . . .

ample, when exposed to mirrors, most chimps rapidly progress from treating the image as if it were another chimp to recognizing it as themselves. They seem to have a sense of "self," perhaps even a primitive cognition sufficient to know who they are. Once thought to be inherent only to humans, self-recognition might be an integral part of the intellectual repertoire of these remarkable apes.

Chimps are also observant copycats. The young learn readily from their elders, as well as from their human trainers. Though their lack of vocal equipment prohibits them from speaking, a few chimps can now communicate symbolically with humans by means of the sign language routinely used by deaf people. Some chimps have even displayed symbolic gestures to communicate with other chimps. Such chimp-to-chimp conversation implies that their intelligence incorporates much teaching and learning ability, meaning that chimpanzees are probably a lot smarter than anyone thought a generation ago. What's more, it may be unfair to label chimps mere copycats. Parrots and seals can also imitate, but there's a difference here: Other animals can be trained to imitate, whereas chimps seem to have a childlike ability to *learn* by imitation.

Chimps are furthermore sociable, though in a highly stratified way. All groups of chimps show a clear social hierarchy, comparable in many ways to today's human groups, whether in the military, business, academic, industrial, or political sector. One or a few males usually dominate a host of subservient chimps, thus ensuring some stability rather than the constant infighting that might otherwise engage a completely normless society. This social structure does not, however, seem to stifle their curiosity or amicability. Some chimps are forthrightly altruistic, sharing food with other members of their group, though most are not. Chimps are thus not wholly self-centered, as implied by earlier studies; some occasionally show affection for others in their group, though seldom for chimps who are not their immediate offspring.

Chimpanzee society is so complex that about fifteen years are needed for a newborn chimp to reach maturity. Like human adolescents, young chimps take many years to learn everything required to become a full member of their social organization. In a certain sense, young chimps are schooled by their parents.

Does that mean that today's chimps display the rudiments of culture? Watching them in the wild hammering a shelled nut with a rock, are we seeing the roots of human culture millions of years ago? The answers involve semantics again: some biologists regard culture as including uniquely human skills such as language, music, or art, but others regard it more simply as any behavior learned from fellow group members of a population rather than inherited through genes. This latter, more general interpretation might then include songs of birds and calls of whales. Although culture was for many years considered unique to the human species, evidence is growing for socially learned traditions elsewhere in the mammalian kingdom, and not just among chimps and bonobos but also including birds, monkeys, rats, and maybe even fish. But do individuals of all these species really learn from one another, with behaviors passed from generation to generation rather than discovering them on their own? True culture, like much else in this Cultural Epoch, is often more tricky—or at least more emotional—to address than some of the earlier, more remote issues of the Galactic and Stellar Epochs.

That chimps really do learn in their formative years demonstrates that environment plays a large role—at least among modern chimps. Other completely unrelated insect species, such as bees and ants, also have organized societies, but they don't really learn. Environment seems to have little bearing on insect knowledge. Laboratory studies show them lacking freedom of individual expression while going about their daily business; insects display little, if any, of the curiosity needed to try new things. Consequently, while insect society is most definitely (and impressively) organized, insect behavior is not overly complex. The social organization of insects is more rigidly controlled, being almost entirely programmed by genes.

Because chimps learn so well, behaviorists cannot easily tell how much innate smarts they really have. What part of their intelligence derives from their biological genes and what part from their cultural environment is simply unknown. We are now in the midst of an ongoing debate concerning the relative importance of the gene and the environment—a debate regarding not only the development of intelligence in chimps. The gene-environment controversy affects all aspects of living beings, especially the cultural evolution of humankind.

The extent to which intelligence is genetically preprogrammed, rather than environmentally endowed, remains controversial. The issue of gene (nature) versus environment (nurture) has triggered an emotional debate for the past quarter-century, indeed has forged a whole new interdisciplinary field of research. Sociobiology—the study of the biological basis of social behavior—aims to know the social instincts within any community of life-forms by appealing to the basic principles of psychology, genetics, ecology, and several other seemingly diverse disciplines. A principal goal of this research seeks to identify the inheritable traits that mold societies and secondarily to unravel the degree of importance between competition and cooperation.

Sociobiology expands the study of biological evolution to include society. Also known as social evolution or evolutionary psychology, whatever it's called it's bound to be a key feature of the Cultural Epoch. In cultural evolutionary terms, the fitness of an individual is measured not just by her own success and survival but also by the contributions made to the success of her relatives, namely, those who share some of her genes. These contributions are often self-sacrificing ones and can be classified under the general heading of altruism—unselfish devotion to the welfare of others—a fancy word for love. Whereas the catchphrase for classic biological evolution is the oft-stated "survival of the fittest *individual*," that for sociobiology and cultural evolution would be something like "preservation of an entire *society*."

Competition is not the sole driving force in evolution; cooperation is also a factor, at least for biological and cultural evolution and at least to some degree. That much was clear regarding the emergence of multicellular creatures from single-celled organisms a billion years ago, as well as during symbiosis that gave rise to eukaryotes billions of years before that, both noted previously in the Biological Epoch. Even earlier, mutual aid might have been essential for chemical evolution if replication of prebiotic molecules was helped along by catalytic surfaces at the time of life's origin. Advanced life, especially insect societies, such as ants and bees and several other animals, also seems to be partly founded on altruistic behavior, meaning that for them cooperation plays a significant role. For instance, wild dogs regularly regurgitate meals in order to feed their young; some species of birds postpone mating to help rear their siblings; "soldier" termites explode themselves, spraying poison over armies of ants, whenever a termite colony is attacked. This behavior is always the same, regardless of where and when the dogs, birds, or termites happen to act. They perform like programmed machines.

Behavior so rigid and uniform has prompted many biologists to argue that it might be exclusively determined genetically, at least among the "lower" forms of life. If so, then each trait, act, or duty likely has its own gene or genes, which are inherited in much the same way as body size, shape, and structure. The principal role of these behavioral genes is to preserve the species. Even while imprisoned within the bodies of life-forms, the genes control all. Extremely interpreted, life-forms exist for the sole purpose of perpetuating the genes—the selfish and unaltruistic genes.

Sociobiology remains controversial mostly because its proponents argue that its central tenets can be extended from insect societies to the societies of "higher" life-forms, including humans. Problems always arise when scientists—or anyone really—make grandiose pronouncements about our own species. Trouble starts because human nature isn't always what we think proper. Bias and value judgments sneak into science, perhaps an inevitable consequence of getting closer to studying ourselves along the arrow of time. The main issue is this: To what extent does human behavior depend on the underlying genes? Which has the dominant influence over the actions of humans, nature or nurture? This is the root of the controversy—human understanding of human affairs.

Researchers generally fall into two groups, both conceding that environmental factors play the greater role in human behavior. One faction maintains that environment is virtually the only important influence: behavioral differences among humans are strongly governed by social, cultural, and political factors—meaning that human control of human behavior is possible. The other group contends that genes are of considerable import: genes may contribute only one-tenth in their contest with the environment, but this is enough for many traits (aggression, envy, sympathy, love, fear, intelligence, among others) to be partly predestined in humans. If so, then societal changes in human behavior are limited because much behavior is biologically dictated by the genes. This second school of thought presumes that, for example, the behavior of humans who go hungry to feed their children or the behavior of people who risk their lives to save a drowning swimmer is not the result of free will. Instead, paralleling an insect's desire to preserve its own species at all cost, such behavior is an unconscious reaction built into and dictated by our genes to ensure survival of our own kind—a nepotistic process known as "kin selection" among interacting individuals who are genetically related. Cooperation then comes to be seen as costly to individuals but potentially beneficial to groups.

Whichever ideas of sociobiology prove valid, it will be important for psychologists and psychiatrists to pay heed. The way people act may be, to some degree, biologically predetermined. Sociologists should also take note, for sociobiology may someday give them quantitative methods by which to test their frequently unsupported assertions. Indeed, economics, law, and politics might eventually become part of the interdisciplinary subject of sociobiology and in turn part of the more inclusive, transdisciplinary worldview of cosmic evolution.

Love, altruism, kinship, and curiosity are attributes associated not only with humans but also with chimps and probably other animals as well. Many contend that these notable qualities of goodwill are less apparent in chimps than in humans, but a glance at the daily newspapers can create doubts.

Not to imply that chimps are nice and gentle all the time; vegetarian pacifists they're not. Chimps do resemble humans in yet another way, namely their occasional desire to exert unnecessary aggression. Some conflict within and among species is a normal, perhaps even essential ingredient of biological and cultural evolution. "Nature red in tooth and claw" may sound politically incorrect among idealists in today's revisionist society, but competition and exploitation are part and parcel of any evolutionary setting. Cooperation and mutualism work more subtly and only to limited extent alongside natural selection, especially as regards individual reproductive success. Even when life-forms do clearly cooperate, it's often done only when in their own best interest. Without some aggression in the guise of competition, few if any species could adapt to changing environments. Reactive aggression probably has deep biological roots, yet unprovoked aggression would seem to be another thing entirely.

Field studies in Tanzania illustrate how some chimps occasionally murder other chimps for no apparent survival-related reason. Premeditated, gangland-style attacks were directed by a large group of male chimps on a smaller group of males and females that had previously broken away from the larger group. Over the course of five years, each member of the splinter group was systematically and brutally beaten. All died. Only young males initiated the attacks, which occurred only when the victims were isolated from the others. Hands, feet, and teeth were often used by the attackers, though sometimes fieldworkers noticed stones being deliberately thrown. The hope, of course, is that comparative studies like these will uncover the reasons behind not only chimp misdemeanors but human belligerence as well, perhaps helping to guide the future survival of the human species, which, it would seem, can no longer tolerate intraspecies aggression.

Despite lingering controversy over details, behavioral studies of modern chimps have aided greatly our understanding of the ascent of humans. As apelike animals resembling chimps nearly ten million years ago began leaving the forests for the savanna (or perhaps it was more that the forests left them as an oncoming ice age made parts of Africa cool and dry), they were probably forced by environmental circumstances to become more sociable in order to survive. The origins of our social organization may well have been shaped by the new, harsher conditions in the open plains, where there would have been less food, reduced protection, and thus greater need for group cooperation—or was it everyone for himself, hence unfettered competition?

These hardships nonetheless gave our ancestors a chance to experiment and to learn as their sights and experiences grew over the course of millions of years. The parochial mentality of forest-living animals was replaced by the wider perspective of our plains-dwelling ancestors. Of much consequence, this suddenly larger world created pressures to evolve bigger brains capable of storing a dramatic increase of raw information.

Change from life in the trees to that in the plains was a renaissance of sorts that likely took a million years or more. Yet once it commenced, the race was on—a race to inhabit entirely new niches, to develop whole new ways of life, and eventually to become technologically intelligent.

∞

The Biological Epoch treated the evolution of the central nervous system from the onset of multicellularity about a billion years ago to the increasingly talented primates living among the trees some tens of millions of years ago. We now resume that history of more recent times, tracing the dramatic rise in cranial capacity among our immediate ancestors of the past few million years. Many of the rapid, sophisticated advances of the Cultural Epoch focus on the brain—either having caused the brain to increase, or conversely being caused by its increase.

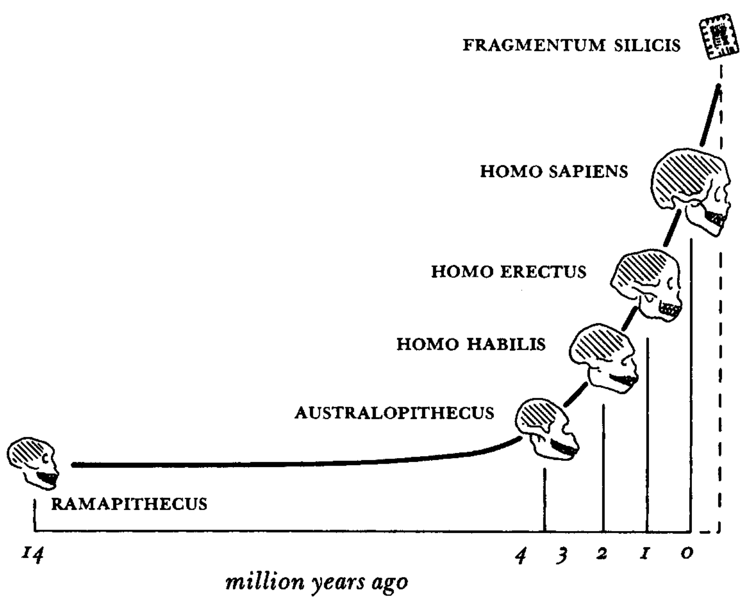
Human beings now have brain volumes of about fourteen hundred cubic centimeters, about the size of a large grapefruit. In mass, that's a little more than a kilogram, or a weight of about three pounds. Sizes do vary from person to person, though no clear behavioral differences are known between people with brains as small as a thousand or as large as two thousand cubic centimeters.

On the other hand, most mental patients having reduced cognitive abilities do have distinctly smaller brains. Often measuring five hundred cubic centimeters, the brain size of these mentally retarded adults approximates that of a normal one-year-old child. Apparently, the brain can be so tiny that its function is much impaired, meaning that a minimum brain volume is likely needed for "adequate" human intelligence as we know it. Once this threshold—probably around a thousand cubic centimeters—is surpassed, normal human behavior is possible.

What about our immediate ancestors? Does the fossil record allow estimates of the brain size of some of the prehumans that paved the way for our existence? The answer is yes, for anthropologists have been able to sketch a rough outline of the recent evolution of the brain. They do so by measuring cranial capacity of the hollow, fossilized skulls of our immediate ancestors, assuming that, as is now true for humans, apes, monkeys, and other modern mammals, the brain matter nearly fills the skull.

The partly bipedal and prehuman australopithecines of three million years ago had brain volumes averaging not quite five hundred cubic centimeters. This is just a bit larger than the brain of a modern chimpanzee and about a third the size of today's average human brain. Thus, the fossil evidence supports the idea that our ancestors could walk on two feet before they evolved large brains.

The first true humans, perhaps *Homo habilis* of two million years ago, had definitely larger brain volumes. Fossils studies show that this ancestor was fully bipedal and had an average cranial capacity of nearly seven hundred cubic centimeters. Not only that, but their fossilized skulls have a distinctly different shape from that of their forebears. Developed substantially were the frontal lobe behind the forehead and the temporal lobe above each ear, those brain regions regarded as sites of speech, foresight, and curiosity, among other useful behavioral traits. Coupled with these ancestors' bipedal posture, the possibility that they may have also made primitive tools implies that at least two significant changes in behavior—toolmaking and bipedalism—were accompanied by significant changes in brain volume. Whether tool use led to bipedalism or the converse remains another of those chicken-or-the-egg conundrums, probably reinforced by positive feedback as noted earlier. At any rate, the fact that bipedalism freed the hands for tasks other than walking connotes a causal link among upright posture, toolmaking, and ultimately brain size.



. . . in the last few million years . . . our ancestors' brains no less than tripled in volume.

Fossils also show that our closest relative, *Homo erectus*, had an average brain volume somewhat less (about a thousand cubic centimeters) than those of our friends and neighbors today. What therefore amounts to a dramatic rate of cranial growth—a fifty percent increase in roughly a million years—coincides with the expansion of early humans to colder climates during the last ice age, an environmental challenge that might have enhanced selection for larger brains to plan the use of seasonal resources, indeed to think more about sheer survival. Large and small circular arrangements of stones found alongside the fossilized remains of this species furthermore imply that our ancestors of a half-million years ago had domesticated fires and constructed homes outside of caves. And cut marks on animal bones suggest that they had shifted from eating fruits and nuts to meat, giving them more energy per unit mouthful. Our recent human ancestors were beginning to *challenge* the environment—to change *it* for a change.

Comparisons of various cranial capacities, then, clearly imply that hominid advances—both biological and cultural—made in the last few million years are at least partly related to enlarged total brain size. During that time, our ancestors' brains no less than tripled in volume. New behavioral functions, increased neural specialization, varied dietary preference, and improved cultural adaptations surely accompanied the steady evolution from *Ardipithecus* through *Australopithecus*, onward to *Homo habilis* and *Homo erectus*, currently culminating in *Homo sapiens.* It was not necessarily the fittest, nor even the strongest, who survived, but those able to best adapt to change.

Absolute brain size is important, but it cannot be the sole measure of intelligence. Small-bodied creatures such as birds have minute brains, especially compared to the much bigger ones of large-bodied creatures such as elephants. Yet in many respects birds act "smarter" than elephants, probably because the former have a lot less body to monitor and control. In fact, much of an elephant's large brain consists of motor cortex—enormous numbers of dedicated neurons enabling those huge hulks to put one leg in front of the other without tripping. Hence the reason why most neurobiologists take as a better measure of intelligence a *comparison* of brain and body sizes.

Ratios of brain to body mass for many animals with similar overall stature show a clear separation of reptiles from mammals. For any given body mass, mammals consistently have higher brain mass, usually ten to a hundred times larger than those of modern reptiles of comparable size. Likewise, the brain masses of our prehuman ancestors (the early primates) also were greater, relative to body mass, than those of all other mammals.

The creature having the largest brain mass for its body mass is *Homo sapiens.* (Our brain-to-body mass ratio equals 0.022.) Dolphins come next (0.016, which is also the value for *H. habilis*), followed by the apes, especially the chimpanzees (0.006). The human brain is about as big as the genes can currently make it and still be safely delivered during childbirth—three or four times bigger, relative to body weight, than the brains of our closest relatives, the great apes. These are data, not sociological sentiments.

Brain-to-body-mass ratios, then, provide a useful index of the intellectual capacities among a range of animals. In this way, the fossil record virtually proves that the evolution of mammals from reptiles about two hundred million years ago was accompanied by a major increase in relative brain size and intelligence. These ratios furthermore show that additional neural evolution paralleled the later emergence of humanlike creatures from the rest of the mammals a few million years ago.

More than any property, the brain most clearly distinguishes humans from other life on Earth. The development of speech, the invention of technology, and the rise of civilization are all products of the human brain's rapid advancement. But what about other forms of life? Are there creatures on our planet today with comparable intelligence—animals having neural capacities enabling them to communicate, act socially, or make tools?

Brain-to-body-mass ratios imply that apart from humans dolphins are the smartest animals now on Earth. As a numerical measure of intelligence, their just-noted brain-body ratio matches that of archaic humans of a couple million years ago and exceeds that of the australopithecines of several million years ago. Laboratory tests do imply that dolphin intelligence, to the extent that it can be realistically gauged, does lie somewhere between that of humans and chimpanzees. Biologically, dolphin evolution seems not too different from ours, yet culturally they are far behind us, perhaps because they live in the water.

Dolphins were not always aquatic creatures. Along with whales and porpoises, dolphins are members of a family of mammals whose ancestors were once land dwelling. Owing to keen competition among many four-legged amphibians some fifty million years ago, the dolphins' ancestors returned to the sea, possibly either in search of food or because land niches were becoming too crowded. Some disadvantages would have undoubtedly accompanied such a seemingly backward move, but that ancestral decision—really an adaptation to change—probably saved them from extinction.

Dolphins, as we know them today, are well adapted to the sea. Their exceptionally strong bodies are streamlined for deep diving and speedy locomotion. They have extraordinary hearing beyond the range of humans, as well as an uncanny sonar system resembling a kind of underwater vision. This advanced system of echo location, now being studied by human naval officials for military purposes, may employ a kind of acoustical radar to map the position and movement of objects in their watery environment.

Interestingly enough, almost every year hundreds of dolphins (and whales too) beach themselves, especially along the outward-jutting Cape Cod off the New England seacoast. Most likely, their navigational beacons go awry, causing them to temporarily lose their way. Or, just perhaps, these dolphins are trying to make their way back onto the land. Are we sure ours is a humanitarian gesture when we so quickly "rescue" them and dump them back into the sea, or are we unwittingly keeping them out of our land-based niche?

Dolphins also have a well-organized social structure. They travel in schools or families and assist each other when in trouble; females often act as midwives for another dolphin. They are not at all hostile, being extremely friendly to other dolphins as well as to humans. Dolphins seem to be the exception to the unwritten rule that all friendly species are inherently aggressive as well—though they certainly are known to ram sharks in a coordinated way if threatened, ganging up on the predator to protect their own.

In addition to their unparalleled ability to navigate underwater, dolphins communicate with one another by means of a series of whistles, quacks, squeaks, clicks, and other noises often resembling Bronx cheers. Although we can hope to communicate with them someday, the human range of generating and hearing noise is relatively limited when compared to the dolphins' much wider auditory range. They are known to be able to produce and hear sounds within our audible range, but to do so requires them to grunt and groan at frequencies lower (bass) than normal. Most of the sounds normally made by dolphins are inaudible to humans, making it improbable that their way of expressing meaning overlaps ours at all. Not inconceivably, dolphins in captivity may have been trying to communicate with us for years. If so, they must be quite discouraged by our lack of response.

Interspecies communication will not be easy, whether among humans, dolphins, or chimps. Empirical findings to date nonetheless suggest that some common ground exists for future cultivation of, especially, dolphin-human links. At the least, it seems that both parties are interested in such a collaboration.

∞

Changes that affected humanity were evolutionary, not revolutionary. Although Darwin saw selection operating on individuals as the primary mode of evolution, biologists today have widened our view to include changes among whole populations on regional, and sometimes global, scales. That evolution still occurs competitively by means of the usual adaptations to changing environments, but broadened, cultural opportunities for living—and thinking—within the past million years quickened the pace of evolution. And it hasn't slowed since.

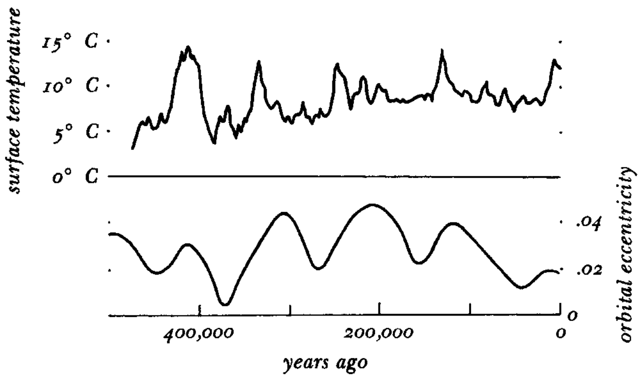
Environmental changes act as the motor of evolution, allowing some life-forms to adapt successfully while forcing others to extinction. There were winners and losers, and no amount of political correctness can alter that. As noted earlier in this Cultural Epoch, some of the most dramatic environmental change on Earth, excepting asteroid impacts, is caused by the climate. Not least, glacial cycling drove biological evolution on our planet for eons and has driven cultural evolution more recently. Harsh climate some thirty thousand years ago, in fact, may have been the reason the Cro-Magnons replaced the Neanderthals; the former had mastered the technical skills needed to survive at the height of the most recent ice age.

Apart from seasonal effects occurring over monthly durations and continental drifts spanning millions of years, planet Earth experiences intermediate-scale changes in global climate over the course of thousands of years. Surprisingly detailed climatic records dating back nearly a million years have been derived by a variety of methods, including analysis of trapped gas and dust in core samples taken from the Greenland and Antarctic icecaps, of pollen in sandy sediments extracted from the seafloor of the North Atlantic, and of land-based geological data exposing freeze-thaw cycles. To give but one example, some snails coil to the left in cold water and to the right in warm water; the proportion of each type in their fossil remains yields a profile of ocean temperature over hundreds of millennia. Since plants and animals are acutely sensitive to changes in climate, their fossils extend our knowledge of climate, albeit less reliably, hundreds of millions of years into the past.

Geochemists can virtually prove that during the most recent million years our planet has cycled through about ten (including four major) episodes of cool, dry climate—intermittent periods commonly known as ice ages. Hence, there was no single Ice Age per se but several of them in recent years, geologically speaking. Though some of the data are incomplete, each cold ice age, as well as its opposite warm interglacial period, apparently lasted several tens of thousands of years. We now reside in an interglacial period—a temporary thaw of sorts before heading back into the deep freeze, though probably not for another twenty thousand years. In fact, all of human history—including the rise of agriculture, nation-states, and technology—has occurred within the current, ten-thousand-year-long interglacial warming trend.

What causes these cycles of heating and cooling on our planet? Some geologists contend that glaciation increases during periods of global volcanic activity when ejected dust reduces the amount of sunlight penetrating Earth's atmosphere. Others maintain that periodic reversals of Earth's magnetic field have caused the protective Van Allen belts to collapse, thereby sporadically allowing unusually high doses of solar radiation to heat the ground and thus decrease glaciation. Still other researchers note that ice ages could have been induced on our planet by variations in the output of the Sun itself, passage of Earth through an interstellar dust cloud, altered circulation of deep water in Earth's oceans, reduction of greenhouse gases in the atmosphere, or any one of a long list of other proposals.

Recently, oceanographers have found convincing evidence to support yet another theory, dubbed the Milankovitch effect after the Serbian mathematician who championed it in the mid-twentieth century. According to this idea, subtle though regular changes in Earth's attitude toward the Sun trigger the ice ages as variable amounts of sunlight hit our planet. These changes are the combined result of three astronomical effects, each in turn cyclical and caused by the normal gravitational torques (or twisting forces) exerted on Earth by the Sun, Moon, and other planets in the Solar System: First, change in the *shape* or "eccentricity" of Earth's orbit about the Sun. Second, *precession* or "obliquity" of Earth's spin axis. Third, change in the *tilt* or "wobble" of Earth's spin.



. . . subtle though regular changes in Earth's attitude toward the Sun trigger the ice ages.

To be a little more technical, Earth's elliptical orbit cyclically alters its shape about every hundred thousand years, becoming more circular, then more oval, and so on. When the orbit is most elliptical, the Earth receives about thirty percent more radiation when it's closest to the Sun than when it's most distant. Also, Earth slowly and smoothly precesses like a spinning top, returning to its starting point roughly every twenty-five thousand years, thereby changing the presentation of Earth's hemispheres toward the Sun. Finally, over a period of about forty thousand years, the tilt of Earth's axis (currently 23.5° relative to its orbital plane) wobbles by a few arc degrees, which is enough to change further the temperature contrast between winter and summer.

The first effect would produce warmer summers and colder winters during periods of high orbital eccentricity. The second also alters seasonal differences, driving extremes in climate when tilt combines with eccentricity unfavorably. And the third likely causes milder winters and cooler summers with reduced melting when Earth's axial tilt is low. The net result of these three effects—for all three operate simultaneously yet over different timescales—sometimes causes abnormal solar heating such as we are now experiencing; yet at other times, that heating is distinctly reduced, producing a decline in global temperature and widespread glaciation.

This theory of astronomically induced ice ages is currently favored among the majority of working scientists mostly because samples of seafloor sediments show that, during the past half-million years, tiny sea plankton have thrived at certain times, while barely surviving at others. Studies of the abundance of fossilized plankton known to prefer warm or cold water (much like the coiling of shellfish noted above) provide estimates of the prevailing water temperature during their lives. This inferred sea temperature correlates well with the expected heating and cooling of Earth by means of the three combined astronomical effects.

Apparently, then, slight changes in Earth's axial tilt and orbital geometry are mainly responsible for triggering the ice ages. Whether they are the only instigator remains to be proved by further research. At least to some extent, the Milankovitch model reinforces yet again a robust astrobiology connection at work in Nature, for the cosmic stirring of the ice ages must have had profound impacts on the evolution of life on Earth.

The most recent major glaciation began nearly a hundred thousand years ago, after which the climate pretty much returned to the way we now know it by ten thousand years ago. At the height of this ice age, some thirty thousand years ago, an ice sheet nearly two kilometers (about a mile) thick extended from the Arctic far enough south to cover much of North America as well as a good deal of northern and central Eurasia. Earth's overall surface temperature then averaged about five degrees Celsius (or nine degrees Fahrenheit) lower than it does today, while the sea, with much water locked up in ice bergs, was about a hundred meters (or three hundred feet) below current levels.

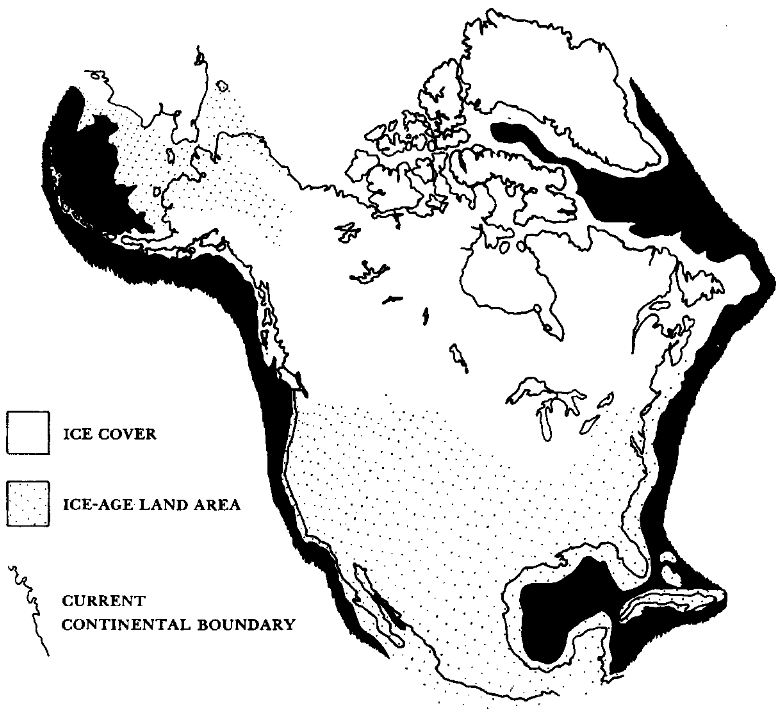
Even thicker, more extensive ice probably covered most, and perhaps even all, of Earth's surface long ago—suggesting that mass extinctions of life could have been caused by cold air and glacial ice, not merely by the fire and brimstone of asteroid hits or the rising tides of oceanic waters. Some geologists have recently argued on the basis of glacial debris and biological tracers in ancient rock that massive glaciers might have completely entombed the entire globe some six hundred million years ago, just prior to the Cambrian outburst of multicellular organisms. Reaching even into the tropics, kilometer-thick sea ice might have encapsulated all the oceans, sealing them off from the atmosphere and potentially cutting off life from its usual source of energy, the Sun. How this "snowball Earth" got itself into such a deep freeze is an unsolved puzzle, but how it got out is an even bigger conundrum. The only reasonable exit would seem to have been volcanoes that belched out enough carbon dioxide to create an enhanced greenhouse warming of the planet—which, in turn and for a while at least, probably caused a brutal episode of heating sufficient not only to melt the ice but also bake the planet. How life survived the rigors of such severe climate reversals is another problem, unless it did so exclusively on geothermal internal energy without recourse to any outside solar energy. An alternative, milder model (called "slushball Earth") contends that, in the tropics at least, the snow didn't freeze solid; if some open waters stayed unfrozen in equatorial refuges, the danger of life's extinction would have been lessened. Scientists are currently troubled about the catastrophic harm potentially done to the biosphere during this multi-million-year-long cold spell that once (and maybe often) wracked our planet's surface—but coming as it did just prior to the Cambrian, perhaps the wave of heat that followed actually stimulated the evolution of diverse animal life more than harming it.

The snowball-Earth scenario of several hundred million years ago brings to mind the so-called faint-Sun paradox of several billion years ago. The issue here, noted earlier midway through the Stellar Epoch, is the slow rise in the Sun's brightness over the course of time; all stars experience this as hydrogen in their cores increasingly converts to helium. Both the theory of stellar evolution and the measured amount of energy reaching Earth today imply that our Sun is currently brightening by about one percent every hundred million years. Extrapolate back some three to four billion years, and the faint young Sun was probably only a third as bright as today. Water on Earth should have been frozen solid for its first two billion years or so. Hence the paradox: How did primitive life survive, or even get started, if the solar energy reaching Earth billions of years ago was insufficient to melt ice? The possible answer, much as for the snowball-Earth scenario more recently, holds that Earth's early biosphere must have been warmed above freezing by greenhouse gases (not only carbon dioxide but also ammonia and methane) released from surface volcanoes and undersea vents. Otherwise, the geological record, whose ancient sedimentary rocks show clear evidence that Earth's climate has always kept oceans in the liquid state, cannot be reconciled. Such global challenges to life may well have been regular, ongoing episodes in the natural history of our planet—much as both natural and anthropogenic events, locally and globally, will likely continue to threaten future life on Earth.

The controversial concept of Gaia is also relevant here—the idea that living organisms can alter their environments and not just the other way around. Modern-day Gaians—disciples of the Gaia hypothesis and cultural descendants of worshippers of the terrestrial Greek goddess—go even further, claiming that planet Earth is a single, vast superorganism, indeed that Earth itself is alive ("strong Gaia"). Reality or metaphor, the notion of Gaia argues for living creatures that affect the composition of Earth's biosphere, where both environment and life act in a coupled way so as to regulate the former for the benefit of the latter ("weak Gaia"). Microorganisms surely excrete metabolic waste products that modify their environmental conditions, a property that enhances their own species' survival, which impressively extends over billions of years. Accordingly, living systems have seemingly prevented drastic climatic changes throughout much of Earth's history, as evolution has endowed organisms with improved ability to keep surface conditions favorable for themselves. For example, as our Sun ages and therefore sends more heat toward Earth, life responds by changing Earth's atmosphere and surface geology to keep the climate fairly constant—a kind of geophysical thermostat that basically adjusts carbon-dioxide levels down by preferentially trapping it in calcium carbonate (calcite) under high temperatures, thereby cooling the atmosphere. If those temperatures drop too low, the reaction to form calcite decreases and the carbon dioxide and the air temperature rise, the whole cycle acting again in feedback fashion to keep water liquid and Earth habitable.

This regulating effect of life on climate will not occur indefinitely; Earth's thermostat will eventually fail. In roughly a billion years—well before the Sun terminates in five billion years—rising solar energy will outpace the offsetting, cooling effects at Earth. Our planet will likely experience runaway, Venus-like greenhouse heating too great to sustain life.

At any rate, ecologists are certain that the last ten thousand years have seen the glaciers retreat, the coastal plains flood, the vegetation bloom, and the ocean and atmosphere warm. The change in climate from the peak of the most recent ice age to its present state occurred quickly, by geological standards. And although global weather has remained anomalously benign during these past ten millennia when civilization developed, ongoing changes in Earth's regional environments, such as those produced by widespread oceanic (especially North Atlantic) circulation and atmospheric (El Niño) oscillation or by local droughts and floods, have surely helped foster the many advances made by our human ancestors during this period of rich innovation. Some of those natural changes, no doubt aided and abetted by anthropogenic actions like internal strife and tribal warfare, may have led to the demise of whole peoples and civilizations, such as possibly the Mesopotamians and later the Mayans—just as innumerable climatic warmings and coolings pushed unfit mammals of old to extinction, all while spurring the evolution of new, better adapted species. Humans were forced to adapt—biologically, culturally, and rapidly—to changes throughout the air, land, and sea. The motor of evolution had indeed quickened.



. . . ice ages must have had profound impacts on the evolution of life on Earth.

In one notable, if provincial, example of climate change, the ice ages themselves may have accelerated the migration to and colonization of North and South America. Anthropologists know that humans didn't evolve in the Americas since fossil evidence has never been found there for apelike creatures from which they could have ascended. Current consensus has it that the so-called Clovis hunters arrived in the New World of the Americas no more than a few tens of thousands of years ago, and possibly as recently as twelve thousand years ago. Their artifacts (notably Clovis arrowheads, first found near the present-day New Mexico town of that name) are strewn across parts of North America, but human remains are few. Kennewick man, who lived in what is now the state of Washington about nine thousand years ago is the most celebrated case. Precisely how they transitioned from the Old World is unknown, but during one of the more recent glaciations, enough water would have been wrapped up in ice to have lowered the sea level by tens of meters, allowing humans to walk dry-shod across the Bering Strait between what is now Siberia and Alaska.

This, then, is the prevailing view: Native North, Central, and South Americans are descendants of Asians who chased and hunted mammoths into the Great Plains hardly more than a hundred centuries ago. Once settled in the Americas, these migrants further developed arts, languages, tools, and many other cultural amenities. Whether civilizations of the Americas experienced cultural evolution independent of those in Eurasia or whether they had some as-yet undiscovered contact with them is another of those contentious controversies in modern anthropology.

∞

Several factors helped make us civilized, sentient human beings. Suffice it to note but a half-dozen such factors in relatively recent human evolution and more or less chronologically: the dawning of rudimentary technology, the discovery of useful fire, the development of symbolic language, the emergence of observational science, the design of mythical stories, and the recognition of self-consciousness. Undoubtedly other factors also helped make us thoughtful, cultured humans, but practical advances and critical thinking such as these are representative of the more important ones.

Anthropologists concur that our ancestors of the past few million years must have survived mainly by hunting, gathering, and scavenging food. The acquired traits of pursuing and eating meat and other high-protein fare were likely exported from the forests to the savanna, whereupon they were enhanced owing to the relative lack of fruit in the open plains. Though most inhabitants of modern civilization no longer regard themselves as hunter-gatherers, this was indeed the job description of all our ancestors from several million years ago until the rise of agriculture some ten thousand years ago. (Then again, who is to say that we don't do the same today when shopping at the supermarket: hunting up and down the aisles, gathering food into our baskets, and scavenging for the best parts of animals that were killed by someone else.)

How do we know that early humans, even the advanced australopithecines, hunted? The evidence is twofold, both found in the fossil record. First, scattered bones of a variety of large animals are often found near those of our ancestors at many dwelling sites along the East African Rift Valley. The animal bones are not intact skeletons but are rather strewn debris suggestive more of a picnic than a natural death. Second and more convincing, tools made from stones are often found alongside the remains of two-million-year-old australopithecines, as well as those of all the more recent human species. Those stone artifacts have endured for millions of years and it seems safe to conclude that even earlier ancestors might have utilized wooden tools that didn't endure.

Judging by the shapes of the stone tools unearthed at Olduvai Gorge, many of these egg-sized implements were used to chop, cut, and prepare food for easy consumption. Many others, though, evoke weapons, especially the rounded stones probably used to maim or kill when thrown, much as modern chimps occasionally do now. Other stones resemble club heads, and they were probably used for exactly that—hunting by killing with a club of some sort. These were possibly the type of "tools" wielded by the advanced *A. africanus* (or *H. habilis*) to exterminate its relative, *A. boisei*, about a million years ago. Whatever their use, these primitive implements evince the beginnings of the technological society that we all now share.

Stones were used not only for tools and weapons. They also provided foundations for early homes. A two-million-year-old site in Olduvai Gorge, for example, contains a circular stone structure conjectured to have been the base of a hut of some sort. This kind of primal, Oldowan stonework predated the so-called Stone Age—a time when our ancestors not only rearranged rocks but also broke them to serve more usefully than mere rockpiles.

Depending upon the place excavated, the duration of the Stone Age spans a period from roughly a million years to hardly ten thousand years ago, after which the bronze and iron ages continued until Renaissance times only five hundred years ago. Distinguished by increasingly intricate stoneware, including various types of hand axes, cleavers, spatulas, and scrapers, the Stone Age displays a steady transition from rather crude tools to more finished ones alongside the advancing fossil record of biological species. Hence the early Stone Age is historically associated with the onset of *Homo*, and toolmaking itself must have accelerated the evolution of the first true human beings.

Much of the long-ago construction of stone implements preceded the enlargement of the brain. The earliest stone-wielding creatures had brains just a little larger than those of modern chimps, not much bigger than five hundred cubic centimeters. Undoubtedly, tool use and bipedal posture, the twin hallmarks of manual dexterity, were powerful evolutionary advances—fundamental changes that triggered whole new opportunities for living. Unbeknownst to our ancestors of the time, their tool-like chips of rock were the start of a manufacturing society, a technological culture. The difference between stony spoons and jumbo jets is only a matter of degree.

The threshold of technology, then, is hard to pinpoint exactly, but it probably occurred more than a million years ago. The beginnings of cultural, as opposed to utilitarian, activities may be nearly as old, for brightly colored mineral pigments have also been found alongside skeletons of the earliest of the true humans. Even the advanced australopithecines may have had some use for ritual, as geometrically arranged pebbles are often found near their remains.

Only toward the end of the Stone Age some dozens of thousands of years ago do more industrious and sophisticated activities become evident, when a genuine outburst of cultural evolution brought forth a wealth of arts and crafts—a wave of abrupt cultural change so dramatic as to be called by some "culture's big bang." Technological advances such as construction of the wheel in central Europe nearly fifty thousand years ago were matched by cultural advances such as the oldest deliberate burials in certain European and Asian locales nearly sixty thousand years ago, as well as the onset of stunning prehistoric art on the cave walls of western Europe and elaborate sculptures of figurines of personal adornment some thirty thousand years ago. These were uniquely human inventions, unarguable signs of behavioral modernity—cultural products of *Homo sapiens*, including some Neanderthal and especially Cro-Magnon peoples.

The discovery of fire's usefulness is perhaps foremost among those factors that accelerated culture, its management of crucial import in the toolkit of the hominids. Our ancestors have been using fire for light, heat, and probably protection from predators for nearly a million years. Archaeological expeditions of caves in France have revealed blackened hearths at least that old. The general benefit of fire, then, was welcomed a very long time ago, at least insofar as it provided warmth in colder climates. But it seems that its functional practicality went unrecognized until more recently. The cooking of food, for example, arose only a couple hundred thousand years ago. Techniques to fire-harden spears and anneal cutting stones are likely to have been still newer inventions. Significantly, the widespread use of fire, especially for utilitarian purposes, was one of the last great steps in the domestication of humans.

Beginning not quite ten thousand years ago, our ancestors learned to extract iron from ore and to convert silica into glass, as well as to cast copper and harden pottery. Many other industrial uses were realized for fire, including the fabrication of new tools and the construction of clay homes. However, archaeologists disagree about the temporal ordering of many of these basic advances. Just when and how one invention paved the way for another is so far unclear. Some researchers argue that, after heat and light, baking clay to make pots was the oldest organized use of fire, even predating the regular cooking of food. Others maintain that the need for pottery arose because cooking was already established since the earliest uses for pottery must have been for cooking and storing food.

When cultural advances are intimately linked in this way, which was the cause and which the effect is never quite clear. Both the motivation for and the exact time of an invention are hard to establish. Some of the pivotal steps may never be pinned down. But of one thing we can be sure: scores of new mechanical and chemical uses of fire were mastered during the past ten thousand years and a few may have been discovered well before that.

Ancient clay, metal, and glass products can still be found in the bazaars and workshops of Afghanistan, China, Iraq, Thailand, Turkey, and other generally Middle Eastern and Asian countries. More modern results of these early technologies are evident in the cities of steel, concrete, and plastic surrounding many of us in the twenty-first century. Cultural changes of this sort are not without problems, however. Increasing amounts of energy expenditure, in the form of fire and its many derivatives, have often been accompanied by grim aftereffects, not least environmental pollution and energy shortages only now becoming evident in our contemporary world.

Development of language—that marriage of speech, cognition, and perhaps a dose of emotion—was another central factor in making us cultured, indeed a unique feature in making us human. Some psychologists regard language as synonymous with intelligence, or at least the "jewel in the crown of cognition." Others, such as the nineteenth-century Linguistic Society of Paris, which banned discussion of the origin of language, regard it even today as too inconclusive for empirical study if only because linguistic behavior doesn't fossilize. At the least, human language would seem to be our main evolutionary legacy—perhaps the most dramatically new trait to emerge since the Cambrian explosion a half-billion years ago. The flowering of grammar, syntax, and higher intellectual function coincides with the change, only about fifty thousand years ago, of anatomically modern humans to those who were behaviorally modern. But the roots of language are likely older than that—and together with the even earlier emancipation of the hominids from the safety of the trees, the development of language probably had a greater influence than either tools or bipedalism on the rapid growth of the human brain.

The record is unclear which effect had greater effect. Surely, early communication probably related in complex ways to hunting and tooling; a jumbled feedback mechanism likely led to the enhancement of both. After all, symbolic communication, such as sign language and arm gestures (or "body language"), would have granted obvious advantages in coordinated, big-game hunting. Likewise, communication of some sort was likely needed to convey such simple skills as toolmaking, tool use, and weaponeering. Equally important, language ensured that experience, stored in the brain as memory, could be passed down from one generation to another. But when did language change from grunts, groans, and symbolic hand waving to the grammatically spoken word? As with much else in our cosmic-evolutionary story, it's often a matter of degree—shades of gray again: Did language emerge gradually over millions of years, or did it originate suddenly when needed much more recently?

Though anthropologists have little direct evidence, some type of primitive language could have been employed a million or more years ago. Linguists have tried to estimate the duration needed to justify the roots of worldwide linguistic diversity among the five thousand currently spoken languages, and they usually find a minimum of a hundred thousand years (even though the hominid breathing and swallowing plumbing were ready for nasalized speech much earlier). Erudite language surely came later than that, though again its origin and evolution are hard to pinpoint with any accuracy. The just-noted fifty-thousand-year-old blossoming of intellectualism—archaeology's "50K bloom"—is commonly supposed to have been caused by a combination of environmental change and genetic mutation—a classic biological-evolutionary advance—that then appended modern language and artistic expression to the behavioral repertoire of a species that had already had a big brain for more than a hundred thousand years. Or, by contrast, did language and art emerge solely as a cultural invention, like agriculture or pottery, having much less to do with biology?

To reiterate the generally accepted scenario: anatomically modern *Homo sapiens*, who looked like us and had virtually our same sized brain, resided in Africa some hundred-fifty thousand years ago, after which they evolved into the behaviorally modern but same human species, who later became dreamers, thinkers, and artists about fifty thousand years ago mostly in Europe. Whether this really was a sudden (punctuated) flowering or a more classical Darwinian (gradual) change is currently unknown. Work by archaeologists in this key transition period is underrated, underfunded, and under attack.

Artifacts made by early humans provide more clues to the cognitive-able and manual-skill prerequisites for such advanced communications as speaking and writing. Excavations of caves, mostly in southern Europe, have revealed a wealth of small statues and bones having distinctive markings or etchings. Predating the most famous and beautiful of the ice-age art depicting buffalos and horses on cave walls—the twenty-to-thirty-thousand-year old expertly rendered lifelike paintings at Lascaux and Chauvet in France—the oldest of these carvings date back about fifty thousand years, including a few stone statuettes displaying symbolic functions of some kind. Since Neanderthal man's larynx couldn't possibly have uttered the full range of sounds of modern, articulate speech, the etchings on these objects are thought by many archaeologists to represent a primitive mode of communication. This idea has been reinforced recently by close examination, which shows the same repertoire of distinct markings repeated on many of the statues and bones. Apparently, the markings are neither accidental nor decorative; they go beyond just art for art's sake and are indeed symbolic. Though sketchy, doodling cave and bone art might be thrice as old as that at Lascaux, these engraved stone artifacts are among the earliest known attempts to better—and to record—the body signaling and mental reasoning used not only now by modern chimps but also long ago by our ancestral australopithecines.

Credit for being the first to write texts usually goes to the Sumerians, those residents of a flat plain bordering the Persian Gulf in what was once called Mesopotamia. Nearly six thousand years ago, this ancient civilization had created an intricate system of numerals, pictures, and abstract cuneiform symbols. Thousands of baked (hence durable) clay tablets have now been unearthed, and each shows that a stylus of wood or bone was used to inscribe a variety of characters. Estimates of the Sumerians' basic vocabulary infer no fewer than four hundred separate signs, each denoting a word or syllable. Animals such as the fish and the wolf, as well as equipment such as the chariot and the sledge, are clearly depicted, but most Sumerian texts remain largely undeciphered. Because they display more than just pictures, the messages on these tablets already represent a reasonably advanced stage in the evolution of writing. Earlier civilizations perhaps responsible for inventing some of the Sumer symbols probably wrote exclusively on papyrus or wood that decayed long ago. Suggestively, modern computer studies that compare shared linguistic roots of known languages and likely rates of divergence among them (much like the biologist's tree of life) have recently revealed that the Hittite spoken language—the forerunner of Indo-European languages including English and all the Germanic, Slavic, and Romance languages—might well have been common among the Neolithic farmers of present-day Turkey as long ago as nine thousand years.

The prevailing view among anthropologists is that writing evolved from the concrete to the abstract, from that needed for survival to the more literary. Sometime between several tens of thousands of years ago and several thousand years ago, the pictures and etchings on the bones, statues, and cave walls became increasingly schematic, using a single symbol to represent an entire idea. This advance could have been deliberate in order to speed the process of record keeping, or it could have been the result of shorthand or carelessness on the part of ancient scribes. Whichever, it suggests that pictures preceded symbolic writing, which in turn led to the alphabetic prose of modern times, as in this book.

Evidence of early culture.

Artifacts made of chipped stone and carved bone date back some fifty thousand years. Here, among several cultural products of early humans found in recent archaeological expeditions, a small piece of reindeer rib shows deliberate markings or engravings of a reclining humanlike figure. Likely more than just doodlings of early artists, these statuettes probably had symbolic functions of some kind. *Source: Smithsonian.*

Statue and bone markings might also reflect the origin of ancient science. Some of the etchings of several tens of thousands of years ago seem to correlate with the periodic lunar cycle depicting phases of the Moon—perhaps among the first attempts to keep track of the seasons. Engravings of this sort, as well as large murals on the walls of caves, suggest that people of the late Stone Age were conscious of the seasonal variations in plants and animals. While some archaeologists prefer to interpret these deliberate bone markings as simple arithmetical games, these artifacts may well have been among the very first calendars—in effect, systematic timekeeping instruments.

The earliest *written* record unambiguously describing the use of scientific instruments doesn't appear for many thousands of years later. Three-thousand-year-old hieroglyphics partially document the Egyptians' knowledge of the sundial—which is really a clock, telling the time of day by noting the angle of the Sun's shadow—but the oldest sundial excavated to date is a Greco-Roman piece of stone built about two thousand years ago. The Greek and Roman sundials reveal key refinements over the earlier Egyptian models, marking both the hour of the day and the day of the year (yet neither the duration of an "hour" nor the design of the calendar were the same as now). These early scientific tools are significant in that they could predict the orderliness of daily, seasonal, and yearly changes on Earth, thus mapping the basic rhythm of the farming cycle.

Megalithic monuments, the most famous of which include the pyramids of Egypt and Stonehenge in Britain, can be used to predict the first day of summer, and possibly solar eclipses as well, by the alignment of big stones with the rising and setting of certain celestial objects. Many other structures like it, though not quite as grand, are scattered across Europe, Asia, and the Americas. In what is now Mexico and its Yucatan Peninsula, ancient "temple" pyramids have miniature portholes through which celestial objects, particularly the Sun, can be viewed at propitious times of the year. Astronomers of the Middle Age Mayan civilization were essentially priests, with the destinies of individuals, cities, and even whole nations apparently determined by the position and movement of celestial objects across of the sky.

By a thousand years ago, these Central American cultures had probably influenced many tribes of North American Indians roaming the plains of what is now the western United States and Canada. Recent studies of numerous "medicine-wheel" land structures made of boulders arranged in various patterns of rings and spokes clearly evoke the possibility of such cross-cultural ties. Though the actual use of these stone contours is unclear, as most Indians had no written language and hence no historical record, some archaeologists regard those structures to have been another kind of calendar that marked the rising of the Sun at certain times of the year. At least one of those tribes was looking up at the time, since it recorded on a rock face in Arizona the eleventh-century supernova event of the Crab Nebula, as noted earlier in the Stellar Epoch.

Thus, although modern scientific research, including instruments like the immensely useful microscope and telescope, dates back hardly more than four hundred years, we can be sure that pre-Renaissance doyens were adept in elementary astronomy, Euclidean geometry, mechanical engineering, and many other practical ventures. Indeed, the roots of technology go far back. Our ancestors of a millennia ago seem to have been a good deal more technically sophisticated than many modern researchers, until recently, have cared to acknowledge.

Mayan astronomer-priests of ancient Mexico bring to mind another factor that aided the cultural evolution of humans—a lofty factor that helps us know who we are and whence we came. The epitome of culture is the search for truth, or at least a reasonable approximation of reality, most notably the need to know ourselves and the world around us. Both the desire and the ability to undertake this search identifies us as humans, distinguishing us from all other known life-forms. To be sure, rational understanding is the major goal of science, though admittedly, science shares this goal with other disciplines. Religion, the arts, and philosophy, among other endeavors, all represent alternative efforts to appreciate our origin and being.

For thousands of years, humans have realized that the best way to dispel mystery is to understand it. Twenty-thousand-year-old cave paintings of southern France may be the oldest traces of early magico-religious ceremonies in Earth's dim recesses. These cave-wall images seem to depict rites in which elaborate myths perhaps linked the hunting men and the animals they killed.

Reliance on the supernatural and the use of mystical activities are more clearly documented near the very beginning of civilized history. Sumerian inscriptions of five thousand years ago record myths explaining how gods had created humans to be their slaves. This system guaranteed that food, clothing, and other necessities of life would be provided to priestly households or temples in order to please the gods, or at least appease them. Such a society divided managers from the managed, priests from the plebeians. Apparently, anyone professing knowledge of even the simplest celestial events was able to subjugate the masses. In the eyes of field workers, anyone who foreknew the seasons, for instance, must have had a special relationship with the gods and therefore deserved to be obeyed. No longer required to spend time producing their own food, the priest-masters of ancient Sumer were able to develop skills and knowledge far greater than humans had ever before attained.

Such mythmaking, perpetrated on the populace, forced further specialization, while assuring that legions of humans labored on social and technical tasks now recognized as vast irrigation projects and monumental templelike structures that even today rise high above the Mesopotamian plain. Likewise, Sumerian poetry of several thousand years ago clearly documents how the Sumer religion incorporated a systematic theology of human, worldly, and cosmic phenomena. Aspects of Nature—Sun, Moon, storms, thunder, and so on—were personified, with humanesque beings playing roles of gods in a divine political society, the whole of which was ruled by the resident god of the sky.

Such a set of beliefs proved powerful, for the system was complex and the populace uninquiring. For several thousand years, the priests of Mesopotamia bamboozled a largely illiterate public with increasingly intricate speculations. Even the surrounding barbarians, the ancestors of the ancient Greeks, Romans, Celts, Germans, and Slavs, were convinced that the gods of Sumer ruled the world. Apparently, myths become truths if upheld long enough.

In some ways, a stratified social system helped maintain cohesion and uniformity; priestly leadership offered a stabilizing influence, as do all religions in principle. But Sumerian inscriptions also tell of arguments, often prompted by water-rights disputes and fostered by rival religious factions and coalitions, that gradually became life-or-death struggles. By a few thousand years ago, dozens of Mesopotamian cities were armed to the hilt, with military organizations rivaling priestly governments. Presumably, the concept of kingship originated when Sumerian clerics decreed that the gods needed a representative among men—a strong-armed chief priest of sorts to adjudicate quarrels or to crush the opposition. Not long thereafter, clustered societies in the form of those cities, united under the influence of one king or one god, began demanding adherence to this, that, or another thing, the objective being to force our ancestors to believe who they were, where they came from, and how they fit into the bigger scheme of things.

Today's plethora of differing religions and philosophies testifies to the fact that theological beliefs and resolute ideas are not subject to experimentation, not open to objective testing, and thus will not ever be universally acceptable. Such dogma offers stabilization locally, perhaps, but how can these ideologies fail to destabilize globally, especially given conflicting sects and variant viewpoints in our modern world? The upshot is that neither belief alone nor thought alone can ever make the unknown known.

Ultimate reliance on the authority of the experimental test is the one key feature that clearly distinguishes the scientific enterprise from all other ways of describing Nature. Initiated in ancient Greece and used sparingly throughout the centuries, most notably by Aristotle and Augustine, the scientific method rapidly blossomed during Renaissance times to become one of the primary criteria in our search for truth in the physical Universe. Today, it is *the* primary criterion used to generate the millennial worldview of cosmic evolution—not a tradition of revealed beliefs demanded of people but a chronology of discovered events offered to them.

If the epitome of culture is our ability to seek the truth about ourselves and our world, then an even more startling advance concerns humankind's *desire* to know the truth—our manifest curiosity about our origin, being, and destiny in the wider Universe. We are unique among all known living things in pondering who we are, whence we came, and where we might be headed. Just what is it that allows us, even drives us, not only to ask the fundamental questions but also to attempt actively to find solutions to them? The answer seems embodied within that hard-to-define refinement of the mind called consciousness, that part of human nature that permits us to wonder, to introspect, to abstract, to explain—the ability to step back, perceive the big picture, and examine how our existence relates to the existence of all things.

How did consciousness originate? When did humans become aware of themselves? Is consciousness a natural, perhaps even inevitable consequence of neurological evolution? Some cognitive scientists think so, but they cannot yet prove it. They maintain that consciousness equates with "mind" and that minds are just what brains do. That is, the mind, or consciousness, is simply the normal functioning of a structured brain—the whole brain, not necessarily a seat of consciousness at any localized neural site. Nothing "extra" infuses the mind or consciousness, just as no vitalism or élan vital informs life. Consciousness is merely accumulated experiences.

Other researchers demur, noting that some specific, perhaps unlikely mechanism is needed for the development of consciousness. They claim that consciousness is likely to be more than the behavior of a huge collection of interacting neurons and that it may well be confined to isolated areas in the brain. The capacity for imagery and imagination does seem to entail something more than a gradual clustering of neurons—something akin to a holistic property resembling a global neural network that emerges when the whole indeed exceeds the sum of its lightning-quick parts.

Records of ancient history are incomplete and unreliable, making difficult any attempt to document the onset of self-awareness. Some psychologists contend that consciousness, as we know it, is not part of ancestral records until a few thousand years ago. This is about the time when some of the writings in ancient texts became abstract and reflective. People may well have wondered about themselves several millennia ago, but it's unclear if that was the first time they began doing so. If consciousness did originate so late, we must be prepared to assume that cultures can become highly refined without evolving personal consciousness. Our early ancestors would have had to invent just about every cultural amenity except consciousness and to have lived until quite recently in a dreamlike, essentially unconscious state.

Yet others argue that human consciousness developed much longer ago, perhaps well more than merely thousands of years back in time. If modern chimpanzees, which display rudimentary self-awareness, do, in fact, mimic our australopithecine forebears, then embryonic consciousness could have emerged millions of years ago. To be sure, it may extend back tens or even hundreds of millions of years, since there's no justification for the widespread assumption that basic awareness is a uniquely human trait. Animal behaviorists have acquired a great deal of evidence showing that simplified consciousness is common among many animals, as amply demonstrated daily by domesticated dogs, cats, and even perhaps invertebrate insects in our homes.

That full-fledged consciousness evolved at some intermediate time—tens of thousands of years ago, at about the time of the invention of the bow and arrow—is another popular assertion. Some neurobiologists have emphasized that the extension of the hand beyond the body—even merely the art of throwing perhaps—might have had profound effects on key brain qualities, such as foresight, planning, and other actions at a distance. Ballistic skills do seem to offer a particularly attractive biocultural route toward the evolution of advanced neural machinery. Ironic indeed if it were out-of-body experiences in the form of the long-distance weapon that acted as a giant step that finally granted humans the freedom to innovatively plan ahead, to begin to evolve culturally, to wonder about space and time.

A reconciliation of these seemingly divergent views holds that crude consciousness originated among the prehuman animal kingdom millions of years ago and that prehistoric humans evolved a better sense of consciousness not more than a million or so years ago, but only much more recently did humanity per se become sophisticated enough to reveal that sense of wonder and self-awareness in their writings. Perhaps. In the absence of objective empirical data and controlled experimental tests, we are left wondering how we learned to wonder—how humankind became so curious about ourselves and our surroundings.

The precise path of human evolution during the past million years is tricky to follow in detail, for the causes of recent evolution include both biological and cultural factors, often the two intertwined as biocultural effects. What's more, whatever truly made us human involved not only such tangible inventions as noted above but also the creative products of emotion and imagination—qualities virtually impossible to define objectively. A jumbled feedback system came into play among the increase in brain volume, the discovery of technical skills, the advance of cultural amenities, and the development of verbal communication and social organization. Changes were slow at first but have markedly accelerated within the past hundred thousand years or so. Whatever the reasons, these many innovations have enabled *Homo* to enjoy unprecedented success as a life-form on planet Earth, for we alone can ask fundamental questions—and attempt to answer them.

∞

The most recent one percent or so of human history—the past ten thousand years—has seen major and rapid cultural innovations. That much is clear from a whole host of prewritten records, from exquisite artifacts to discarded trash, from ornamental beads to crude weaponry, and much, much more. The glaciers had retreated, creating warmer and wetter environments, thus allowing the land to flourish. Our hunter-gatherer forebears followed the spreading flora and fauna, occupying, even if sparsely, every part of the globe except the poles. All the while they were toying with tools and refining ideas to enhance their survival. But of all the factors that contributed to the rapid rise of modern humans, the invention of agriculture was surely among the most important—some say *the* most important, as it represents a pivotal milestone on the road to modernity. Tilling the land made available a reliable source of food to feed growing numbers of people on Earth. In short, scavengers had transformed into agriculturists, beginning with the domestication of plants and animals some hundred centuries ago.

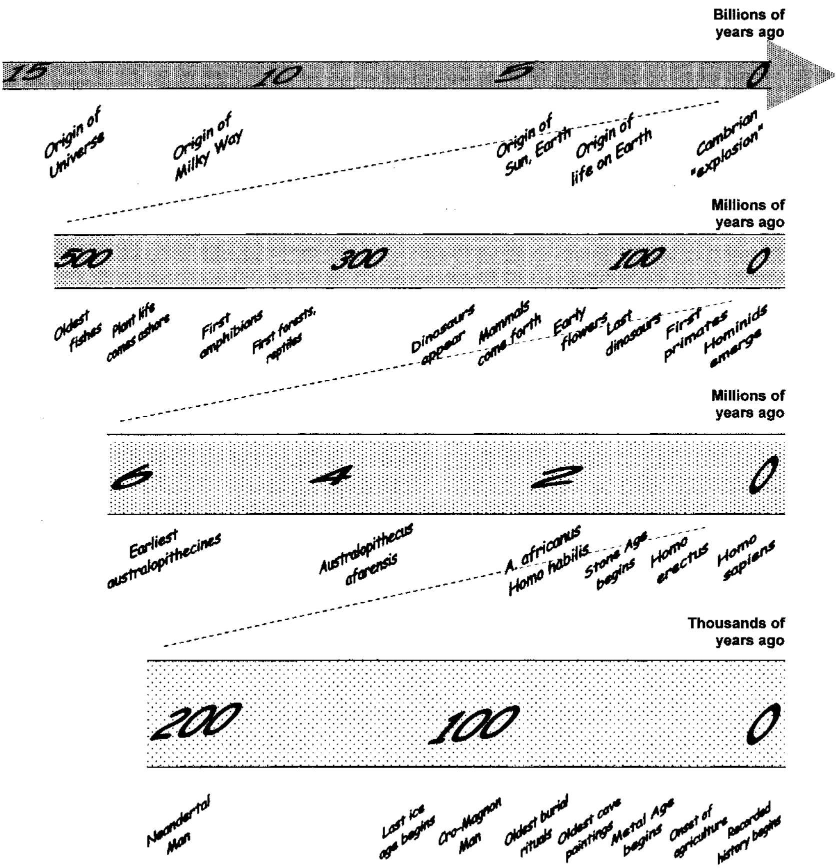
Archaeological data show clear evidence for whole new methods of subsistence by eight thousand years ago, initially among the hills of the Fertile Crescent, from the eastern Mediterranean to the Caspian Sea, spreading quickly thereafter into the Near East and western Europe. Systematic crop planting and livestock raising near stable village settlements fostered swelling populations in agricultural locales (eventually to become cities) across Eurasia. Not only did absolute numbers of additional people survive, but many more were migrating to eventually colonize virtually every nook and cranny on the planet. The tricks of the trade of agriculture spread like a weed on the wind from Asia and the Aegean, reaching within a few thousand years many distant places (or perhaps emerging independently in them), including China, Mexico, and South America. Hunter-gatherers had indeed given way to farmers and herders of mainly cattle, sheep, maize, rice, and wheat, the last of these the most valuable single crop in today's modern world. Food-production jump-started society: trading flourished; economy rose; population soared. Change was rampant as the so-called Neolithic Revolution was well underway—though even here those cultural changes were probably more gradual—to repeat that trite though apt phrase, more evolution than revolution. Urbanization and ultimately industrialization were not far behind.

Civilization had moved into high gear. It had taken an awfully long time after life originated, but highly organized and manipulative life-forms had finally arrived. One thing led to another: Lifestyles multiplied; cultures proliferated; technologies advanced. Aided by irrigation systems built alongside river valleys, farming skills developed dramatically. The human population rose yet more rapidly, especially in urban areas along waterways such as the Nile River in what is now Egypt and the Tigris and Euphrates rivers running through what is now Turkey, Syria, and Iraq. Specialized crafts were refined to serve the populace of these growing communities: metalwork, ceramics, shipbuilding, and woodcuts all show up clearly in the archaeological record beginning some six thousand years ago. That record is best documented for southwestern Asia (the Middle East), yet pragmatic and artistic progress likely ensued at other geographical centers as well.

Although much of this preceded recorded history, urban societies and mature economies, as well as complex social and political systems, were surely the rule, not the exception. Agriculture, industry, and commerce were fully established several thousand years ago. And they have persisted to this, the twenty-first century. We have reached the here and now in our cosmic-evolutionary narrative.

Be sure to place into perspective these later developments of civilization's ramping up toward greater complexity. Thousands, tens of thousands, even millions of years tend to merge into a temporal blur after a while. To appreciate the time frame for some of the greatest highlights of natural history, consider the following analogy:

Imagine the entire duration of Earth to be fifty years, instead of nearly five billion, thereby making each megacentury a "year." Comparable to a human life span, this compressed time scale then allows salient features of Earth's history to become more comprehensible. In this analogy, no record whatever exists for nearly the first decade. Rocks hardened relatively soon thereafter, as the environment subsided appreciably during those first ten years. Life originated some thirty-five years ago, when Earth had hardly become a teenager in our analogy. The planet's middle age is largely a mystery, though we can be reasonably sure that life continued to evolve, or at least persist, and that tectonic events continued to build mountain chains and oceanic trenches.



. . . a microscope would be needed to see the highlights of recorded history and cultural advance . . .

Not until about six years ago, in our fifty-year analogy, did abundant life flourish throughout Earth's oceans. This same life didn't come ashore until about four years ago. Leafy plants and squeaky mammals prevailed across the surface no more than two years ago. Dinosaurs reached their peak about a year ago, only to disappear suddenly eight months ago. Humanlike apes became apelike humans only last week, and the latest major ice age occurred only yesterday. Modern *Homo sapiens* didn't appear until several hours ago. In fact, according to the compressed timescale of this analogy, the onset of agriculture is only about one hour old, all of recorded history a half hour, and the Renaissance a mere three minutes in the past.

In short, a microscope would be needed to see the highlights of recorded history and cultural advance on a temporal scale spread across any one page of this book. Yet it's within that shrunken duration of contemporary times that we humans have richly probed, reasoned, and discovered much about our conscious selves and the expansive Universe beyond.

∞

Social and cultural evolution, on the opposite side of the evolutionary spectrum from galactic and stellar evolution, has brought us to us—third-millennium *Homo sapiens sapiens*—the most intricate cluster of natural matter yet known. This is no anthropocentric statement; no evidence whatever implies that humankind is the pinnacle or endpoint of cosmic evolution, nor are we likely the only sentient beings in the Universe. Yet, while pondering the grand synthesis of radiation, matter, and life in our undoubtedly incomplete inventory, today's technological society is currently poised at the most complex extreme (thus far) of the cosmic-evolutionary scenario. That's not to say that the scenario is done, if ever it will be. Nature continues to write the story, and we continue to unravel it.

Cultural evolution tracks the changes in the ways, means, actions, and ideas of society, including their transmission from one generation to another. Called "memes" by some, in loose analogy to genes, these are cultural replicators, such as a new word or song invented by one person and mimicked by others. To be sure, many of the cultural traits already noted, including the construction of useful tools, the teaching of elaborate language, the practice of viable agriculture, and not least the discovery of controlled energy, have been imitated and refined over scores of generations. The bulk of this newfound knowledge was transmitted to succeeding offspring not by any genetic-directed inheritance but by use and disuse of information available to intelligent beings.

A mostly Lamarckian process (based on the ideas of the nineteenth-century French evolutionist, Jean-Baptiste de Lamarck), cultural evolution proceeds via the passage of acquired attributes. Like galactic, stellar, and planetary evolution before it, cultural evolution involves no molecular reactions as in chemical evolution and none of the genetic inheritance typical of biological evolution. Culture enables animals to transmit favorable traits and survival skills to their offspring by non-genetic routes. Information gets passed on behaviorally, from brain to brain. Human culture itself is shaped by the sum total of human minds, often acting imitatively and cooperatively, sometimes over the ages but at other times in a single generation. The upshot is that cultural evolution acts much faster than biological evolution. Genetic selection operates little, if at all, in evolutionary realms sandwiching neo-Darwinism, where adaptive and selective pressures clearly dominate. Even so, a kind of selection was, and is, at work culturally. The ability to start a fire or throw a spear, for example, would have been a major selective asset for those hominids who possessed them, an asset transmitted not by genes but by memes. Perhaps more than anything else, memes are what sets us apart from other species.

As different as they are, biological and cultural evolution are not unrelated, as might be expected for two adjacent phases of cosmic evolution. Somewhat surprisingly, though, these two phases enjoy a subtle reciprocal interplay. Discoveries and inventions may well have been made by talented individuals having the "right" combination of genes, but once made, an invention such as lighting a fire or sharpening a tool would have, in turn, granted a selective (i.e., reproductive) advantage to those better endowed genetically to handle the skill. The two kinds of evolution thus partially complement one another, although in the recent history of humankind, Lamarckian (cultural) evolution has clearly dominated Darwinian (biological) evolution. Cultural acquisitions spread much faster than genetic modifications. Our gene pool differs little from that of the Cro-Magnons some twenty thousand years ago, yet our cultural heritage is a good deal more robust in the knowledge, arts, traditions, beliefs, and technologies acquired and transmitted during the past thousand or so generations.

That cultural and social changes represent the most complex phenomena in the known Universe is undeniable. Human behavior, now engulfed by heavy energy use and rapidly changing environments, is what makes social studies so difficult. Unlike in much of the physical and biological sciences, experiments in cultural evolution—humans interacting (social psychology), cities functioning (urban planning), or nations warring (political economics)—are virtually impossible to model objectively. Just observing social behavior, let alone experimenting with it, is vastly harder to accomplish than probing molecules in a chemistry laboratory or sending spacecraft to the planets. Likewise, the number and diversity of factors influencing the outcome of a human interaction is far greater than those affecting the birth or death of a star. Although a physicist or chemist might never have a concern for an individual atom or molecule, sociologists often treat the human behavior of a single individual as paramount—and that's what makes their task all the more difficult, and complex, for not even statistical reasoning can help much.

A few examples of Lamarckian-style cultural change toward greater complexity will suffice. Consider first the earlier-noted and paramount exemplar of culture: language. Language is transmitted largely through the media of teaching and schooling, ensuring that knowledge and experience stored in the brain as memory is accumulated by one generation and transferred to the next—not perfectly but adequately (including imitatively) over the years. Not only do we transfer knowledge to our children in this way, but the body of available knowledge itself also grows with the acquisition of new ideas, data, and stories. And because that knowledge accumulates faster than it's forgotten (especially with the onset of recorded history), the sum total of culture passed along builds, indeed grows copious yet convoluted. That's why it now takes a third of a human lifetime to train for a doctorate in science, despite narrow specialization. Human knowledge today far exceeds that of any one individual. Hardly any of our cherished educational facts, models, and methods are transmitted via biochemistry of the genes. Those genes do grant a hard-wired ability to learn from other human beings, but learning itself is often a surprisingly long, tough struggle, usually overcome by hard work—which is energy. The story of cosmic evolution itself is a cultural myth hereby told in the form of this book—a myth, because it's admittedly a simplification of an extremely elaborate approximation of reality.

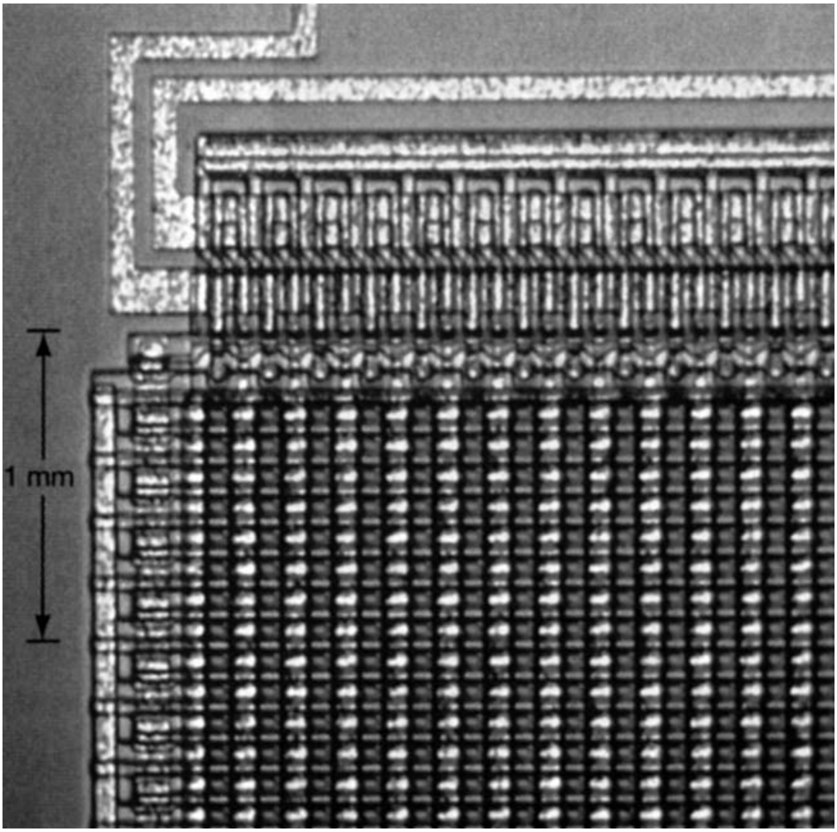
Industrial development is another cultural practice that increases order locally in the form of artificially manufactured products, yet only with the sweat and toil of spent energy, which inevitably increases the disorder (entropy) of the larger environment of raw materials used to make the goods. Modern automobiles, for instance, are better equipped, sounder mechanically, and basically safer than their decades-old precursors not because of any internal tendency to improve but because manufacturers have constantly experimented with new features, keeping those that worked well while discarding the rest—a clear case of acquiring and accumulating successful features from one generation of cars to the next. A kind of selective pressure functions by means of dealer competition and customer demand in the social environment—selection as human preference in the marketplace—the evolutionary mechanism being more Lamarckian than Darwinian. Lamarckian tinkering can surely improve technology: use and disuse engenders gradual change in automotive style, operation, and safety, all of which feed back to increase the pace of our lives and the thrust toward even more complexity. Would anyone deny that today's gadget-filled Mercedes (with computer-assisted fuel injection, electronic valve timing, and microprocessor-controlled turbochargers, not to mention all those widgets on the dashboard) is more complicated than Ford's Model-T of nearly a century ago, or that more energy is expended (per unit mass) to drive it?

Cultural evolution, though richer and more complex than either physical or biological evolution, need not be treated any differently than any other aspect of cosmic evolution. Energy flows through open systems help create and maintain rising complexity throughout our fast-paced human society. Cultural evolution occurs mostly when societies alter their organizational posture while responding to changes in the use of energy, for it's increased order that helps dynamically stabilize a far-from-equilibrium society such as ours. Energy expenditure per capita clearly rose in relatively recent times—from hunter-gatherers of a couple of million years ago, to ancient forebears who managed to control the use of fire hundreds of thousands of years ago, to pioneering agriculturists some ten thousand years ago, to industrial revolutionists of a few centuries ago, in turn to our fossil-fuel-driven society of today. Per capita energy use has increased by nearly a factor of a hundred during the past few million years, much of that change in fact occurring during just the past few millennia.

The rise in normalized energy flow has been an evolutionary, competitive process in which selection has once again played a role. New technologies drove older ones to extinction, thereby benefiting humankind over the ages. Throughout the past few centuries, for example, businessmen chose shorter travel times, lower transportation costs, and heavier shipping loads; steam-powered iron ships replaced the wind-powered clipper ships, and now air travel has replaced them both. Likewise, "horsepower" provided literally by horse and mule was first marginalized and then eliminated as steam and eventually gas engines became the impetus on most farms the world over; people elected to concentrate their energy for greater efficiency. Typewriters, ice boxes, and slide rules, among many other innovative advances in their own time, were selected out of existence by the pressures of customer demand and commercial profit, often replaced initially by luxuries that eventually became necessities, such as word processors, electric refrigerators, and pocket calculators. Yet all this progress, which did better the quality of life (as measured by health, education, and welfare), came at the expense of greatly increased energy consumption and its consequent toll on the environment.

Cities, states, and nations can be treated in thermodynamic terms, for these are also open structures regulating the flow of energy in and out. They acquire and consume resources, produce and discard wastes, all while employing energy for a variety of services: transportation, communication, construction, medicine, comfort, and entertainment, among a whole host of maintenance tasks. Modern cities are as much a product of evolutionary events as any galaxy or organism, and many are still developing, seeking to establish dynamically stable communities within our planet's larger, vibrant ecological system. Their populations are dense, their structures and functions highly complex; cities are voracious users of energy.

Economies, too, are products of evolution. They are social organizations that seek to manipulate the environment for increased resources, enhanced efficiency, and greater productivity. The emerging interdisciplinary subject of ecological, or evolutionary, economics highlights the celebrated concept of energy flow (including material resources), much as energy flow commands the interdisciplines of astrophysics and biochemistry. All such ordered systems exist uneasily "on the edge," from unstable giant stars to struggling life-forms to endangered ecosystems. This is precisely the way that all physical, biological, and cultural systems act as dynamic steady states—as sources of novelty and creativity, enabling them to take advantages of opportunities to advance along the



Evidence of technological intelligence.

Silicon chips are now the symbol of our modern, energy-consuming society. There are probably more such chips now functioning in today's world than bricks in all the world's buildings and walkways. Chips manage the information age in which we live, from the computers we use to the automobiles we drive; their presence is ubiquitous, including in cameras, televisions, phones, and so many other miniaturized gadgets relied on daily. Machines, and the energy they consume, drive our modern civilization. *Source: Dept. of Defense.*

arrow of time and the scale of complexity. The twin elements of randomness and determinism, once more, are also why realistic economies will never be predictable in detail but will remain largely dynamic, flexible, and always evolving. Chance and necessity mix yet again, much as they've guided perpetual change from the big bang to humankind.

Throughout the past tens of thousands of years, biological and cultural evolution have been inextricably interwoven. Their interrelationship is natural, much as particulate and galactic evolution interacted, stellar and planetary evolution overlapped, and chemical and biological evolution too, for the development of culture admits heavily one of those key factors affecting all of evolution—the environment. Here, though, cultural inventions enabled our immediate ancestors to circumvent some environmental limitations: Hunting and cooking allowed them to adopt a diet quite unlike that of the australopithecines. Clothing and housing permitted them to colonize both drier and colder regions of Earth. Tools and equipment enabled them to explore places for which they were not biologically adapted. Increasingly, human life-forms learned to manipulate their localities—to alter the environment as much as the environment altered life—a hallmark of the Cultural Epoch.

Likewise, cultural innovations now enable present-day humankind not merely to end-run the environment but also to challenge it directly—indeed, to expand our environment, for we are not only an exploratory species but an expansionary one as well. Technology allows us to fly high in the atmosphere, to live deep within the oceans, to probe the subatomic world, to communicate across continents, even to journey beyond our home planet. Change now quickens yet more, and with it, the pace of life. Culture, it would seem, is a catalyst, speeding the course of evolution toward an uncertain future.

What's more, not only does change on Earth continue, it has recently done so more rapidly than a lifeless Nature would have. Humankind itself has now become part of that change—altering it, selecting it, accelerating it. We are now less at the mercy of the environment than conversely, for technological beings have gained some mastery over matter. We have become the agents of change, the drivers of cultural evolution.

If any one factor has characterized the evolution of culture, it's probably an increasing ability to extract energy from Nature—not merely to capture energy, rather to store it, to transfer it, to use it more efficiently. Over the course of the past ten thousand years, humans have steadily mastered wheels, agriculture, metallurgy, machines, electricity, and nuclear power. Soon, solar power will emerge in its turn; all intelligent civilizations, anywhere in the Universe, likely learn to exploit the energy of their parent star. Each of these cultural innovations has channeled greater amounts of energy into society; energy use grew sixteenfold during the twentieth century alone.

The ability to harness energy and thus order our daily lives are defining characteristics of modern society. But energy use is also a source of rising disorder in our surrounding environment—global pollution, waste heat, and social tumult, among other societal ills. Ironically, the need for increased energy and natural resources so vital to our technological civilization is also a root cause of many of the sociopolitical problems now facing humankind at the dawn of the new millennium.

Earth now finds itself in a delicate balance. Our planet harbors a precarious collection of animate and inanimate systems on localized scales amid a complex web of global energy flows in a larger, cosmic setting. All these systems—whether entirely natural or humanly built—need to heed the laws of thermodynamics as an unavoidable ground rule. Consciousness, too, including societal planning and technological advances likely to dominate our actions well into the future, must embrace an evolutionary outlook, for only with an awareness and appreciation of the bigger picture will we likely survive long enough to have a future.

∞

Physical, biological, and cultural evolution span the spectrum of complexity, each forming an essential part of the greater whole of cosmic evolution. Stars, planets, and life, as well as culture, society, and technology, all contribute to a magnificently coherent story of ourselves, our world, and our Universe. All these systems, among many other examples of order and organization in the richly endowed cosmos, share common features, common drives, and a common evolutionary epic.

The precise path of human evolution during the past few million years of the Cultural Epoch is controversial in its details. The scientific data are sketchy, the historical record incomplete; specifics have been lost to the march of time. As best we know presently, modern humankind likely spread from Africa some hundred-fifty thousand years ago—much as we are again today dispersing, however haltingly, while exploring the Solar System—but little is certain about the fossil remains or genetic diversity of our immediate forebears. What's more, that which truly made us human involved the creative traits of emotion and imagination—qualities difficult to discern objectively. Understanding the tempo and mode of human evolution, including neural evolution of the brain, remains anthropology's greatest challenge.

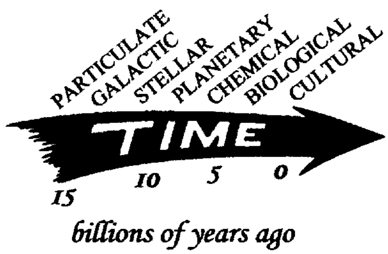
The causes of recent evolution include not only biological factors but cultural ones as well. An intricate biocultural interplay accompanied the increase in brain volume and the invention of technical skills, the development of verbal communication and the onset of social organization, the creation of subjective art and the discovery of objective inquiry. These changes were slow at first, but they have markedly accelerated within the past quarter-million years. Reduced to essentials, culture is the only difference between hunter-gatherers and ourselves. Cultural innovations have enabled *Homo sapiens* to become unprecedentedly complex as a life-form on planet Earth. We alone can ask, moreover explore—religiously, philosophically, and scientifically—the fundamental questions regarding deep origins.

The matter networked within the human brain is more complex than anything in the known Universe. Many breakthroughs were needed along the road to intelligence, most notably the ability of cells to cluster and interact starting about a billion years ago. More recently, tree swinging, manual dexterity, binocular vision, fire, tools, speech, writing, foresight, and curiosity, among many other evolutionary advances, all had a clear effect on the brain: it got bigger. Neurons have evolved steadily to the point where we now use them to steer our technological civilization, to unlock secrets of the Universe, and to reflect upon the material contents from which we arose.

Without a brainy seat of consciousness, galaxies would twirl and stars would shine, but no one would know it. Nothing could likely comprehend the majesty of the reality that is Nature. By contrast, with a conscious, curious brain, we probe the past and discover our history, striving to decipher our celestial roots and the routes that brought us to the here and now, all the while searching for a better understanding of both the cosmos and ourselves. What emerges is nothing less than a cosmic-evolutionary heritage—a plenary worldview of who we are, whence we came, and how we fit into the universal scheme of all material things.

Epilogue

A WHOLE NEW ERA



THE SCENARIO OF COSMIC EVOLUTION is a human invention. It's a long and spectacular story, an evolutionary epic that includes the storyteller. Despite its seven major epochs, this grand narrative was not handed to us on a stone tablet atop some mountain. The scientific community has gradually deciphered the story, is now telling it forthrightly, and continues to refine it as we learn more.

Nor is the idea that we are children of the Universe a new one. That notion may be as old as the earliest *Homo sapiens* to contemplate existence. Nor is the underlying concept of change especially novel, persistent as it was through the ages, to be sure throughout all the epochs arrayed along the arrow of time. The idea of ceaseless, everlasting change—somehow causing, forcing, or allowing beings to become—was philosophically embraced thousands of years ago.

While entering the new millennium, we can now begin to identify scientifically some of the major astrophysical and biochemical events, indeed changes, that demonstrate the cosmos as the origin and source of our reality. The intellectual approach is decidedly interdisciplinary, interweaving knowledge from virtually every subject universities offer. And—of great import—many parts of the lengthy and ongoing scenario sketched here have recently been confirmed with experimental and observational evidence.

Cosmic evolution is an inclusive working hypothesis that strives to integrate the big and the small, the near and the far, the past and the present, into a unified whole. Though many details remain outstanding, the overall conceptual framework of existence, including the rise of complexity from radiation to matter to life, seems reasonable and comprehensible.

Consider once more the broadest view of the biggest picture. In the earliest epoch of the Universe, radiation dominated matter. During the Radiation Era, intense light ruled all. Whatever matter existed at the time did so only as widely scattered elementary particles in a sea of blinding radiation. Vast amounts of radiant energy produced a spectacularly bright fireball inside which no atoms, stars, or ordered structures of any kind could have formed.

As the Universe expanded, it naturally cooled and thinned. Having evolved from radiation, matter gradually began assembling into individual atoms and then eventually clusters of atoms. An event of incomparable significance occurred when matter first began coalescing a few thousand centuries after the Universe originated in a naked singularity. The emergence of organized matter as the leading entity is the first great transformation in the history of the Universe. This change was fundamental and preeminent, an absolutely key part of the big picture.

From the start of the Matter Era, matter then dominated radiation; it controlled most events, even in the presence of radiation. And matter has governed radiation ever since, successively and successfully forming galaxies, stars, planets, and life. The means by which these myriad structures evolved, triggered by chancy fluctuations in changing environments yet guided by energy exchanges obeying deterministic laws, are neither magical nor mystical, for we nearly understand them.

There's no stopping the arrow of time, that manifest yet indefinable flow against which cosmic evolution unfolds. Born of a titanic event some fourteen billion years ago, time itself, as best characterized by the expanding Universe, is a Prime Mover—an underlying, neo-Platonic driver that permits order, shapes structures, and fosters complexity, at least for localized systems within a cosmos irreversibly and relentlessly decaying toward immense disorder. We see in our data, and increasingly so, a rich natural history of past phenomena steadily unveiling. The Universe itself may not be making progress, but we sentient beings most certainly are while discerning it.

Of all the known clumps of matter in the Universe, life-forms are surely the most fascinating, especially those enjoying membership in advanced technological civilizations. And that's not an anthropocentric statement; the arrow of time is not pointing at us. Technically competent life differs notably from lower forms of life and from other types of matter strewn throughout the Universe, not only because we can manipulate matter and radiation but also because we can tinker with evolution itself.

Given enough time, even evolution evolves.

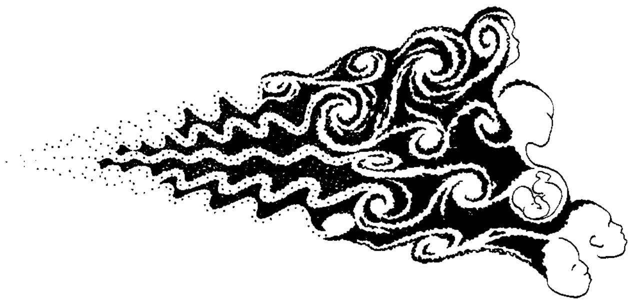
Be assured, stellar evolution continues unabated in the cores of stars everywhere. Chemical evolution occurs in such remote sites as galactic clouds and exotic moons. Biological evolution persists for most species on Earth and possibly on distant planets as well. And cultural evolution endures in many corners of our world and conceivably on alien worlds beyond. But for technologically intelligent life, evolution per se is undergoing profound change.

Whereas previously the gene (strands of DNA) and the environment (whether physical, biological, or cultural) guided evolution, we humans on planet Earth are rather suddenly gaining control of both these agents of change. We are now tampering with matter, diminishing our planet's resources while constructing the trappings of utility and comfort. And we are now beginning to manipulate life itself, potentially altering the genetic makeup of human beings. The physicist unleashes the forces of Nature; the biologist experiments with maps of genomes; the psychologist influences behavior with drugs. We are, in fact, forcing a change in the way things change.

The emergence of sentient, technological beings, on Earth and perhaps elsewhere, heralds a whole new era—the Life Era. Why? Because technology enables life to begin to control matter, rivaling that previous transformation when matter began uncoupling from, and then dominating, radiation more than ten billion years ago. In turn, matter is now losing its dominance, if only at those isolated places where technologically intelligent life resides. To be sure, given the gadgets and goods of our modern lives, human beings have literally taken matter into our own hands, granting ourselves the option of a grand and glorious future, or perhaps one marked by self-destruction, devolution, and death.

For clarity, the onset of the Life Era coincides not with the origin of life itself, nor even with the emergence of humanity or consciousness.

Rather, it's an event in spacetime when technological life-forms begin manipulating matter more than matter influences life, much as matter eventually came to dominate radiation earlier in the Universe. For humans, this novel event is here and now.



The emergence of sentient, technological beings, on Earth and perhaps elsewhere, heralds a whole new era—the Life Era.

The transition from the Matter Era to the Life Era will not be instantaneous. This change is a demonstrably evolutionary, unrevolutionary, process. Just as much time was needed for matter to conquer radiation in the early Universe, long durations will likely be needed for life to best matter. Life might not, in fact, ever fully dominate matter, either because civilizations fail to gain control of material resources on truly galactic scales or because the longevity of technological civilizations everywhere is inherently short.

Though a mature Life Era may never come to pass, one thing seems certain: our generation on planet Earth, along with any other neophyte technological life populating the Universe, stand on the verge of slowly becoming a meaningful factor in the future evolution of the Universe. This is a transition of astronomical significance—a fundamental change from matter's dominance to life's dominance—indeed the dawn of the second great transformation in all of history. A quintessential event in the evolution of the Universe, it's the threshold beyond which life-forms can begin to fathom at least their position, and perhaps their role, in the cosmos. Accordingly, we have an obligation, a moral responsibility to survive, especially if we are alone in the Universe. The great experiment that intelligent life represents must not end in failure.

∞

Humans are undeniably the highest form of intelligence on planet Earth. That's not to say that we shall inherit the Earth; the microorganisms, however meek and submissive, may well outlast us, indeed overwhelm us. Yet, currently, we are the only species able both to communicate culturally and to construct technologically. We are the only ones capable of knowing our past and worrying about our future. Even so, intelligence is one trait, wisdom quite another. Just how wise are we, and how shall we use that wisdom to ensure the survival of life in general and humanity in particular?

What about the future? Where do we go from here? Though these are not easy questions, there is one surety: our Sun is destined to run out of fuel, balloon into a red-giant star, and engulf several of its planets, perhaps even Earth. That will doubtless end civilization on our planet and perhaps extinguish any life everywhere in the Solar System.

On the other hand, Earth will not become unbearably hot for another billion years and the Sun itself not perish for at least another five billion years, times so remote by human standards as to be nearly incomprehensible. That leaves plenty of opportunity for intelligent life on Earth, should it endure despite our technological intelligence, to undertake galactic engineering projects and other grand ventures literally out of this world.

What about shorter timescales, say a million, a thousand, or even a few hundred years ahead? Is there any way at all to predict further evolution of *Homo sapiens* while extending our trek along the arrow of time? Frankly, it's doubtful, as chance inevitably mixes with necessity, randomness with determinism, in ways that make outcomes uncertain. Yet actions and attitudes can potentially aid our longevity—practical steps meant to bolster our viability, our survivability.

Surely, we would be conceited, vainglorious, and downright pretentious to regard ourselves as the final product of cosmic change, the very pinnacle of cosmic evolution. Anthropocentrism—we must repeat—is not part of our story, expressed or implied. Nor will change stop now with our emergence, however commanding our being may be. Change has accompanied time's passage since the very start of the Universe. Just because technologically intelligent life has achieved dominance on a single planet, there's no reason to think that change will now cease. Change has been, and will likely remain, a hallmark of Nature, quite synonymous with time itself.

Change must furthermore persist for all tomorrows if we are to survive as a civilization. Change or perish: that's the primary code for the continued viability of all matter, including life—an essential feature of our cosmic-evolutionary scenario, and a vital message to take from it.

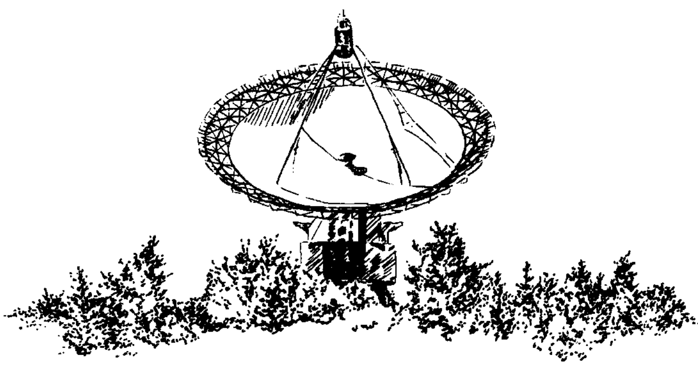
The future is a tricky subject and comment about it runs the risk of saying nothing concrete. Futures are especially troublesome to foresee when life is involved. As a case in point, predicting the fate of our civilization is actually harder, and a good deal more foolhardy, too, than predicting the fate of the Universe. It may sound ludicrous that we can know more about the future of the Universe in toto than about the future of life on our own planet. But the behavior of human life weighs heavily on our civilization, while likely not at all on the whole Universe. And while the Universe obeys the laws of physics, civilizations legislate their own laws.

The fate of the Universe depends on only one factor—its energy (or mass) density, a term that scientists are struggling to understand better and whose value we are now trying to estimate. The destiny of civilization also seems to depend heavily on a single term. But that term is *humanity*, a complex state surprisingly tricky to specify and nearly impossible to quantify. Even humanity's nature is defined tautologically: "the condition of being human, the quality of being humane; the kind feelings, dispositions, and sympathies of humankind."

Here's another way of perceiving the riddle of the future before us. As noted earlier, the business of the classical physicist is to comprehend Nature well enough to predict the response of matter to a variety of circumstances. The route of a baseball moving through air, for example, is now precisely understood. Knowing the mass of the ball and of the Earth, the gravitational force between them, the air pressure and resistance, the ball's momentum and spin, and a few other physical factors affecting the ball, we can model with great accuracy the future trajectory of this piece of matter through space. By contrast, to predict the "trajectory" of *life* through *time* is a much tougher puzzle. Too many nonphysical causes—individual and group sociology, national and international politics, biological and cultural attitudes, among a host of other unquantifiable parameters—will doubtlessly affect the future of civilization.

Once again, at issue is timing—truly a central theme throughout this book. Can our civilization get its act together in time to ensure its own survival? In the language of the evolutionist, shall we be selected by Nature to endure? Not inconceivably, a grander selection process may well be at work in the cosmos—a principle of cosmic selection, akin to Darwin's natural selection, that operates on larger scales, beyond biology and on into the cultural, indeed astronomical, realms, to wit: those technological civilizations anywhere in the Universe that develop in time a planetary society, or global ethics, will survive, and those that do not, will not.

Or, is it in the natural scheme of things that sentient, organic beings are merely way stations on the path to systems of greater complexity? A symbiotic merger of silicon-based systems—machines—with carbon-based systems—humans—has often been cited as the next dramatic expression of increased complexity in the Universe. Perhaps such cyborgs will be among our descendants, or perhaps something else entirely inhuman is our destiny. Does humankind as we know it, in whole or in part, have a future?



Those technological civilizations anywhere in the Universe that develop in time a planetary society, or global ethics, will survive, and those that do not, will not.

Or—it must be considered—do complex, technical systems like ourselves naturally self-destruct? At roughly the same time as any technological society develops, its citizens and its nation-states gain the ability both to discover Nature as well as to destroy themselves, ironically with many of the same tools. Simultaneous with the invention of the science and technology used to explore our origins, today's civilization is confronted with the means of mass destruction, indeed mass extinction. Many of the problems now threatening our species' longevity—from the foremost issues of overpopulation and nuclear warfare to genetic degeneration and environmental pollution, among a host of other ills—are unlike any encountered by our ancestors. For ours are now *global* problems, the likes of which will confront us forevermore. Although we can envision the Life Era as a phase in spacetime when intelligence could conceivably dominate, it's unclear if we humans have the wisdom to achieve and sustain that lofty plateau in the cosmic hierarchy, given the quickening onslaught of *le monde problematique.* Regrettably, perhaps no one, anywhere in the Universe, makes it into the Life Era.

Ethics, it would seem, now takes center stage in the cosmic-evolutionary story, at least for humankind on planet Earth. Ethics itself—broadly defined as conduct collectively recognized regarding all classes of human actions forming our global culture—may well be the beacon that charts our way into the Life Era. To be sure, ethical evolution will likely become the next great epoch along the arrow of time, potentially enabling our descendants to take another evolutionary leap forward toward a higher-ordered state of complexity. At the least, an awareness of cosmic evolution aims us, indeed like an arrow, toward an Ethical Epoch as a firm prerequisite for our future well-being.

The idea of ethics, and especially its relevance as an instrument of behavior, have been around since the beginning of recorded history. Philosophers of old probably invented it and theologians have warmly embraced it (or maybe it's the other way around), but who among them today speaks for planet Earth? Nor is it likely that the needed ethics will arise from science alone, what with our heavy reliance on technology and our dogmatic determination to probe deeper and farther, beyond the world without end. The notion of a worldly ethic, broadly conceived, is easy to grasp in principle, including a mandate for society to embrace global morality and planetary citizenship as a means to survival. But, in practice, it would seem that only an amalgam of these three powerful institutions will together engender, or if necessary demand, the required ethics—a kind of evolutionary advance possible only when we harmonize the agendas of science, philosophy, and religion, indeed one that broadens still more our cosmic-evolutionary scenario.

Now is the time for scientists to become more eclectic and less specialized—to explore holistic worldviews systemically and synergistically and not merely to undertake the reductionistic science that has been so heavily supported by funding agencies for decades and just as myopically honored by our colleges and universities. The scientific community needs to welcome synthesis as well as specialization, teaching as well as research, dissemination as well as discovery. This is by no means a call to abandon the focused research and development that have been the hallmark of a productive economy for decades, but a recognition that it's now appropriate to widen the span of intellectual effort in science and beyond, indeed to engage the larger philosophical and religious communities in an ambitious attempt to understand truly who we are, whence we came, and where we are headed as wise, ethical human beings.

Evolution, energy, and ethics are the core elements that will guide us along the challenging path toward the Life Era: the first—*evolution*—because a good understanding of our universal roots and of our place in the cosmic scheme of things will help us create a feasible future course; the second—*energy*—because our fate will bear strongly on the ways that humankind learns to use energy efficiently and safely; and the third—*ethics*—because global citizenship and a planetary society are crucial factors in the survival of our species.

To employ cosmic evolution as an intellectual as well as practical map to the Life Era is to think in dynamic rather than static terms, to forge a link between natural science and human history, to realize the evolutionary roots of human values, to renew a sense of hope.

∞

We have journeyed far and wide in our cosmic-evolutionary narrative. Life now contemplates life. It probes matter and energy. It seeks to know deep history. It explores the planetary system we call home. It searches for alien life. It quests for new understanding, for meaning, for a raison d'etre.

Told herein has been a surprisingly integrated universal history, or modern *weltgeschichte*, that people of all cultures can understand and embrace—a big-bang-to-humankind story about the awe and majesty of twirling galaxies and shining stars, of buzzing bees and redwood trees, of a Universe that has come to know itself. But it's also a story about our human selves—our origin, our existence, and perhaps our destiny.

In the process, sentient life discovers a meaning, a relevance to cosmic evolution, an underlying motive for universal change. Life comes to recognize that countless billions of stars were born and have died to create the matter now composing our world. We ourselves are made of matter forged in the hearts of generations of stars, annealed in the crucible of billions of years of evolution—a kind of cosmic reincarnation driven solely by an expanding Universe. We have moreover, with our star-stuff brains, become smart enough to ponder the material contents and myriad changes that gave us life. And what we find, quite literally, is that we are more than products of the Universe, more than life *in* the cosmos. We are agents *of* the Universe—animated, cultural instruments commissioned by the Universe to study itself.

Depressing? Frightening? Absolutely not. Cosmic evolution is a wonderfully warm and inviting synthesis, imploring us to embrace our cosmic heritage, to make fuller use of our human potential, and to explore yet more to further unlock the secrets of Nature for the enlightenment of our ordered selves at home in a vastly disordered Universe.

Provided civilizations continue to seek new knowledge, provided they are wise enough to survive, provided above all they remain intellectually curious, then it's not inconceivable that life could someday evolve sufficiently to dominate matter, just as matter overthrew radiation in the early Universe. Indeed, the destiny of part of the Universe may well be determined not only by matter and energy but also by the life that arises from them. Together with our galactic neighbors, should there be any, we may eventually gain control of the resources of much of the Universe, redesign it to suit our purposes, and, in effect, ensure for our being a sense of immortality.

As we enter the new millennium, the integrated, coherent story of cosmic evolution—a powerful and noble epic—can act as a viable intellectual vehicle to involve all our citizens as participants, not just spectators, in the building of a whole new legacy. Perhaps we are indeed becoming wise, ethical, humane human beings. Perhaps we are now on the path toward ethical evolution, arguably part of a cosmological imperative to help us address the many varied challenges along the future arrow of time.

"We are brothers of the boulders, cousins of the clouds."

—Harlow Shapley, an American astronomer of the twentieth century

Further Reading

Prologue

Eric Chaisson, *Relatively Speaking*, W. W. Norton, 1988.
Steven Hawking, *A Brief History of Time*, Bantam Books, 1988.
Martin Rees, *Our Cosmic Habitat*, Princeton Univeristy Press, 2001.

Particle Epoch

Brian Greene, *Elegant Universe,* W. W. Norton, 1999.
Alan Guth, *The Inflationary Universe*, Perseus Press, 1997.
Steven Weinberg, *Dreams of a Final Theory,* Vintage, 1994.

Galactic Epoch

Debra Elmegreen, *Galaxies and Galactic Structure*, Prentice Hall, 1997.
Robert Kirshner, *Extravagant Universe*, Princeton University Press, 2002.
Joseph Silk, *The Big Bang*, W. H. Freeman, 2001.

Stellar Epoch

Eric Chaisson and Steve McMillan, *Astronomy Today*, Prentice Hall, 2005.
Robert Jastrow, *Red Giants and White Dwarfs*, W. W. Norton, 1979.
James Kaler, *Stars*, W. H. Freeman, 1988.

Planetary Epoch

William Broad, *The Universe Below*, Simon & Schuster, 1997.
Richard Fortey, *Earth: An Intimate History*, Knopf, 2004.
William Hartmann, *Moons and Planets*, Wordsworth, 1993.

Chemical Epoch

Christian de Duve, *Vital Dust*, Basic Books, 1995.
Paul Lurquin, *Origins of Life and the Universe*, Columbia University Press, 2003.
William Schopf, *Cradle of Life*, Princeton University Press, 1999.

Biological Epoch

Richard Fortey, *Life: An Unauthorized Biography*, Knopf, 1997.
Andrew Knoll, *Life on a Young Planet*, Princeton University Press, 2003.
Lynn Margulis and Dorian Sagan, *What Is Life?*, University of California Press, 1995.
Ernst Mayr, *What Evolution Is*, Basic Books, 2001.

Cultural Epoch

David Christian, *Maps of Time*, University of California Press, 2004.
Daniel Dennett, *Consciousness Explained*, Little-Brown, 1991.
Jared Diamond, *Guns, Germs, and Steel*, W. W. Norton, 1998.

Epilogue

Eric Chaisson, *The Life Era*, iUniverse.com, 2000.
Hubert Reeves, *Hour of Our Delight*, W. H. Freeman, 1991.
Brian Swimme and Thomas Berry, *The Universe Story*, Harper, 1992.

More technical works on this same subject include:

Eric Chaisson, *Cosmic Evolution*, Harvard University Press, 2001.
Armand Delsemme, *Our Cosmic Origins*, Cambridge University Press, 1998.
Cesare Emiliani, *Planet Earth*, Cambridge University Press, 1992.
David Layzer, *Cosmogenesis*, Oxford University Press, 1990.
Geerat Vermeij, *Nature: An Economic History*, Princeton University Press, 2004.

Glossary

Active galaxy A galaxy that radiates large amounts of energy quite differently in character than that of a normal galaxy.

Adaptation The response to a changing environment of an organism's structure or function in a way that improves its ability to survive and reproduce; any property of an organism that adds to its fitness.

Agriculturist The main job description of our ancestors beginning some 10,000 years ago, namely those who mastered farming.

Amino acid An organic molecule containing carboxyl (COOH) and amino (NH_2) groups, of which twenty different types (plus two rare ones) form the building blocks of the proteins that direct the metabolism in all life-forms on Earth.

Angular momentum The tendency of an object to keep spinning or moving in a circle; rotary inertia.

Anthropic principle The idea that the Universe is the way it is because we (intelligent beings) are here to observe it; the Universe is made for us.

Anthropocentrism The idea that events can be viewed and interpreted in terms of human activities and values.

Anthropomorphism The attribution of human characteristics to other organisms or objects.

Anthropology The study of humanity, including its origin, evolution, development, culture, race, customs, and beliefs.

Antimatter A form of matter having an opposite charge than is normally the case; for example, a positively charged positron is the antimatter opposite of a negatively charged electron.

Archaeology The study of old artifacts that ancient humans left behind.

Arrow of time *In thermodynamics*: the irreversible and inexorable increase in entropy for all natural events. *In cosmology*: the regular and apparent increase in complexity throughout the history of the Universe.

Artifact An object made by humans that has been preserved and can be studied to learn about a particular time period.

Asteroid A small, rocky object revolving around the Sun, sometimes called a minor planet or planetoid.

Asteroid belt A region of space between Mars and Jupiter where the great majority of asteroids are found.

Astrobiology The study of the origin, evolution, and distribution of past and present life in the Universe; also known as bioastronomy or exobiology.

Astronomical unit The average distance between the Earth and the Sun, about eight light-minutes.

Astronomy The study of material events in the Universe beyond Earth's atmosphere.

Astrophysics The study of interactions between matter and radiation in space.

Atmosphere Those gases surrounding the surface of a planet, moon, or star.

Atom A submicroscopic component of matter, composed of positively charged protons and neutral neutrons in the nucleus, surrounded by negatively charged electrons.

Atom period A time in the early Universe when elementary particles began to cluster, thus fashioning the first atoms.

ATP An acronym for adenosine triphosphate, an organic molecule that acts as energy currency in life-forms; the central conveyor of phosphate-bond energy in a cell's metabolism.

Australopithecus The designation given to those prehuman creatures having a mixture of apelike and humanlike qualities, and who lived several million years ago.

Autotroph Any organism capable of self-nourishment by feeding on inorganic matter and external energy, such as plants with the help of sunlight.

Axon The main extrusion of a neuron that acts as a transmitter of information.

Barred-spiral galaxy A type of spiral galaxy having a linear extension or "bar" made of stars and interstellar matter passing through its center.

Baryons Matter made mainly of protons and neutrons; "normal" matter (as opposed to "dark" matter) composing stars, planets, and life-forms.

Big Bang A popular term describing the sudden, expansive start of the Universe.

Binary-star system A pair of stars in orbit about their common center of mass and held together by their mutual gravitational attraction.

Biochemistry The study of chemical processes in living organisms.

Biological evolution The changes experienced by life-forms, from generation to generation, throughout the history of life on Earth.

Biology The study of life in all its forms and phenomena.

Biosphere That part of Earth's crust, water, and atmosphere capable of sustaining life.

Bipedal An adjective meaning "having two feet," "an ability to walk on two legs," or both.

Black hole A region containing a huge amount of mass compacted into an extremely small volume, thus making its pull of gravity so strong that not even light can escape—hence its name.

Brain That part of the central nervous system enclosed in the cranium of humans and other vertebrates, made of a soft, convoluted mass of gray and white matter acting to control and coordinate mental and physical actions.

Carbonaceous chondrite A meteorite having embedded pebble-sized granules that contain significant quantities of organic (carbon-rich) matter.

Catalyst A facilitator or accelerator of a chemical reaction without itself being consumed or changed in the process.

Cell A minimal, usually microscopic, structural unit made of chemicals that can be considered alive.

Central dogma *In biology*: the assertion that information in biological systems passes unilaterally from nucleic acids to proteins, but not conversely. *In physics*: the assertion that energy is conserved in all systems, in all environments, and at all times in the Universe.

Chance A happening without known cause; fortuitous, accidental, contingent, unpredictable.

Change To make different the form, nature, and content of something; the transformation of one system into another that is different in at least one respect.

Chaos In the old sense, unconstrained randomness, disorder; in the new sense, the behavior of a deterministic system under conditions that allow for the possibility of multiple outcomes.

Chemical compound A tightly knit cluster of elements, also sometimes called a molecule.

Chemical evolution The prebiological changes that transformed simple atoms and molecules into the more complex chemicals needed for the origin of life.

Chemistry The study of the properties, compositions, and structures of substances and elements and the ways they interact with one another.

Chemosynthesis The production of organic matter by microorganisms that use chemical energy stored in certain inorganic substances, such as hydrogen sulfide.

Chromosome A threadlike molecule in the nucleus of a cell, consisting of mostly DNA and containing the bulk of the cell's hereditary genes; in humans, twenty-three types vary in length from 50 million to 250 million nucleotide base pairs.

Civilization An advanced stage of development in the arts and sciences accompanied by corresponding social, political, and cultural complexity.

Closed system A system able to exchange energy, but not matter, with its surrounding environment.

Closed Universe A model Universe that stops expanding at some time in the future, after which it contracts to a point much like that from which it began.

Collisional process An event involving a collision of objects; for example, the excitation of a hydrogen atom when hit by another hydrogen atom.

Collision model The idea, now out of favor, that planets form from hot streaming debris torn from a star during a near-collision or close encounter with another passing star.

Comet A small ball of rock and ice from which extends a long wispy tail of gas and dust while nearing the Sun.

Complexity A state of intricacy, complication, variety, or involvement, as in the interconnected parts of a structure—a quality of having many interacting, different components; operationally, a measure of the information needed to describe a system's structure and function or of the rate of energy flowing through a system of given mass.

Condensation model The idea, broadly accepted today, that planets originate from gravitationally contracting and chemically condensing eddies as a natural by-product of the formation of a star.

Consciousness That property of human nature generally, or of the brain specifically, that grants us self-awareness and a sense of wonder.

Conservation of mass and energy A basic principle of science stipulating that the sum of all mass and energy in a closed system remains constant during any event.

Constellation A geometric pattern of bright stars that appear grouped in the sky, named after gods, heroes, animals, and mythological beings by ancient astronomers.

Continental drift The movement of Earth's crust over geological times caused by the plate tectonics within Earth's mantle.

Convection The transfer of heat via circulation, resulting from the upwelling of warm matter and the concurrent downward flow of cool matter to take its place.

Core The central region of a planet, star, or galaxy.

Cosmic abundances A standard listing of the relative numbers of the various elements, determined by studies of the spectral lines in astronomical objects and averaged for many stars in our cosmic neighborhood.

Cosmic background radiation A weak, nearly isotropic electromagnetic (mostly microwave) signal permeating all of space, thought to be a remnant of the big bang.

Cosmic evolution A grand synthesis of all the many varied changes in the assembly and composition of radiation, matter, and life throughout the history of the Universe.

Cosmic-ray particle A charged, subatomic particle of matter (not radiation) that races throughout interstellar space and that regularly strikes Earth's atmosphere.

Cosmological principle The idea (really an assumption) in modern cosmology that the Universe is homogeneous (uniform at every point) and isotropic (uniform in every direction) on scales larger than galaxy superclusters.

Cosmology The study of the structure, evolution, and destiny of the Universe.

Cosmos A complete, orderly, harmonious system; from the Greek, *kosmos*, meaning an orderly whole.

Critical universal density The density of matter above which the Universe is closed (and will collapse back on itself) and below (or equal to) which the Universe is open (and will expand forevermore).

Cultural evolution The changes in the ways, means, actions, and ideas of society, including their transmission from one generation to another.

Culture The totality of activities, artifacts, values, and behavior acquired by members of society through learning.

Cyclic Universe A model Universe that continuously oscillates between expansion and contraction.

Cytoplasm The contents of a cell around its nucleus.

Dark dust cloud A region of interstellar space containing a rich concentration of gas and dust in an irregular but well-defined cloud that obscures the light from stars beyond it.

Dark matter Unseen mass in galaxies and galaxy clusters whose existence is only inferred indirectly but that has not been confirmed directly by any observations.

Darwinism The idea that life-forms evolve by descent with modification among individuals within populations, by means of natural selection of those best adapted to survive environmental changes.

Decoupling An event in the early Universe when atoms first formed, after which photons moved freely in space, causing matter and radiation to behave differently.

Dehydration condensation The linking of two or more amino acids by means of removing water.

Dendrite One of a network of extrusions of a neuron that acts as a receiver of information.

Density A measure of compactness, namely the quantity of something in a unit of volume.

Determinism The idea that all events have specific, definite causes and obey precise, natural laws, making their outcomes completely predictable; from any particular initial state, one and only one sequence of future states is possible.

Development Any process of change, usually of growth or elaboration, between a system's origin and its maturity.

Differentiation The separation of heavy matter from light matter, thus causing a variation in density and composition.

Disorder An irregularity in arrangement or behavior; a synonym for entropy; an absence of order.

DNA An acronym for deoxyribonucleic acid, a self-replicating, double-helical molecule resident chiefly in biological nuclei, mainly responsible for storing hereditary information needed for the building of proteins.

Doppler effect The apparent change in the wavelength (or frequency) of a wave, caused by line-of-sight motion of the source or of the observer (or both).

Dwarf star Any star comparable to or smaller in size than the Sun.

Earth Humankind's home planet in space, third out from the Sun.

Earthquake A sudden dislocation of rocky matter near Earth's surface.

Ecology The study of the interrelatedness among all systems and between those systems and their environment; most common in biology where the systems comprise all living things.

Ecosystem A community of systems and their shared environment, regarded as a unit, all interacting so as to perpetuate the grouping more or less indefinitely; most common in biology, where the systems are plants, animals, and other organisms and the environments are often seafloor, forest, and grassland areas.

Electromagnetic force The force that binds charges of opposite electrical charge and repels charges of identical electrical charge.

Electromagnetism The phenomenon of electricity and magnetism studied together.

Electromagnetic spectrum The entire range of all the various kinds of radiation; light (or the visible spectrum) is just one small segment of this much broader spectrum.

Electron A negatively charged elementary particle that resides outside (but is bound to) the nucleus of an atom.

Element A substance comprising one and only one distinct kind of atom; one impossible to separate into simpler substances by chemical means.

Elementary particle A basic building block of atoms.

Elliptical galaxy A galaxy having a spherical or elliptical shape, some more than others, and composed mostly of old stars and little interstellar matter.

Emergence The appearance of entirely new system properties at higher levels of complexity not preexisting among, nor predictable from knowledge of, lower-level components; the process of a system "becoming" from its environment at certain critical stages in its development or evolution.

Energy The ability to do work or to produce change; an abstract concept invented by nineteenth-century physicists to quantify many different phenomena in Nature.

Energy density A measure of compactness of energy; an amount of energy per given volume.

Energy rate density The amount of energy (available to do work) flowing through a system per unit time and per unit mass.

Entropy A measure of randomness, or disorder, of a system, reaching a maximum state of inert uniformity at thermodynamic equilibrium; a lack of information about a system's organization.

Environment Any part of the Universe not included in a system; a combination of all things, conditions, and influences surrounding a system.

Enzyme Any of numerous complex proteins that catalyze specific biochemical reactions.

Equilibrium A state wherein a system's gradients are negligible, its probability maximized, and its free energy minimized; one that constantly reacquires any and all of its possible configurations randomly, thus from which it exhibits no tendency to depart.

Escape velocity The minimum speed needed for an object to escape the gravitational pull of a massive object.

Euclidean geometry The terrestrially familiar geometry of "flat space" that all of us learn in high school.

Eukaryote A life-form whose cells have well-developed biological nuclei; all organisms above the level of prokaryotes, including protists, fungi, plants, and animals.

Event Any occurrence in spacetime defined by its location and date; a happening.

Event horizon A region within which no event can ever be seen, heard, or known by anyone outside; also termed the "surface" of a black hole.

Evolution Any process of growth and change with time, including an accumulation of historical information; in its broadest sense, both developmental and generational change.

Extraterrestrial An adjective meaning "beyond the Earth."

Fauna The species of animals living in a given geographical area at a given time.

Fermentation The extraction of energy via the capture and chemical breakdown of small molecules.

First law of thermodynamics A principle stipulating that in any real process, energy is conserved, that is, never created or destroyed but allowably changed from one form to another.

First-generation star A star made of primordial matter and thus having no heavy elements.

Fission A nuclear process that releases energy when heavyweight nuclei break down into lightweight nuclei.

Flora The species of plants living in a given geographical area at a given time.

Flow The movement of an entity from one place to another; to issue or proceed from a source.

Force An agent of change in or on any system.

Form The structure, pattern, organization, or essential makeup of anything.

Fossil The hardened remains of a dead organism whose skeletal outlines or bony features are preserved in ancient rocks.

Fragmentation Developing inhomogeneities in the gas density of a cloud that eventually break down into smaller clumps of matter within the cloud.

Function The ability of a system's components, beyond its mere structure, to execute an internal action, role, or job assignment, such as breathing, running, writing, or reproducing.

Fusion A nuclear process that releases energy when lightweight nuclei combine to form heavyweight nuclei.

Galactic center The hub or core of the Milky Way, some 30,000 light-years from the Sun.

Galactic cluster A loose collection of tens to hundreds of relatively young stars spread over several light-years, sometimes termed an open cluster.

Galactic evolution The changes experienced by galaxies, either intrinsically because of localized changes among myriad stars or environmentally because of merges, acquisitions, and close-encounters among neighboring galaxies.

Galactic halo A nearly spherical region that surrounds the Milky Way Galaxy and extends some 50,000 light-years from the galactic center.

Galactic plane A relatively thin disk or plane in which most of the Milky Way's stars and interstellar matter now reside.

Galaxy An open, coherent, spacetime structure maintained far from thermodynamic equilibrium by a flow of energy through it—a colossal system of billions of stars and much loose gas held together by gravity.

Galaxy cluster A group of galaxies held together by their mutual gravitational attraction.

Galaxy period A time in the relatively early history of the Universe when the bulk of most galaxies formed.

Galaxy supercluster A truly huge cluster of galaxy clusters, often stretching over a hundred million light-years or more.

Gaseous nebula A region of ionized gas (plasma) surrounding one or more young, hot stars (sometimes termed an emission nebula).

Gene A segment of any DNA molecule containing information for the construction of proteins, hence responsible for directing inheritance from generation to generation; in humans, approximately 30,000 such genes vary in size from 3000 to 2.4 million nucleotide bases.

Gene pool The spread or distribution of the variations or traits among a given population of a species; all the genes in a given population.

Genetic code An encyclopedic blueprint of the physical and chemical properties of all of an organism's cells and all of its functions.

Genetics The study of heredity and the biological processes by which inherited characteristics are passed from one generation to the next.

Genome The sum of all genes carried by a single organism; humans, for example, have nearly 30,000 genes (comprising about 3 billion nucleotide bases), bacteria, a few hundred genes (including about a million bases).

Genotype The sum of all genes of an individual.

Geocentric An adjective meaning "centered on the Earth."

Geography The study of positions, shapes, sizes, and numerous other qualities of Earth's continents.

Geology The study of the physical history of Earth, especially the rocks of which it is composed and the events it has undergone.

Geometry The study of the size, shape, and scale of things.

Giant star Any star much larger in size than the Sun.

Globular cluster A tight-knit collection of many thousands, sometimes even millions, of old stars spread throughout the halos of galaxies.

Grand-unified theory An idea that three forces—the electromagnetic force, the strong nuclear force, and the weak force—are different manifestations of one and the same force.

Gravitational force The force that holds matter together on a large scale, such as stars within galaxies, atoms within stars, and people on Earth.

Gravitational instability A condition whereby an object's (inward-pulling) gravitational potential energy exceeds its (outward-pushing) thermal energy, thus causing the object to infall.

Gravity An attractive force that any massive object exerts on all other massive objects.

Gravity wave The gravitational analog of an electromagnetic wave whereby gravitational radiation is emitted at the speed of light from any mass that undergoes rapid acceleration.

Greenhouse effect The trapping of radiation by an atmosphere (or greenhouse), thus causing greater heating than would normally be the case.

Habitable zone A three-dimensional region of "comfortable" temperatures that surrounds every star.

Hadron period A very early time in the history of the Universe when heavy, strongly interacting, elementary particles, such as protons and neutrons, were the most abundant type of matter.

Heat The amount of energy transferred to or from a substance; the thermodynamic state of an object by virtue of the random motions of the particles within it.

Heliocentric An adjective meaning "centered on the Sun."

Heredity The transmission of genetic traits from parents to offspring, thus ensuring the preservation of certain characteristics among future generations of a species.

Heterotroph Any organism requiring organic matter for food, such as primitive cells that survived by absorbing acids and bases floating on primordial seas or most animals today.

Hierarchical clustering The idea that large objects are built from small objects, including, for example, galaxies having originated partly by collecting already-made star clusters.

Holism The idea that a whole entity, as a basic component of reality, has an existence greater than the mere sum of its parts.

Hominid Both our erect-walking human ancestors (present and extinct *Homo*) and their predecessor near-relatives (bipedal austrolopithecines) arising after the split from the gorilla-chimp lineage roughly six million years ago.

Homo erectus The species designation given to all human creatures who lived from roughly 200,000 to 1 million years ago; literally, the Latin means "erect man."

Homo sapiens The species designation given to all human creatures who lived during about the past 200,000 years, including ourselves; literally, the Latin means "wise man."

Hubble's constant The proportionality factor between the distance of a galaxy and the velocity with which it recedes; currently, its best estimate is 20 kilometers/second/million light-years.

Hubble's law An empirical finding linking the distance of a galaxy and the velocity with which it recedes.

Humankind The human race on Earth, considered collectively.

Hunter-gatherer The main job description of our ancestors during most of the past few million years, namely those who survived by hunting and gathering food.

Hydrosphere The liquid part of a planet's surface, including any lakes, streams, oceans, and rivers.

Ice age A period of cool, dry climate that intermittently plagues planets, causing, in the case of Earth, a long-term buildup of glacial ice far from the poles.

Inflation A period of extremely rapid expansion of the Universe very shortly after the beginning of all things.

Information The number of bits needed to specify a message or structure; the difference between the maximum possible and actual entropies of any given system.

Intelligence The capacity to comprehend relationships and address multiple tasks simultaneously; a biological adaptation for complex behavior, probably synonymous with language.

Intergalactic space Regions outside galaxies and especially galaxy clusters where matter has never been conclusively found.

Interplanetary matter Debris in the great spaces between the planets of the Solar System.

Interplanetary space Regions among the planets, moons, and related objects of the Solar System.

Interstellar matter Sparse gas and dust in the vast domains among the stars.

Interstellar space Dark regions among the stars of any galaxy.

Invertebrate An organism without a backbone.

Invisible radiation Those kinds of radiation to which the human eye is not sensitive, including radio, infrared, and ultraviolet waves, as well as X rays and gamma rays.

Ion An atom with one or more electrons removed (or added), giving it a positive (or negative) charge.

Ionosphere A radio-reflecting region high in Earth's atmosphere in which solar radiation has removed an electron from some of the atoms.

Irregular galaxy A strangely shaped galaxy, often rich in interstellar matter but apparently not a member of any of the major classes of spiral or elliptical galaxies.

Isotope An atom having more or fewer neutrons than normal.

Jovian planets The four, big, gassy planets in the outer parts of the Solar System: Jupiter, Saturn, Uranus, and Neptune.

Kelvin scale An international temperature scale, equal to the Celsius (or Centigrade) scale plus 273 degrees.

Kepler's laws Three principles, discovered empirically by a seventeenth-century German astronomer, that describe the motions of the planets in their orbits about the Sun.

Kinetic energy The energy of an object or system due to its mass and motion; the ability to do work actively via motion.

Kin selection In individuals related by common descent, such as siblings, altruistic selection for the shared parts of their genotypes.

Lamarckism The idea that an organism is a result of environmental influences rather than genetic inheritance; traits can be acquired through habit, use, or disuse during a single lifetime and then passed on intact to the next generation.

Lepton period A very early time in the history of the Universe when the lightweight, weakly interacting, elementary particles, such as electrons and neutrinos, were the most abundant type of matter.

Life An open, coherent, spacetime structure kept far from thermodynamic equilibrium by a flow of energy through it—a carbon-based system operating in a water-based medium, with higher forms metabolizing oxygen.

Life Era An advanced period in the history of the Universe when technological life-forms manipulate their genes and their environments more than conversely.

Light The kind of radiation to which the human eye is sensitive.

Light-year The distance traveled by light in a full year; equal to some 10 trillion kilometers.

Linear momentum The tendency of an object to keep moving in a straight line.

Lithosphere The solid part of a planet's surface, including any continents and seafloor.

Local Group The specific name given to the galaxy cluster that includes the Milky Way Galaxy as a member.

Luminosity The rate of electromagnetic energy released from any object, sometimes called the absolute brightness.

Magnetism An attractive or repulsive influence that a magnet exerts on another magnet (or on a charged particle).

Magnetosphere A region of space, usually high above a planet's atmosphere, where charged particles are magnetically deflected or trapped.

Mantle The interior of a planet, namely that matter below the crust yet above the core.

Mass A measure of the total amount of physical substance, or "stuff," contained within an object.

Mass extinction The destruction of a large part of the biosphere of Earth by means of climatic, geologic, cosmic, or other environmental events.

Materialism The idea that there is nothing other than matter, energy, and their various arrangements, motions, and changes in the Universe.

Matter Anything that has mass and occupies space.

Matter Era A mature period in the history of the Universe when the density of energy contained within matter exceeds the density of energy contained within radiation.

Matter-antimatter annihilation A highly efficient process in which equal amounts of matter and antimatter collide and destroy each other, thus producing a burst of energy.

Meiosis The division of a cell nucleus, whereby the chromosomes do not divide but are equally shared by the two new (sex) cells.

Meme In analogy with gene, a cultural replicator—an idea, behavior, style, or usage that spreads from person to person within a culture.

Metabolism The sum of all chemical reactions that energetically support a living organism, starting from energy sources that are either chemical (environmental nutrients) or physical (sunlight).

Meteor A heated, glowing object streaking through Earth's atmosphere but not yet having hit the surface.

Meteorite A meteoroid that manages to survive passage through an atmosphere to collide ultimately with the surface of a planet or moon.

Meteoroid On average, a meter-sized boulder that has probably escaped from the asteroid belt and thus roams the Solar System.

Microbe A unicellular microorganism, or bacterium, distinguishable from plant or animal.

Milky Way Humankind's home galaxy, comprising some hundred billion stars of which the Sun is one and so named because its stars resemble a milky band running across the dark night sky.

Mitochondria An organelle of a eukaryotic cell that transforms nutrients into energy, generally by oxidation; thought to be an ancient bacterium, originally captured by symbiosis as a parasite by an infected cell.

Mitosis The division of a cell nucleus into two equal parts, in which all the chromosomes divide equally, yielding two cells identical to the first.

Modern synthesis A conceptual unity in contemporary biology, based on Darwinian evolution, including natural selection, adaptation, diversity, and Mendelian genetics; also termed neo-Darwinism.

Molecular clock The regularity in the change of a gene or a whole genotype over geological time.

Molecular cloud A relatively dense, cold region of interstellar matter where molecules are abundant.

Molecule A bound cluster of two or more atoms held together by electromagnetic forces.

Multicell A group of cells that collaborate with other cells.

Mutation A microscopic change in the DNA base sequence of any gene, transmissible by replication.

Natural selection *In general*: a normative process whereby environmental resistance tends to eliminate nonrandomly those members of a group of systems least well adapted to cope and thus, in effect, choose or "select" those best suited for survival. *In biology*: the Darwinian process whereby a population's life-forms having advantageous traits are able to adapt to a changing environment, thereby surviving, reproducing, and passing on to their descendants those favorable traits which then accumulate in the population over time.

Nature The Universe, including all its natural phenomena.

Nebular model The idea that the Solar System originated in a contracting, swirling cloud of gas that left behind a concentric series of rings from which the planets formed.

Necessity The inevitable "force" of circumstances.

Neo-Darwinism A combination of traditional Darwinian evolution and Mendelian genetics; also termed "the modern synthesis."

Neuron A biological, or nerve, cell in a brain.

Neutrino A neutral, weakly interacting elementary particle having almost no mass.

Neutron A neutral elementary particle having slightly more mass than a proton and that resides in the nucleus of most atoms.

Neutron star An extremely compact ball of neutrons having the mass of a star but a size smaller than a planet.

Non-equilibrium A state characterized by nonnegligible gradients and a regular energy flow, allowing for further change, growth, and evolution.

Non-thermal radiation Radiation released by virtue of a fast-moving charged particle (such as an electron) interacting with a magnetic force field; heat has no role in this process.

Normal galaxy A galaxy that radiates energy much as expected from a large accumulation of stars.

Nova An unstable star that rapidly brightens while expelling a small fraction of its matter after which it slowly fades back to normal.

Nuclear force The force that binds atomic nuclei.

Nuclear transformation Changes in atomic nuclei owing to the reaction of one nucleus with another nucleus.

Nucleic acid A class of long-chain, organic molecules, made by grouping many nucleotides with sugars and often inhabiting the biological nuclei of cells.

Nucleotide base An organic molecule, of which five different types constitute the building blocks of all nucleic acids within genes that transmit hereditary characteristics from one generation of life-forms to the next.

Nucleus *In physics*: the positively charged core of an atom where nearly all of its mass resides and that comprises protons and (except for in hydrogen) neutrons, around which electrons orbit. *In biology:* the inner, central part of a eukaryotic cell, containing the genetic recipe (DNA) for making similar cells.

Oceanography The study of the ocean's motion, history, and physical and chemical behavior.

Ontogeny The developmental history of an individual system; in biology, the developmental life cycle of an individual organism.

Open system A system able to exchange both energy and matter with its surrounding environment.

Open Universe A model Universe that expands forever.

Order A regularity in arrangement or behavior; a restriction on the number of possible states; an absence of disorder.

Organism Anything that lives—plant, animal, or microbe—or has ever been living.

Organization Relations existing among the components of a system for it to be a member of a specific class.

Origin A coming into being; a process whereby a given state precedes all other such states in time.

Ozone layer A layer in Earth's atmosphere rich in triatomic oxygen (O_3) molecules, which block high-frequency radiation.

Paleoanthropology The study of prehistoric humanity.

Paleomagnetism The study of ancient magnetism.

Paleontology The study of the fossilized remains of dead organisms.

Panspermia An idea stipulating that germs are everywhere in the Universe and that primitive life on Earth originated when some of these germs came to our planet from outer space.

Particulate evolution The changes among elementary particles, including photons, in the early Universe.

Periodic table of the elements A systematic listing, according to increasing mass, of all the known kinds of atoms.

Phenotype The sum of all traits of an individual (anatomical, physiological, behavioral) resulting from the interaction of its genes with the environment.

Photon A packet of pure quantum energy; the massless, chargeless carrier of electromagnetic radiation and mediator of the electromagnetic force.

Photosphere The visible portion of the Sun or any star, although this "surface" is made exclusively of hot, opaque gas.

Photosynthesis The production of organic matter by (usually) green plants that use sunlight to make glucose from carbon dioxide and water, the byproduct being oxygen.

Phylogeny The evolutionary history of a group of systems; in biology, the evolution of ancestral relations among species, often illustrated by a "tree of life" branching diagram where organisms are connected by the number of mutations separating them.

Physics The study of matter, energy, space, and time.

Planet An open, coherent, spacetime structure maintained far from thermodynamic equilibrium by a flow of energy through it—a rocky and/or gaseous system, more massive than an asteroid yet less massive than the star about which it orbits.

Planetary evolution The changes in the physical or chemical properties of planets during the course of their histories.

Planetary nebula A twofold object comprising an old yet hot white-dwarf star surrounded by a thin, ionized, spherical shell of expanding gas.

Planetesimal An asteroid-sized blob of matter that gradually collided with others in the formative stages of the Solar System, thus fabricating the planets.

Plasma A state of matter wherein all atoms are ionized; a mixture of free electrons and free atomic nuclei, often called the "fourth state of matter," after solids, liquids, and gases.

Platonism The idea that the changing, shifting world of physical phenomena masks a deeper reality—an underlying set of eternal ideas and unchanging forms, and it is these alone that grant true knowledge.

Positron A positively charged antiparticle of the electron.

Potential energy The energy of an object or system due to its mass and position; the ability to do work passively stored.

Power The rate at which work is done; an amount of energy transferred per unit time.

Primate The order of mammals that includes monkeys, apes, and humans.

Primordial nucleosynthesis Element building that occurred in the early Universe when the nuclei of primordial matter collided and fused with one another.

Principle of equivalence The idea that the pull of gravity on an object and the acceleration of that object (i.e., its gravitational and inertial forces) can be viewed as conceptually equivalent.

Probability The likelihood of an event, expressed by the ratio of the number of actual occurrences to that of possible occurrences.

Process The change of a quantity over time; the act of proceeding.

Prokaryote A life-form whose single cell lacks a well-developed biological nucleus, such as various types of bacterial microorganisms.

Protein A class of long, folded organic molecules, made of typically hundreds of amino acids and inhabiting the cytoplasm of cells; a major structural component as well as a functional enzyme in both plants and animals.

Proteinoid microsphere A microscopic proteinlike cluster rich in amino acids, artificially produced in the laboratory.

Proteome The sum of all proteins in a cell.

Protist A general name for any of the vast variety of unicellular eukaryotes, bigger and more complex than any prokaryote.

Protogalaxy A forerunner of a present-day galaxy, also sometimes termed a "baby galaxy."

Proton A positively charged elementary particle that resides in the nucleus of every atom.

Proton-proton cycle A series of nuclear events whereby hydrogen nuclei (protons) are converted into helium nuclei, releasing energy in the process.

Protoplanet A forerunner or progenitor of a genuine planet.

Protostar An embryonic condensation of interstellar matter perched at the dawn of star birth.

Pulsar A compact, celestial object that emits rapid and periodic pulses of radiation and that is thought to be a rotating neutron star.

Punctuated equilibrium The idea that life's species remain essentially unchanged for long periods of time, after which they change rapidly in response to sudden, drastic changes in the environment.

Quantum physics A branch of physics dealing with submicroscopic parts of a system, including its inherent uncertainty.

Quark A fractionally charged, basic building block of protons, neutrons, and many other elementary particles.

Quasar An acronym for quasi-stellar source; a high-red-shift object whose image resembles a star but whose energy budget seems comparable to or larger than that of a normal galaxy.

Radiation *In physics*: a form of energy that travels at the velocity of light, of which light itself is a special kind. *In biology*: divergence of members of a single evolutionary line into different niches.

Radiation Era An early period in the history of the Universe when the density of energy contained within radiation exceeded the density of energy contained within matter.

Radioactivity The spontaneous decay of certain rare, unstable, heavyweight nuclei into more stable lightweight nuclei, a natural by-product of which is the release of energy.

Red-giant star An old, bright star, much larger in size and cooler than the Sun.

Red shift The Doppler lengthening of the wavelength of radiation (or the shifting of spectral lines toward longer wavelengths) caused by some net motion of recession.

Reductionism The idea asserting that all natural phenomena can be understood only by reducing them to their smallest component parts.

Relativistic physics A branch of physics dealing with matter moving at high speeds (special theory) or in strong gravitational fields (general theory); the worldview according to Einstein.

Relativity A theory of physics that describes the dynamical behavior of matter and energy under peculiar circumstances, especially at very high velocities and very high densities.

Reproduction The natural process among life-forms by which new individuals are generated; a copy, duplicate.

Respiration A chemical process whereby cells use oxygen to release energy; a technical term for "breathing."

Revolution The orbital motion of one object about another.

RNA An acronym for ribonucleic acid, a single-stranded organic helix found chiefly in the cytoplasm of cells, often instrumental in protein synthesis.

Rotation The spin of an object about its own axis.

Scientific method An investigative technique used by all natural scientists throughout the world. In general, some data or ideas are first gathered, then a theory is proposed to explain them, and finally an experiment is devised to test the theory.

Second law of thermodynamics A principle stipulating that in any real process, the entropy of the Universe increases, that is, irreversibly tends toward greater disorder; energy naturally flows from hotter to colder systems, and not the reverse.

Secondary atmosphere Gases that a planet exhales from its interior after having lost its primary or primordial atmosphere.

Second-generation star A star having some heavy elements and that is thus made of matter that has been previously processed through other stars.

Selection A process of Nature that causes some systems having certain properties, which are not the norm for a population of systems, to preferentially adapt to their environment and thus to enhance their state; those things that work well survive, and those that don't, don't.

Shock wave A rapidly rushing shell of gas that tends to push aside and sometimes implode matter in its wake.

Simplicity A state free of complexity or of the possibility of confusion.

Singularity *In physics*: a superhot, superdense state of matter, where the known laws of physics are likely to break down. *In mathematics*: a point where a mathematical function ceases to be defined, usually because it becomes infinite.

Sociobiology The systematic study of the biological basis of social behavior in humans as well as in other life-forms.

Solar core The region immediately surrounding the center of the Sun where nuclear reactions release vast quantities of energy.

Solar System Humankind's home planetary system, comprising nine planets, dozens of moons, and countless smaller asteroids and meteoroids, all orbiting the Sun.

Space An indefinitely great three-dimensional expanse in which all material objects are located and all events occur.

Spacetime A synthesis of the three dimensions of space and of a fourth dimension, time; a hallmark of relativity theory.

Speciation The change of a single species into two or more new species; also termed "disruptive selection."

Species Any organism—plant, animal, or microbe—of a single kind; a fundamental biological classification denoting a group of individuals not only structurally similar but also able to mate among themselves and produce fertile offspring.

Spectroscopy An observational technique designed to disperse radiation into its component wavelengths in order to study in fine detail the way that matter emits or absorbs radiation.

Spiral arm Part of a pinwheel structure of young stars and interstellar clouds usually winding out from a galaxy's center.

Spiral galaxy A galaxy having a pinwheel shape, some more than others, and composed of a mixture of old and young stars as well as loose interstellar matter.

Spontaneous generation The theory, now refuted, that life-forms have suddenly emerged fully developed from peculiar arrangements of nonliving matter.

Standard model *In physics*: an acknowledged description of microscopic phenomena, bolstered by accelerator experiments and the quantum field theory of particles and forces. *In cosmology*: an acknowledged description of macroscopic phenomena, bolstered by observations of galaxy recession, background radiation, and elemental abundances.

Star An open, coherent, spacetime structure maintained far from thermodynamic equilibrium by a flow of energy through it—a glowing ball of gas held together by its own gravity and powered by nuclear fusion at its center.

State The status or condition of a system as specified by certain dynamical variables.

Statistical fluctuation Continual change from one course or condition to another; an irregularity or instability.

Statistical physics A branch of physics dealing with vast numbers of particles and their probable states, enabling some averaging of properties within macroscopic systems.

Stellar evolution The changes experienced by stars as they originate, mature, and terminate.

Stellar nucleosynthesis Element building that occurs in stars when nuclei collide and fuse with one another.

Stellar period A time in the history of the Universe (including now) when the stars form.

Structure The arrangement of the basic components of a system, including form but not function.

Sun Humankind's parent star, a resident of the Milky Way.

Supernova An explosive death of a massive star whose glowing debris produce for a few weeks a great brightening comparable to a whole galaxy.

Supernova remnant The remains of a supernova, namely glowing debris scattered over a light-year or more.

Symbiosis The living together, in a mutually beneficial union that aids survival or evolution, of two organisms of different species.

Symmetry The ordered repetition of identical parts of a structure or state.

Synapse A microscopic gap separating an axon of one neuron from a dendrite of another neuron.

System A finite assemblage of interdependent things in the Universe, separated from its surrounding environment by topological and organizational boundaries; any entity of interest, usually one having interconnected components acting as a unitary whole.

Tectonics The study of crustal displacements and deformations of large continental plates on a planet's surface; for Earth, popularly termed "continental drift."

Temperature A measure of the heat of an object, by virtue of the random motions of the particles within it.

Terrestrial An adjective meaning "of the Earth."

Terrestrial planets The four small, rocky planets in the inner parts of the Solar System: Mercury, Venus, Earth, and Mars.

Thermal energy The energy of an object or system due to its heat, a measure of which is temperature.

Thermal radiation Radiation released by virtue of an object's heat; namely, by charged particles interacting with other charged particles.

Thermodynamics The study of the macroscopic changes in the energy of a system, for which temperature is a central property and meaning literally "movement of heat."

Time The fourth dimension that distinguishes past, present, and future; a quantity easily measured yet hard to define.

Turbulence The disordered, irregular motion of matter, so complex as to defy description except in a statistical manner.

Unicell A single cell that does not collaborate with other cells.

Universe The totality of all known or supposed objects and phenomena, formerly existing, now present, or to come, taken as a whole.

Van Allen belts Zones of intense radiation surrounding Earth's midsection, caused by charged particles trapped in Earth's magnetic force field.

Velocity of light The fastest speed that any object can move, approximately 300,000 kilometers per second.

Vertebrate An organism having a backbone, including fishes, amphibians, reptiles, and mammals.

Virus The smallest and simplest entity that sometimes appears to be alive.

Visible spectrum The narrow range of wavelengths in the electromagnetic spectrum to which the human eye is sensitive; namely, light.

Volcano The site of hot lava upwelling from below the crust of a planet or moon.

Wavelength The distance between successive crests of a wave.
Weak force The force that governs the change of one kind of elementary particle into another.
White-dwarf star An old, dim star, much smaller in size and hotter than the Sun.

Index

About the Author

Eric J. Chaisson has published more than a hundred scientific articles, most of them in professional journals. He has also written ten books, including *Astronomy Today* (with S. M. McMillan), the nation's most widely used college astronomy textbook, and *Cosmic Dawn*, which won the Phi Beta Kappa Prize, the American Institute of Physics Award, and a National Book Award nomination (finalist) for distinguished science writing. Another recent book, *The Hubble Wars*, also won the Science Writing Award of the American Institute of Physics. His most recent book, *Cosmic Evolution: The Rise of Complexity in Nature*, was published in 2001 by Harvard University Press.

Trained initially in atomic physics, Chaisson received his doctorate in astrophysics from Harvard University, where he spent a decade as a member of Harvard's Faculty of Arts and Sciences. While at the Harvard-Smithsonian Center for Astrophysics, he won fellowships from the National Academy of Sciences and the Sloan Foundation as well as Harvard's Bok Prize (for scientific research) and Harvard's Smith-Weld Prize (for literary merit). Thereafter, he spent several years as a senior scientist and division head at the Space Telescope Science Institute at Johns Hopkins University, where he was also adjunct professor of physics and associate director of the Maryland Space Grant Consortium.

Dr. Chaisson currently holds research professorships in the departments of physics, astronomy, and education at Tufts University, where he directs the Wright Center for Innovative Science Education; he also teaches a course on cosmic evolution at Harvard University and is affiliate director of the Massachusetts Space Grant Consortium at MIT. He lives with his wife and three children in historic Concord, Massachusetts.