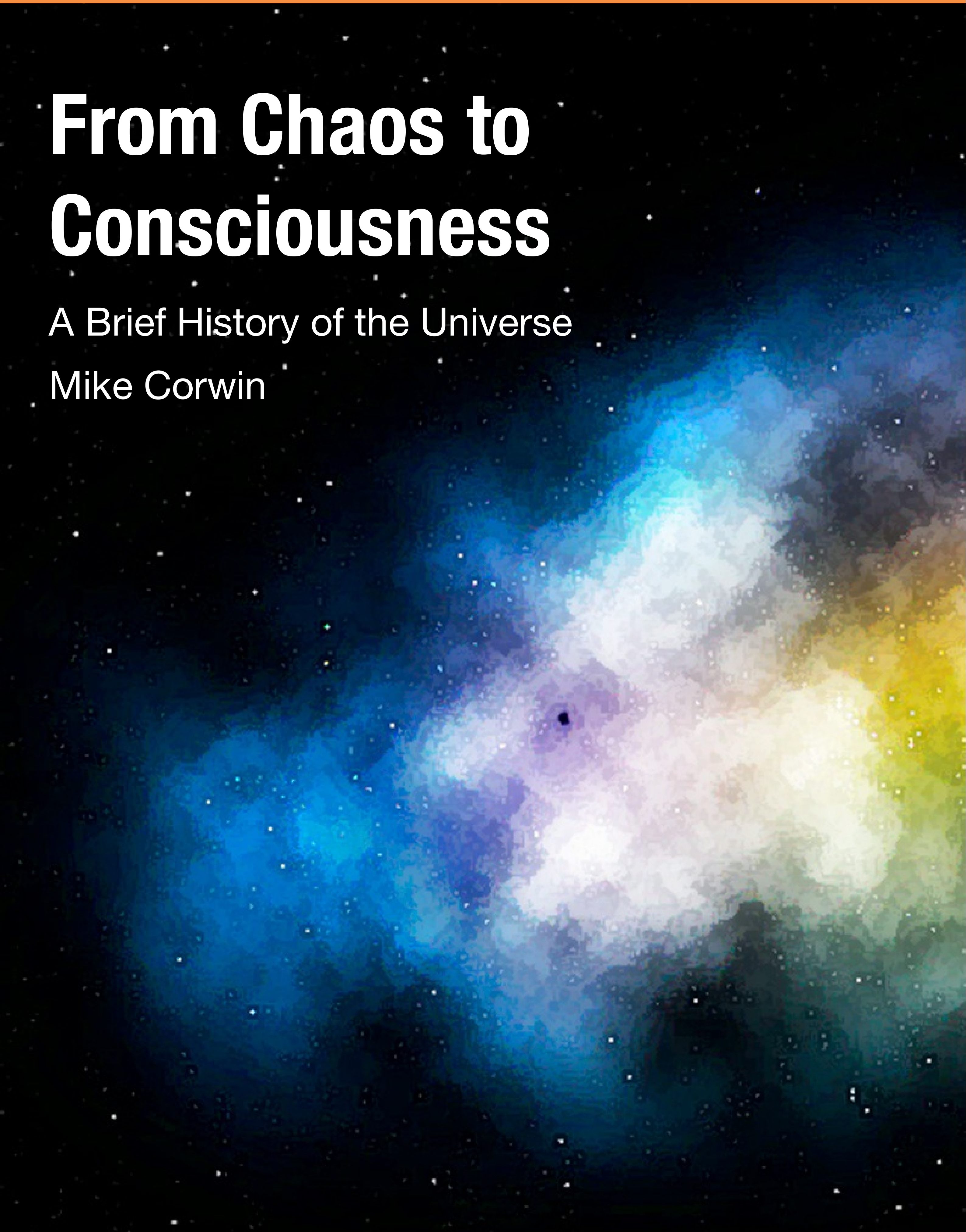
From Chaos to Consciousness

A Brief History of the Universe

Mike Corwin



bookboon
The eBook company

Mike Corwin

From Chaos to Consciousness

A Brief History of the Universe

From Chaos to Consciousness: A Brief History of the Universe
1st edition

ISBN 978-87-403-0937-9

Contents

1 The Greeks

Your goals for this chapter are to know the following:

- What is the most significant factor that distinguishes Greek thought from earlier ways of thinking about the universe? Give some examples.
- Briefly describe the ancient (Greek) world-view. Include the views of Thales, Democritus, Plato, Aristotle, and Ptolemy.
- What were some of the most significant discoveries and understandings about the Universe that were made by the Greeks?
- What most differentiates the philosophy of Socrates from that of the pre-socratics? What most differentiates the philosophy of Aristotle from that of Plato?
- What is the meaning of the Greek word "cosmos?"
- What was the role of Islamic culture in the events between the ancient world view and the Scientific Revolution?

1.1 Introduction

The evolutionary origin of consciousness is not well understood. But it is clear that at some point in our evolutionary past, an inner world free from passive captivity in sensory impressions came to be – a life of the mind, if you like. The task of this inner world was to impose order on the content of our senses; that is, there came a point in our evolutionary past when our minds began to demand that the universe make sense.

Between 250,000 and 30,000 years ago, *Homo neanderthalensis*, commonly known as Neanderthals, inhabited a territory from England across southern Europe into Asia. They built hut-like structures, hunted big game, and developed ceremonial rituals. The remains of this relatively short-lived group reflect a consciousness akin to our own. They clearly had a sense of the passage of time and of physical mortality. Ceremonial burials show a concern with life after death. We find in the Neanderthals the signs of a greater consciousness of the subtleties of the world and, in this greater consciousness, the roots of our own religious and scientific thinking.

The species to which all living humans belong, *Homo sapiens*, probably emerged somewhere around 100,000 years ago. We really don't know much concerning the specific worldviews of early humans. We have only some burial sites, figurines, and cave paintings to provide tantalizing hints. These suggest that their thinking was of a type that would now be called magical or superstitious. It seems clear that these stone-age people sought to employ magical rituals to influence the external world, hoping to positively affect hunting, fertility, and other survival-related aspects of their lives. Though such thinking may seem primitive to us, it clearly reflects a mind attempting to bring order into the universe.

Many scholars believe that the magical, animistic approach to understanding the world led directly to the mythical approach. Repeated attempts at magic would probably give birth to ritual. Even when such a ritual outlived its original purpose or when its true meaning had become obscured over time, it might still remain a psychologically essential part of the culture. An appropriate myth could provide justification for its continuance. In any case, with the coming of civilization some 10,000 years ago, it is clear that the magical thinking of the earliest peoples had already evolved into mythology.

Myths, though differing in their local details, have some common threads running through them. Often powerful non-human, but anthropomorphic, figures create and control the world and its inhabitants. Myth is always related to creation; it tells how something came into existence (the universe, human beings), or how a pattern of behavior was established. This "History" is considered to be absolutely true and sacred.

Surely myth arose from our need to make sense of the world as a whole, and, particularly, of our place as human beings in it. We see in myth attempts to find cause and effect explanations for the experienced world. Early people wove basic sensory knowledge of the world into a pattern that seems reasonable. For instance, the Mesopotamian creation myth used their knowledge of how silt deposits form land where fresh and salt water meet. Thus, although there are some obvious differences between the mytho-poetic approach and the scientific approach, we can also see connections. Myths are the first rungs on the ladder of discovery. Embedded within them are basic truths about both the universe and the human condition.

1.2 The Birth of Science

The next step in our understanding of the universe was taken in ancient Greece. Although it is probably an exaggeration to think in terms of "the Greek miracle" or of "motherless Athena," as is frequently done, it is clear that about 600 BCE, a new approach to understanding the universe emerged. Although the Greeks had their myths, they went beyond the myths to search for physical explanations. Unlike earlier cultures, they were not content to explain the universe in terms of the actions of the gods; the Greeks insisted on thinking in terms of natural processes. This attitude is exemplified in the statement of a writer belonging to the Hippocratic school on the nature of epilepsy.

> "It seems to me that the disease is no more 'divine' than any other. It has a natural cause, just as other diseases have. Men think it divine merely because they do not understand it. But if they called everything divine which they do not understand, why, there would be no end of divine things!"

These proto-scientists made the remarkable assumption that an underlying rational unity and order existed within the flux and variety of the world. Nature was to be explained in terms of nature itself, not of something fundamentally beyond nature, and in impersonal terms rather than by means of personal gods and goddesses. Science was born here, not motherless, to be sure, but nonetheless a new and distinctly different way of looking at the world.

Map 1. The Greek World about 450 B.C.

Figure 1: Map of Greece.

Thales (624–547 BCE) was born in the Greek city of Miletus across the Aegean Sea from the Greek mainland. The inhabitants of this region were known as Ionians (Greeks who fled the Dorian invasion). Its location on the coast of Asia Minor provided Thales with exposure to the cultures of both the Babylonians and the Egyptians, and in fact, he visited both regions. It was his knowledge of Babylonian astronomy that gave rise to the story, probably apocryphal, that he predicted the solar eclipse of May 28, 585 BCE.

We consider Thales the first scientist because, as far as we can tell from the admittedly incomplete historical record, he was the first to approach the world from a scientific perspective. He wondered how the universe came to be and came up with an answer far different from that depicted in the creation of the gods myth of Hesiod's Theogony (8th century BCE). It seemed to him that all things either came from moisture or were sustained by moisture. He concluded that the universe grew from water. According to Thales the earth is a flat disc floating on a sea of water. The unique element in the cosmology of Thales is the idea that the universe developed over time through natural processes from some undifferentiated state. The first recorded use of a physical model in explaining a natural phenomenon is Thales belief that earthquakes are caused by disturbances in the water that supports the earth.

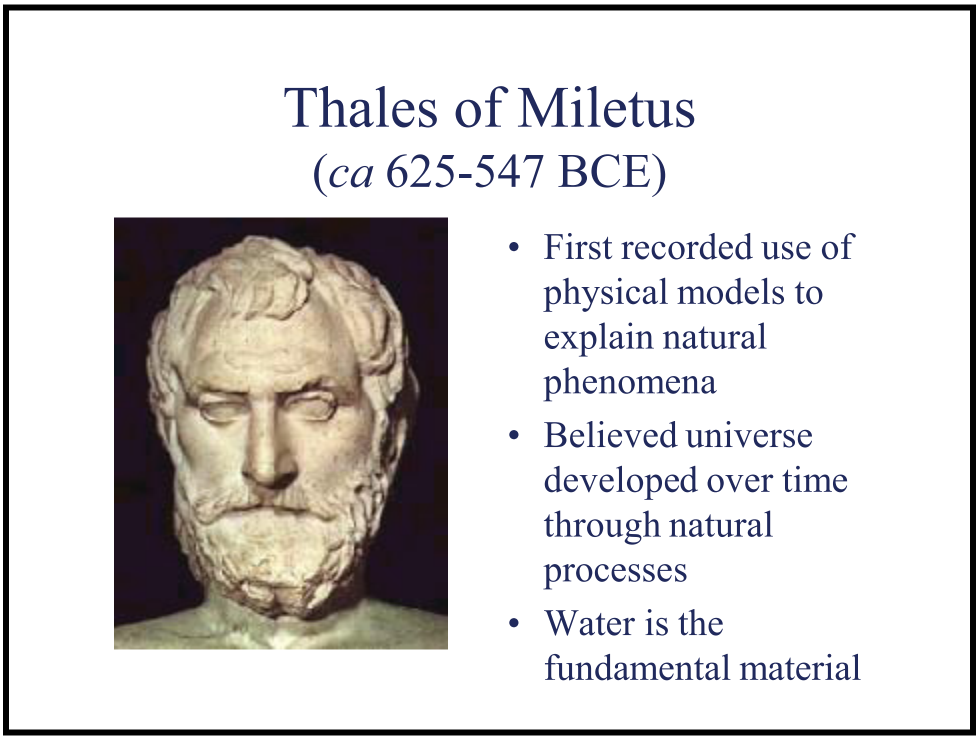


Figure 2: Thales of Miletus.

As with any human being, Thales was constrained by the level of knowledge available at the time and by the cultural and intellectual context in which he found himself. It is clear that the earlier mytho-poetic tradition exerted a strong influence on him. For example, from Aristotle we learn that included in Thales' metaphysical and cosmological doctrines was the idea that inanimate objects that move and are moved (magnets and iron, amber and wool) possess souls. It's hard to know what Thales meant by this exactly, but on the surface of it, it doesn't strike us as particularly scientific.

Thales had a student named Anaximander (610–546 BCE). He introduced the notion of a spherical universe, an idea that survived for more than 2000 years. He saw the earth as suspended in space (rather that floating on water). He also believed that living creatures arose from the moist elements when it had been partially evaporated by the sun. According to Anaximander, humans in the remote past resembled fish. Perhaps the first theory of biological evolution.

In the second half of the fifth century, the approach of Thales and Anaximander was adopted and extended by Leucippus of Miletus (*fl.* 440 BCE) and Democrutus of Abdera (*c.* 470–*c.* 400 BCE). Democritus constructed a complex explanation of all phenomena as the result of material interactions. He taught that the world was composed exclusively of uncaused and immutable material atoms. These invisibly minute and indivisible particles perpetually moved about in a boundless void and by their random collisions and varying combinations produced the phenomena of the visible world. This concept is known as materialism. In the words of Democritus, "nothing exists except atoms and the void; all else is mere opinion."

It is interesting to note that a central concept in the thinking of Thales, Anaximander, and Democritus is that there is no real distinction between the terrestrial and celestial realms. Only later did Greek thinking regress to needing a fifth essence (the quintessence) for celestial objects. These early Greek thinkers, known as the presocratics, were the first we know of to systematically seek natural explanations of natural phenomena. The Babylonians and ancient Hebrews had a literature embodying stories expressing awe about the heavens and the earth. However, they concerned themselves with the question 'why', and did not attempt to answer the question 'how.'

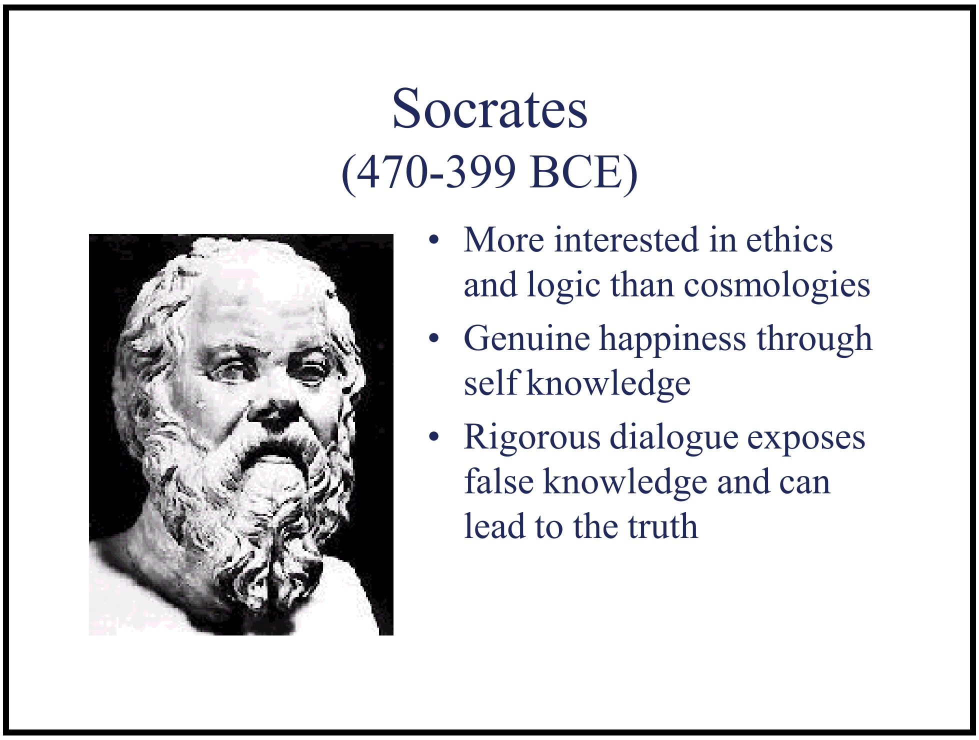


Figure 3: Socrates.

This earlier and simpler phase of Greek thought terminates in the fifth century with a thinker of an entirely different type, Socrates (470–399 BCE). With Socrates and his student Plato (427–347 BCE), we have a unique synthesis of Greek science and Greek religion. They taught that the visible world contains within it a deeper meaning, in some sense both rational and mythic in character, which is reflected in the material world but which emanates from an eternal dimension that is both source and goal of all existence. This is described in some detail in *The Dual Legacy*, pp. 69–72 in Richard Tarnas' book, *The Passion of the Western Mind*.

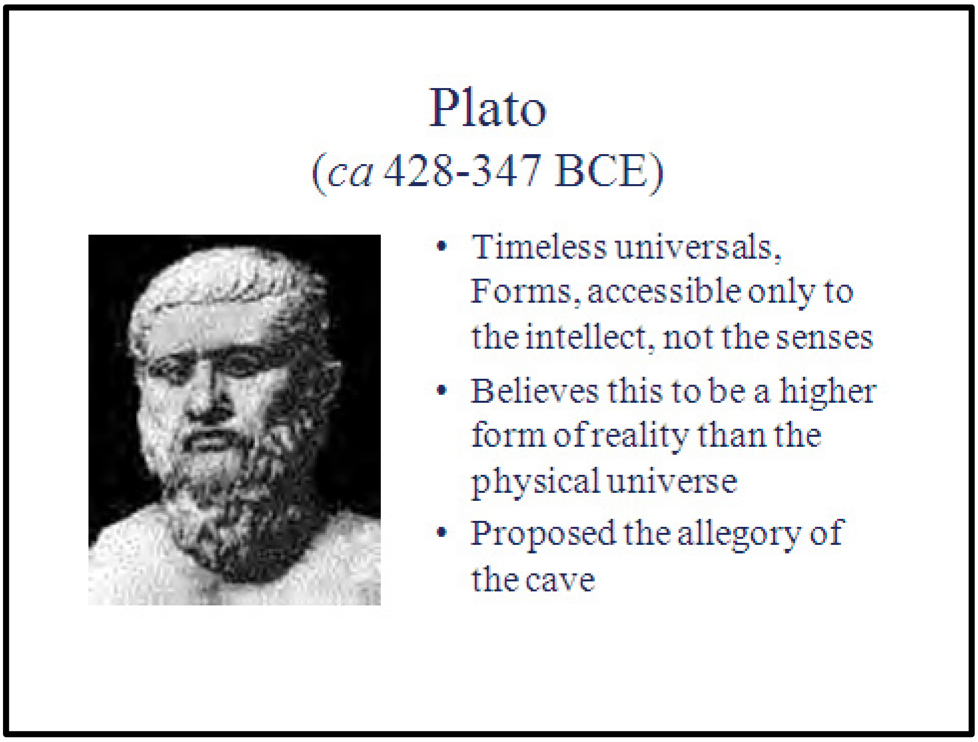


Figure 4: Plato

With Aristotle (348–322 BCE), a student of Plato's and a teacher of Alexander the Great, the pendulum began to swing back toward the more down-to-earth perspective of the presocratics. Plato asserted the existence of archetypal Ideas or Forms as primary, while the visible objects of conventional reality are their direct derivatives. These Ideas, according to Plato, possess a quality of being, a degree of reality, that is superior to that of the concrete world. On the other hand, Aristotle assumed that true reality was the perceptible world of concrete objects, rather than the imperceptible world of Plato's eternal Ideas.

Aristotle placed a new and fruitful stress on the value of observation and classification. Aristotle's writings were the first to create a comprehensive system of philosophy, encompassing morality and aesthetics, logic and science, politics and metaphysics. He provided a language and logic, a foundation and structure, and, not least, a formidably authoritative figure without which the philosophy, theology, and science of the West could not have developed as they did.

Aristotelian philosophy, in a more or less modified form, was absorbed by the various philosophical schools of antiquity and, as discussed in a later chapter, played a very important role in the history of Christian thought.

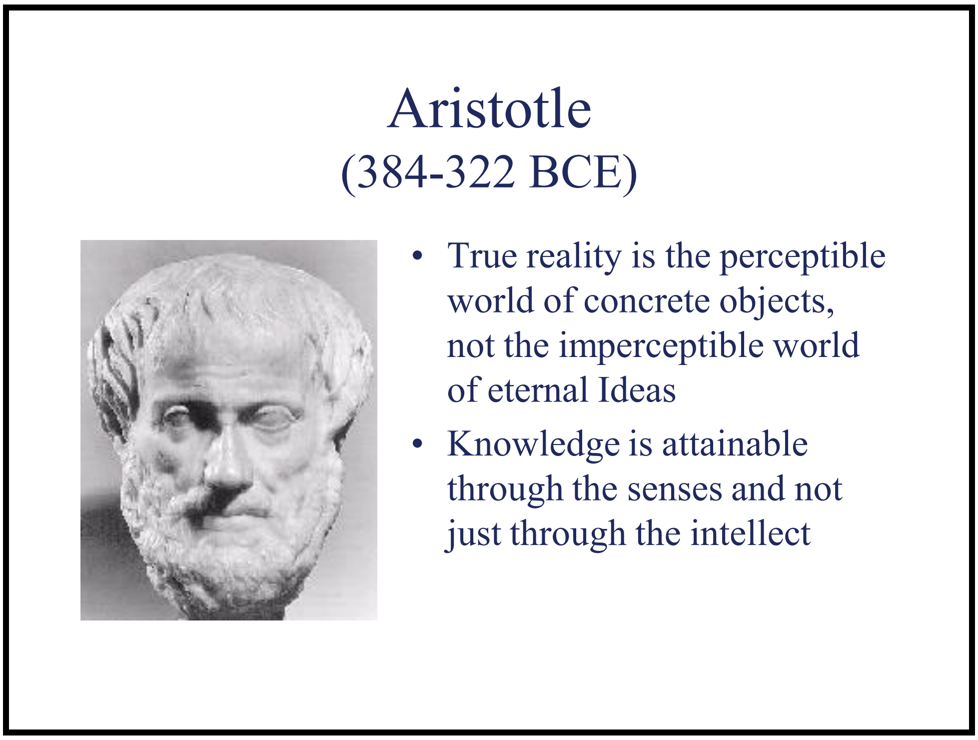


Figure 5: Aristotle

1.3 The Greek Worldview

The Greek worldview is based primarily on the teachings of Aristotle and its defining characteristic is that it is geocentric. It is the most long-lived cosmological model in history, lasting into the 17th century.

Aristotle taught that the universe was spherical and finite, with rotating spheres carrying the moon, sun, planets, and stars around a stationary earth at its center. To support the fact that the earth did not move, he pointed out that if the earth were in motion, an observer on it would see the fixed stars as shifting their positions with respect to one another, a phenomenon known as parallax. However, parallax was not observed.

Aristotle offered several proofs that the earth was a sphere. One such proof involved lunar eclipses. At the time, it was known that lunar eclipses were caused by the shadow of the earth falling on the moon. The fact that the shadow was always circular showed that the earth was a sphere. He also pointed out that when one travels northward or southward, the position of the North Star changes with respect to the horizon. To further bolster his argument that the earth was the immovable center of the universe, he proposed a comprehensive theory of motion that required all earthly substances to naturally move toward the center of the earth.

Aristotle accepted Empedocles' view that there are four earthly elements, earth, air, fire, and water. Aristotle added an additional element called aether or quintessence (fifth element), which he believed composed the celestial bodies. However, Aristotle rejected, on grounds of logic, Democritus' view of atoms. Because atoms have extension in space, they could not be indivisible. Extension in space also implies composition and therefore atoms could not be elementary. Aristotle proposed that matter is continuous and infinitely divisible.

Aristotle's view of the universe was hierarchical, and he made a sharp distinction between the sublunar world of change, and the eternal and immutable heavens. Not only were they composed of different materials, they obeyed different laws.

A Greek thinker living after Aristotle, Aristarchus (*c.* 310–230 BCE), used geometric arguments based on eclipses and the phases of the moon to estimate the relative sizes of earth, moon, and sun. He showed that the sun was many times larger than the earth and that the moon was much smaller. Reasoning that the smaller object should orbit the larger object, he concluded that the sun not the earth was the center of the universe, and that the earth orbited the sun once a year while spinning on its axis once every 24 hours.

This idea was taken seriously by Greek thinkers but ultimately rejected. If the earth were to orbit the sun, the positions of the stars must show parallax, which was not observed. Aristarchus argued that it was not observed because the nearest stars are at extremely great distances compared to the distance from the earth to the sun. In fact, the correct explanation. The nearest stars are so far away that parallax was not observed until the early 1800s, long after it had been established that the earth did orbit the sun.

Eratosthenes (*c.* 276–195 BCE), a younger contemporary of Aristarchus, used geometry to determine the circumference of the earth. It was known that at noon on a particular day of the year, the sun shone directly down a vertical well in Syene in southern Egypt. At that same time in Alexandria, north of Syene, the sun was not directly overhead, Eratosthenes was able to measure the angle that the sun made with the vertical as one-fiftieth of a circle (about 7°). From this, he was able to conclude that the earth's circumference is 50 times the distance between Alexandria and Syene.

It is not possible to evaluate precisely the accuracy of Eratosthenes' solution because there is some uncertainty about the length of the unit he used. However, it is certain that his value was within 20% of the correct answer and may have been as close as 1%. (Eratosthenes' value was much closer to correct than the one used by Columbus almost 2000 years later. Columbus thought the earth was much smaller than it is. Had he known and accepted Eratosthenes' value, it would have been obvious that he could not reach China by sailing west.)

Hipparchus (*c.* 190 BC–*c.* 120 BC), refined the method of Aristarchus for measuring the relative distances and sizes for the earth, moon, and sun, and obtained better results. He compiled an accurate catalog of the positions of over 850 stars. Hipparchus had access to 1000 years of Babylonian records of star positions. By comparing these with his own, he was able to determine that the earth's axis of rotation sweeps around in a cone much like the motion of a top inclined with respect to the vertical. The earth's axis takes about 26,000 years for one complete sweep. This motion is referred to as the precession of the equinoxes.

The contribution made by Hipparchus represent the last significant contribution to science made by the Greeks until the time of Ptolemy in the second century of the Common Era. By that time Greek science and culture had spread to the entire Mediterranean region. This spread was due primarily to Alexander the Great.

Alexander's father, Philip II was king of Macedonia, a region directly north of Greece, from 359 BCE until his assignation in 336 BCE. Philip chose Aristotle as tutor for Alexander. Upon ascending to the throne upon his father's assignation, Alexander continued his father's military campaigns. He eventually extending the empire from India to Spain, including Persia and Egypt, resulting in the universalization of Greek culture throughout the Mediterranean region..

Claudius Ptolemaeus, better known as Ptolemy, lived in Alexandria, Egypt in the middle of the second century CE. Relying heavily on the work of Hipparchus, Ptolemy compiled 13 volumes containing all known astronomical knowledge. Some of the material was original with Ptolemy.

Ptolemy is best known for his detailed geocentric model of the universe. The notion that celestial motions must always be represented by constant speeds and circular orbits was well established in Greek thought. Ptolemy retained this concept but added additional circles within circles to more accurately represent the observed locations of the planets in the sky. Although the model was very cumbersome, it was sufficiently accurate to remain the accepted view of the universe until the 17th century.

The works of Ptolemy and other Greek tinkers were stored in the great library at Alexandria. One of the last guardians of this knowledge was Hypatia (c. 375–415 CE). An accomplished astronomer, she invented astronomical navigational devices and wrote a commentary on Ptolemy's work. At the time there was considerable conflict between the newly emerging Christian community and science, which they viewed as a pagan philosophy detrimental to the Christian faith. Hypatia was a pagan. She was murdered by an anti-intellectual mob in one of the riots that plagued Alexandria during its decline.

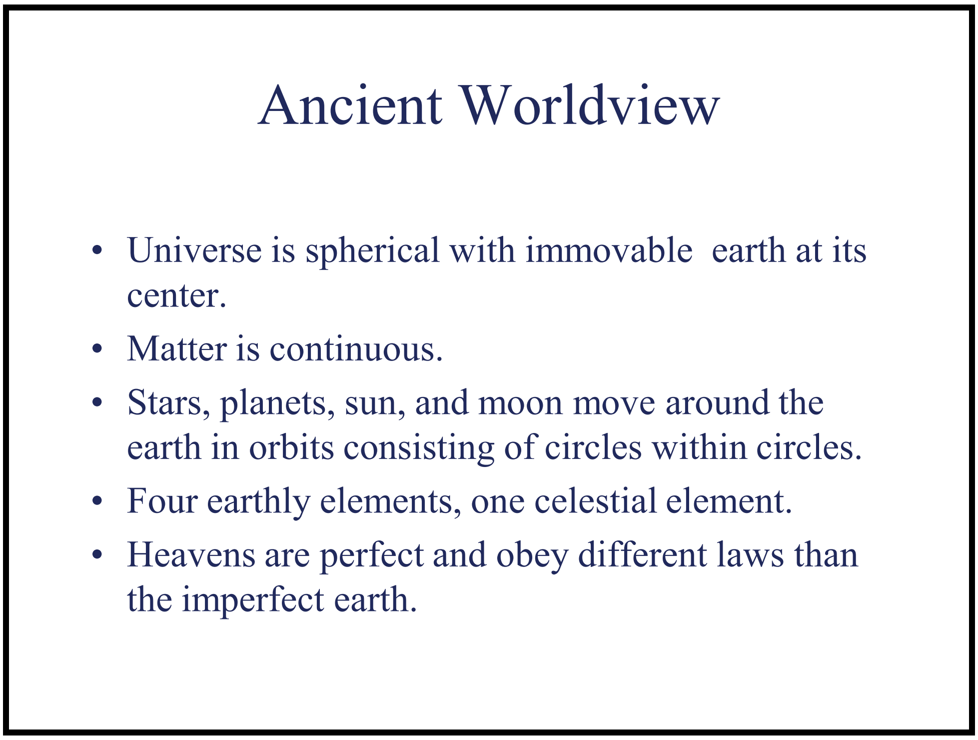


Figure 6: Ancient worldview.

1.4 Islam & Science

With the fall of the Roman Empire, science temporarily came to an end in Europe. However, the Greek tradition was carried on in the Islamic cultures. In the century, following the death of Muhammad (*c.* 570–632 CE), his followers conquered all the Middle East to India, as well as North Africa and most of Spain. By 750, they had become tolerant of diverse ideas and attitudes, and conditions for intellectual life improved significantly in their domains. For the next 250 years, the caliphs of Baghdad became patrons of science and the city a center of learning. By the eleventh century the library of the caliph of Cairo contained roughly 150,000 volumes. By contrast, a Western monastery considered itself fortunate to have 150.

An important factor in the development of Islamic science was the old Greek writings that the Arabs found in the lands they conquered. Within just a few decades after 750, the major Greek scientific works were translated into Arabic. By the end of the tenth century essentially all the known Greek manuscripts had been translated. The atmosphere of tolerance was such that this work was done by Christians, Jews, and pagans, as well as by Islamic scholars.

The religious requirements of Islam were a powerful impetus to the study of astronomy. For example, Islam adopted (and generally still uses) a strictly lunar calendar, which, however, begins not at the time of the new moon, but at the first sighting of the crescent moon just after sunset. Calculating when this occurs requires fairly complex geometry. Since the lunar year is about 354 days long, the months of the Islamic year cycle through the seasons in about 33 years. Mosques had to be oriented toward Mecca, and Muslims were to pray facing in that direction. Furthermore, time-keeping was required to properly set the five times for daily prayer.

In attacking such problems, Islamic scholars developed mathematics, especially trigonometry and spherical geometry, far surpassing the Greeks. They learned of the 'sine' trigonometric function from India, and invented the other five. They also derived some of the relations among these trigonometric functions, such as the law of sines. With these developments, it was far easier to solve the geometric problems of astronomy. The Arabs also used a system of numbers that included the concept of zero, which they had acquired from India. This was later adopted by the West. These Arabic numerals, as they are called, replaced Roman numerals, primarily because they are far easier to calculate with. Arab texts describing these new mathematical developments, as well as various summaries and commentaries on Greek and Roman texts, eventually found their way to the West, where they became standard works for centuries.

From the scientific point of view, however, the Muslims' most important contribution was that they preserved much of the Greek learning. Beginning around the year 1000, these texts made their way back into Europe. Cities near the boundaries between Islamic and Christian domains, such as Toledo in Spain, became centers of a "translation industry" (from Arabic to Latin) and Arabic words such as zenith, nadir, alchemy, algebra, and algorithm entered our language, along with star names such as Algol, Aldebaran, Alcor, Vega, Deneb, and Betelgeuse.

Perhaps the best known, and most significant, text recovered from the Islamic culture is the *Almagest*, the English name derived from the Arabic, *al-majisti*, meaning 'the greatest.' The *Almagest* was written by Ptolemy in about 150 CE and translated into Arabic about 827. In the last half of the 12th century, it was translated from Arabic to Latin and served as the basis for European astronomy until about the beginning of the 17th century.

2 The Middle Ages, the Renaissance, and the Reformation

Your goals for this chapter are to know the following:

- Characterize and give approximate dates for the following historical eras: the Greco-Roman (Classical) era, the Early Middle Ages (Dark Ages), Late Middle Ages, the Renaissance, and the Reformation.
- What were some of the advances that occurred in the Late Middle Ages the differentiated it from the Early Middle Ages and eventually led to the Renaissance?
- What roles did the Christian church play in history from its beginnings up through the Early Middle Ages?

The Romans began building their empire in the 4th century BCE and eventually controlled the entire Mediterranean world. They conquered Greece in the 2nd century BCE. However, the Romans recognized the superiority of much of Greek thought and culture and incorporated it into their culture. The resulting civilization is known as Greco-Roman.

In 313 CE the Roman emperor Constantine issued the Edit of Milan recognizing the right of Christians to practice their religion within the Empire. Gradually Christianity became its official religion. By the 4th century, the Empire was beginning to collapse. Commerce, particularly seaborne commerce, was in decline as was the population, especially in urban areas. The area under the control of Rome was shrinking and Rome itself was being subjected to barbarian invasions. This period of time is usually referred to as Late Antiquity. This is a transitional period, during which the ancient world slid slowly into the medieval.

2.1 Late Antiquity and The Early Middle Ages

The Middle Ages is a derogatory term coined after the fact to reflect the view that this period was a time of intellectual stagnation between that of classical (Greco-Roman) culture and its later reestablishment during the Renaissance (rebirth). At the time it first appeared in the 15th century, it referred to the period from the fall of the Roman Empire (~500 CE) to the beginning of the Renaissance (~1350). Later, the term referred to the division of history into Classical, Medieval, and Modern. More recent scholarship has somewhat modified the earlier divisions. Recognizing that the Roman Empire was in serious decline before 500 and that some aspects of Greco-Roman culture continued after 500, the era Late Antiquity was added.

Various dates have been given for these eras. One version accepted by many scholars is: Late Antiquity (300–700), Early Middle Ages (700–1000), Late (or High) Middle Ages (1000–1350), and Renaissance (1350–1600). These dates are somewhat arbitrary and, of course, do not correspond to abrupt, discontinuous cultural changes.

Although the term Dark Ages is no longer used to characterize Late Antiquity and the Early Middle Ages, these eras were none the less a time of wide spread political insecurity, famine, and illiteracy. As the authority of secular Rome declined, the vacuum was slowly filled by the Roman Church. By 500 the Church was essentially the only authority in Western Europe.

Almost from the beginning of Christianity, safeguarding the faith was the Church's almost exclusive priority. Dialogue was often curtailed all together least faith be undermined and the authority of the Church questioned. There was a reaction against the pagan philosophers. Tertullian, a 3rd century Church leader, said "What has Athens to do with Jerusalem?" One consequence is that the value of observing, analyzing or understanding the natural world was greatly diminished. "The pluralism of classical culture, with its multiplicity of philosophies, its diversity of polytheistic mythologies, and its plethora of mystery religions, gave way to an emphatically monolithic system – one God, one Church, one Truth." Richard Tarnas, *The Passion of the Western Mind*, p. 118–119.

Needless to say, not much progress in the understanding of the universe was made during this period. In fact, just the opposite. Much of what was known in the classical era was lost. Books were burned. For the Church authorities, direct study of the natural world was seen as a threat to the integrity of religious faith and thus to salvation.

One of the more absurd results of this was a rejection of the idea of a spherical earth, in part because the idea was supported by the pagan philosophers. Instead, the universe was seen as shaped like the Holy Tabernacle; the earth enclosed within a domed rectangular box. Based on the story in Genesis that the firmament is enclosed by water, the model included super-celestial waters resting on top of the tabernacle. Not until about the end of the 9th century was the idea of a spherical earth reinstated.

2.2 The Late Middle Ages

In the later Middle Ages, Christianity's earlier need to distinguish and strengthen itself by the more or less rigid exclusion of pagan culture lost some of its urgency, and a more relaxed attitude toward secular learning developed. "It seems to me a case of negligence if, after becoming firm in our faith, we do not strive to understand what we believe." Saint Anselm of Canterbury, (1033–1109). Contrast this with the attitude of Tertullian in the 3rd century, "All curiosity is at an end after Jesus, all research after the Gospel. Let us have Faith and wish for nothing more."

The change in the Church's attitude was accelerated by the rediscovery of a large body of Aristotle's writings that had been preserved by Islamic and Byzantine cultures in the East. The Arabic and Greek text were translated into Latin and widely circulated in the new universities of the West. At first the Church tried to suppress the teachings of Aristotle as they were in conflict with those of neo-Platonism which Saint Augustine, in the early 5th century, had established as the philosophical foundation of Christianity. This turned out to be impossible, and finally in the mid-13th century, Saint Thomas Aquinas was able to integrate Aristotle with Christian theology, in much the way that Augustine had before with Plato.

Aquinas blended Aristotelian philosophy and Christian doctrine by suggesting that rational thinking and the study of nature, like revelation, were valid ways to understand truths pertaining to God. According to Aquinas, God reveals himself through nature, so to study nature is to study God. The writings of Aquinas became the new philosophical foundation for the Christian religion and Aquinas' view of the relationship between God and the universe became the cosmological worldview of the Late Middle.

For Aquinas, there is really no difference between the creation of the universe and its continued existence. Aquinas argued that the relation of a clock to the clockmaker is very different from the relation of the universe to its Creator. Once the clock is made, it no longer has a relationship of dependence on its maker. However, since the universe was created from nothing rather than pre-existing materials, as is the case with the clock, the universe would fall back into non-being without the Creator's continuing omnipotent support. God sustains the world. Nature is infused with God's power and intentions. Thus, for Aquinas, knowledge of God can be obtained through natural science due to God's omnipresence in the universe.

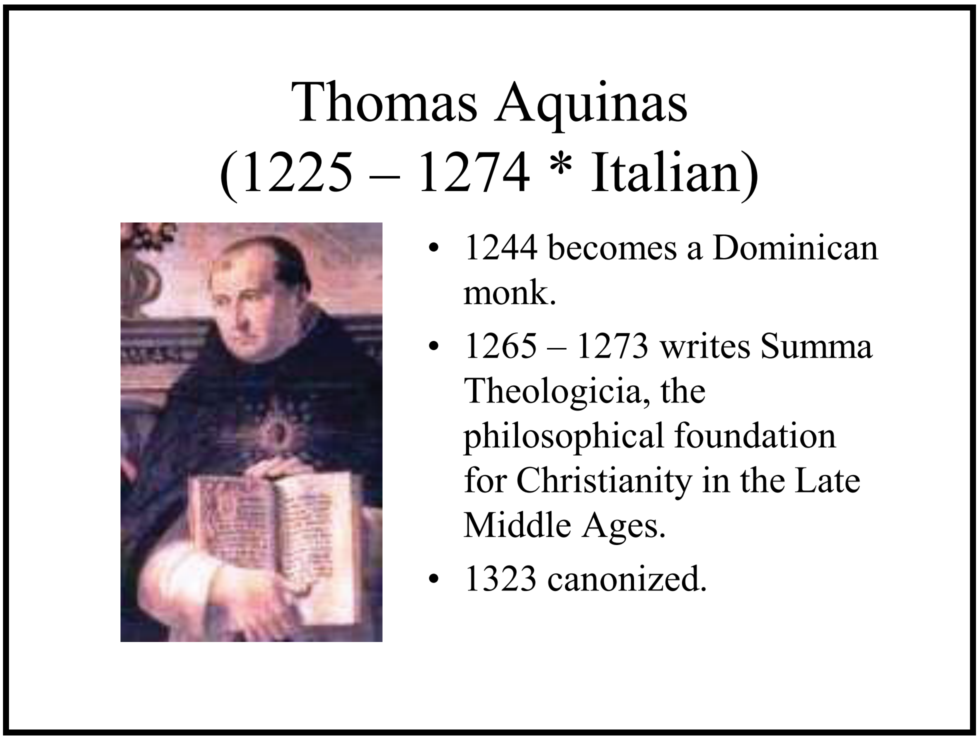


Figure 1 Thomas Aquinas

We will return later to the relationship between a clock and its maker. For some people of the Enlightenment, the relationship between the universe and its Creator is precisely the same as that between the clock and the clockmaker. This idea is called Deism. For others, the clockmaker analogy is used as an argument for God – the cosmological argument or the argument from design. This argument has recently been made by proponents of Intelligent Design who have proposed that it be taught in the public schools as an alternative to evolution.

Through the work of Aquinas and other Scholastics, the Aristotelian-Ptolemaic cosmology was reintroduced to Europeans and, at the same time, permeated with Christian meaning. This model was enthusiastically embraced by the Italian poet Dante Alighieri. In his epic poem La Divina Commedia, he presents a moral, religious, and cosmological paradigm whose cosmological architecture created a comprehensive Christian mythology, one encompassing the whole of creation. The Commedia beautifully summarizes the late medieval worldview.

In retrospect, the blending of Aristotelian thought with theology may have been a mistake. Because the cosmology of Aristotle was so fundamentally a part of his philosophy, the synthesis had the (perhaps unintended) consequence of making the Greek view of the physical universe a part of Christian dogma. The Church became locked into a geocentric model of the universe. During the Scientific Revolution this would bring the Church into direct conflict with science, one factor that eventually eroded its authority.

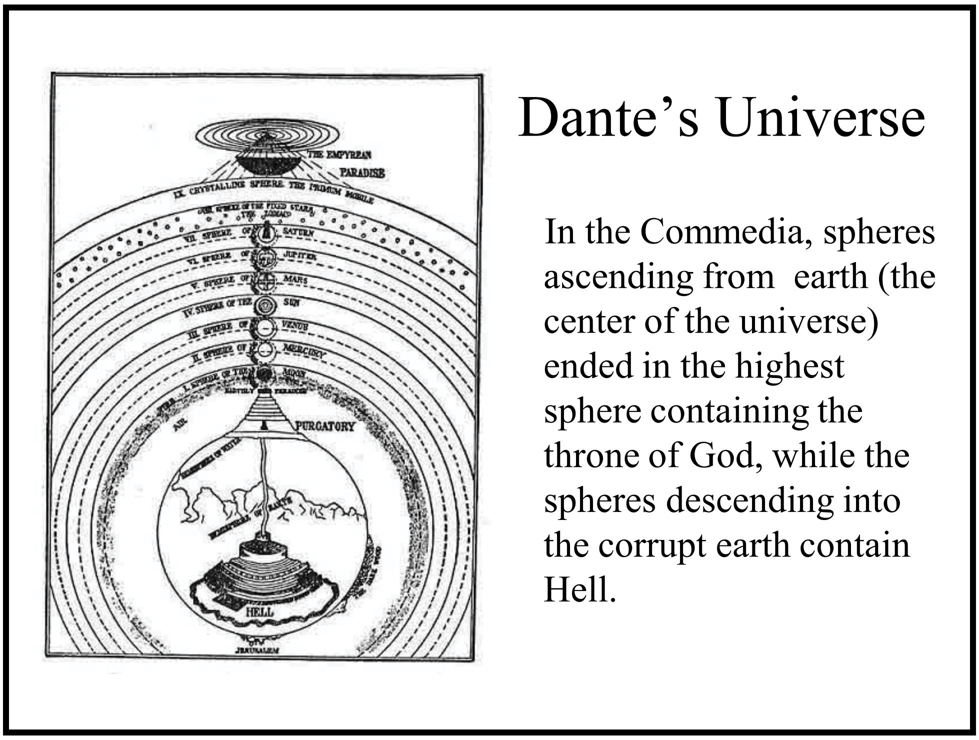


Figure 2: Dante's Universe

In the meantime, events were occurring that would transform the medieval period into the Renaissance. With the rise of principalities and nation states, there was increasing political security and a gradual increase in secular power. Innovations in agriculture resulted in population increases, particular in the cities. A middle class began to develop. There was a corresponding increase in literacy.

2.3 The Renaissance

The Renaissance began in Italy sometime around the mid-14th century. It can be characterized by a number of cultural shifts. Primarily it was a rebirth of Greco-Roman culture, the literal meaning of renaissance. The Italian commercial and aristocratic elites saw classical culture as a source not just for scientific knowledge and rules for logical discourse, as it had been for the Scholastics of the Late Middle Ages, but for the deepening and enrichment of the human spirit. There was a reaction against Aristotle and a revival of Platonism, in part because of his superior literary style. Humanist scholars and artists flourished in the new cultural climate. Forsaking the ideal of monastic poverty, Renaissance man embraced the enrichment of life afforded by personal wealth. "There was…an emphatic emergence of a new consciousness – expansive, rebellious, energetic and creative, individualistic, ambitious and often unscrupulous, curious, self-confident, committed to this life and this world." Richard Tarnas, *The Passion of the Western Mind*, p. 231. All in all, the Renaissance thinkers had a much more positive view of humanity and its capacities than was true in the Middle Ages.

Although many of the cultural changes, particularly the new curious, self-confident individualism, would eventually lead to renewed interest in the universe, this was not the case initially. The humanists favored human-centered subjects like politics and history over the study of natural philosophy or applied mathematics. Later, however, interest did turn to restoration of the ancient knowledge of the universe. Some refer to this period as the Scientific Renaissance. This in turn, led to the Scientific Revolution of the 17th century. The emphasis of the Renaissance was on the recovery of knowledge, whereas in the Scientific Revolution the emphasis shifted from recovery to discovery.

2.4 The Reformation

The Reformation is the era of European history when the monolithic Roman Church was splintered into many different denominations. Because the Church was such a powerful influence in all aspects of medieval life, the Reformation was a significant social and political event as well as a religious one.

There were many factors that led to the Reformation. The Roman Catholic hierarchy had become wealthy, secular, and often openly corrupt and immoral. This, in sharp contrast to the deep piety and poverty of the Church faithful. In a sense, the straw that broke the camel's back was the selling of indulgences. The indulgences could be used to absolve sins and to lessen after-death punishments for the contributor or any of his or her relatives. It was clearly a scam. In 1517, Martin Luther, a German monk, nailed his 95 theses against the selling of indulgences to a church door in Wittenberg. The rise of nationalism was an important factor in Luther's success as he was protected from papal authority by German princes. Within the next few decades, several other religious discenters would break with the Roman Church to form what would come to be called Protestant denominations.

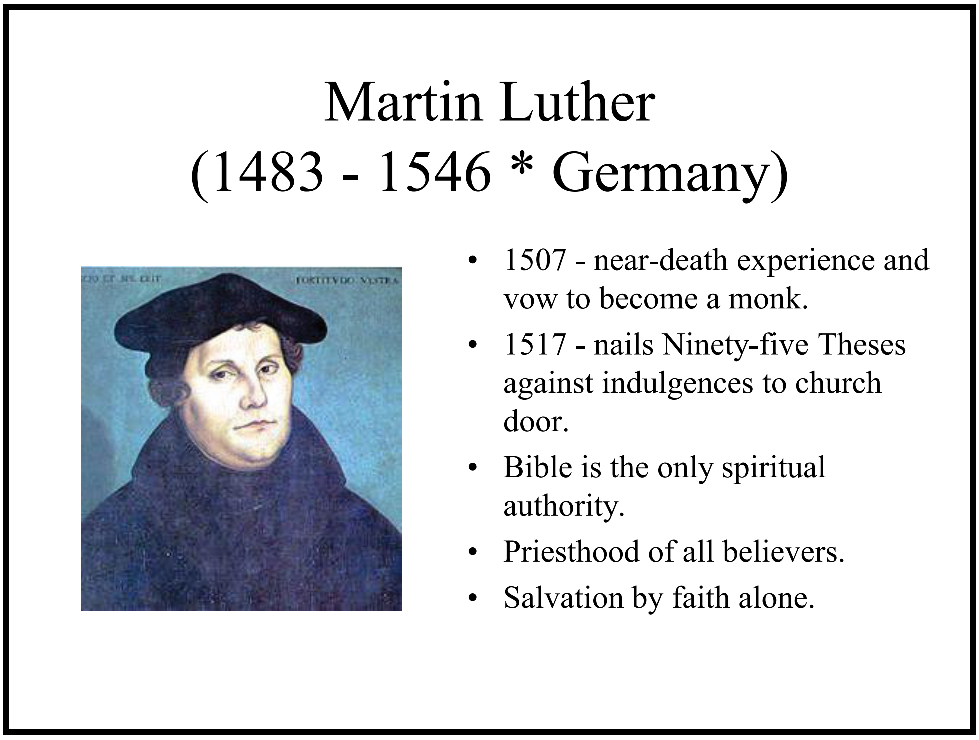


Figure 3: Martin Luther

Although the Reformation had primarily religious and social rather than scientific consequences, many believe that it was an important step leading to the Scientific Revolution and the return to serious attempts, for the first time in over 1000 years of Western European history, to come to terms with the physical universe. Perhaps the most important factor was the break Protestant theology made with Aquinas' view that the universe is infused with God's presence and hence sacred in its own right. In the Protestant worldview, the supernatural was almost completely disconnected from the natural. However, if God is radically transcendent, and thus isolated from nature, nature becomes simply an object to be used for human purposes. This change in worldview is a shift in the direction of the modern worldview produced by the Scientific Revolution.

3 The Scientific Revolution

Your goals for this chapter are to know the following:

- What is meant by the Scientific Revolution that took place between the times of Copernicus and Newton? In what ways did the modern world-view that resulted from the Scientific Revolution differ from the ancient world-view?
- How did the science of Galileo differ from that of the Greeks (Aristotle)? What were Galileo's most significant contributions?
- What was the most significant contribution of Copernicus? Brahe? Kepler? Newton?
- What is the scientific method? What does it mean to say that in order for a statement to be scientific it must be falsifiable?
- What is meant by a mechanistic-deterministic worldview? Why is a clockwork a good analogy for the mechanistic-deterministic worldview? What are Diests?

3.1 Nicolaus Copernicus

History I generally divided into eras each of which can be characterized in a way that differentiates it from all others. But, in fact, these eras blend, more or less gradually, into one another. For example, there is a sense in which the Scientific Revolution is simply a continuation of the Renaissance. History also has a tendency to date the various eras in round numbers. For instance the Late Middle Ages is usually said to have occurred between the years 1000 and 1350 and the Renaissance between 1350 and 1600. The Scientific Revolution is usually said to have occurred in the 17th century. However, all treatments of the Scientific Revolution inevitably begin with Nicolaus Copernicus, a Polish canon in a Catholic cathedral in the first half of the 15th century – a thoroughly Renaissance man in both time and spirit.

Copernicus was a scholar rather than a scientist in the modern sense of the word. As scholars of the time did, he immersed himself in the newly translated classical literature, not with the intention of making new discoveries, but of recovering old discoveries. Copernicus is sometimes credited with discovering the heliocentric model of the solar system. In fact he read about it in a book. Greek thinkers had proposed it centuries before the common era, principle among them, Aristarchus.

Copernicus read of this suggestion and realized that it explained in a simple manner many things about the motion of the planets that had complex, implausible explanations using Ptolemy's geocentric model. Copernicus felt that a satisfactory representation of the solar system should be coherent and physically plausible, not requiring a different construction for each phenomenon, as Ptolemy's system did. To him, Ptolemy's system was ugly and therefore could not represent the work of the Creator.

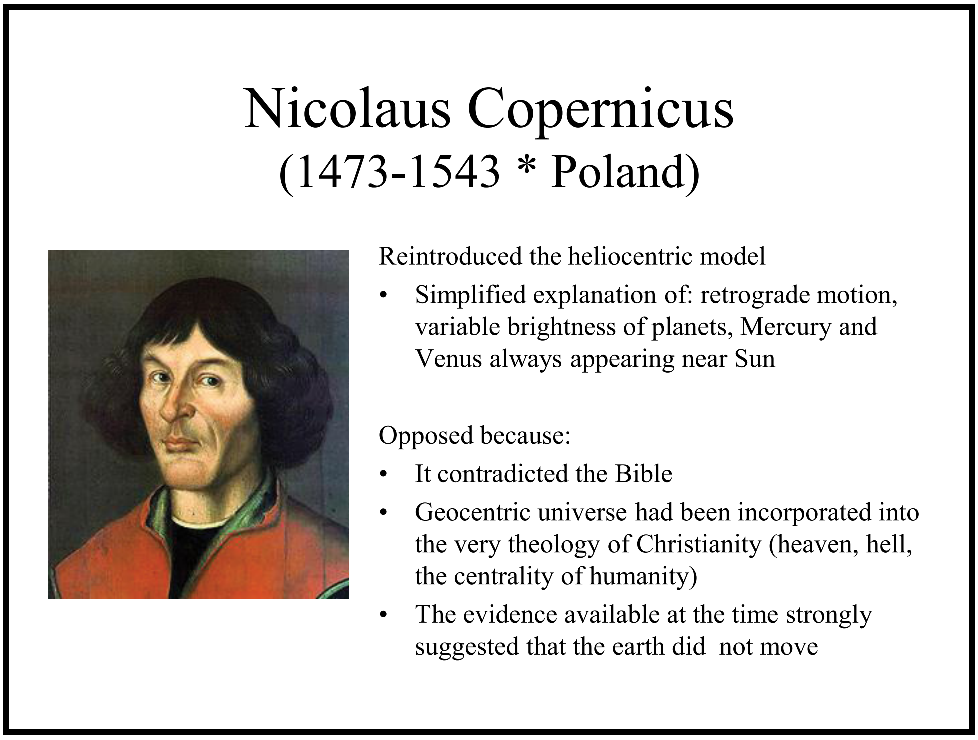
(Upon hearing in the late 13th century of Ptolemy's model of the universe and of the extremely complicated mathematics it required, Pope Alfonso X, is said to have replied, "If the Lord Almighty had consulted me before embarking on creation, I should have recommended something simpler.)

As early as 1514, Copernicus circulated among his friends a short manuscript describing his heliocentric views. However, he was reluctant to publish. Most scholars believe that it was not fear of the Church that caused his reluctance. The Church did not take a hard line on the issue at the time. In fact, it was in general supportive of Copernicus. It was only later, during the counter-Reformation, that people such as Giordano Bruno and Galileo Galilei suffered retribution for their views on the nature of the universe.

In 1533, Johann Widmannstetter, personal secretary to Pope Clement VII, delivered a series of lectures in the Vatican gardens, outlining Copernicus' theory. Clement and several Cardinals heard the lectures and were interested in the theory. Clements successor, Paul III, probably heard of Copernicus' ideas from Cardinal Schonberg, a confident of popes Leo X, Clement VII, and Paul III. At Paul's urging, Schonberg wrote Copernicus on 1 November 1536, saying in part, "Therefore, learned man, without wishing to be inopportune, I beg you most emphatically to communicate your discovery to the learned world…"

In spite of this support, Copernicus waited six years after receiving Schonberg's encouraging letter to publish. Finally, however, his friend and student, George Rheticus, convinced him that it was time to do so. Copernicus died in 1543. He is said to have received a copy of the printed book, consisting of about 200 pages written in Latin, for the first time on his deathbed. The book is entitled, *On the Revolutions of the Celestial Spheres*.

The likely explanation for his reluctance to publish is that Copernicus was concerned about how his ideas would be received by both the devout masses and by his fellow scholars, all deeply committed to the Aristotelian point of view. To put this outrageous idea forward with no evidence, other than its greater simplicity, would invite heavy criticism. In his dedication of the book to Pope Paul III, he mentions his concern that after people heard of his views, he would be "hissed off the stage." Additionally, Copernicus was a perfectionist, always striving to improve his presentation of the theory.



Fighre 1: Nicolaus Copernicus

The initial religious reaction against the theory came not from the Catholics but from the Protestants. The Copernican hypothesis contradicted several passages in Holy Scripture concerning the fixity of the earth, and Scripture was Protestantism's one absolute authority. Even before the publication of the book, Martin Luther heard of the theory and is reported to have said, "The fool wants to turn the whole art of astronomy upside-down. However, as Holy Scripture tells us, Joshua bid the sun to stand still and not the earth." The Catholic Church did eventually, 73 years after its publication, put the book on the Index of books Catholics were forbidden to read.

3.2 Tycho Brahe

Copernicus got it right about the earth going around the sun, but he got the orbits in which the earth and other planets orbit the sun wrong. Copernicus' model continued to use the circles-within-circles orbits of Ptolemy's model. His model consisted of a moving earth in a cosmos otherwise ruled by Aristotelian and Ptolemaic assumptions. Observations soon showed that Copernicus' model was somewhat better at predicting the exact locations of the planets in the sky at some future date, but still not completely accurate. Both models had to be wrong. The Danish astronomer, Tycho Brahe, set himself the task of coming up with the correct model.

Brahe realized that progress in astronomy required systematic, rigorous observation, night after night, using the most accurate instruments available. This program became his life's work. Brahe improved and enlarged existing instruments, and built entirely new ones. The telescope had not been inverted yet, so all of Brahe's instruments were naked-eye. Brahe began making observations and recording data in 1572 and continued to do so until his death in 1601.

His model of the solar system was a hybrid of the geocentric and the heliocentric. The simpler explanations for the locations of Mercury and Venus in the sky and for the retrograde motion of Mars and the outer planets convinced Brahe that the planets must orbit the sun as in Copernicus' model. However, he was convinced that the earth did not orbit the sun, in part because he was never able to measure parallax, an unavoidable prediction of a moving earth. As accurate as his measurements were, the distances to the nearest stars required more than 200 years of technological advances before parallax was observed. In Brahe's model, the sun, with its orbiting planets, orbited the earth. It was a geocentric model.

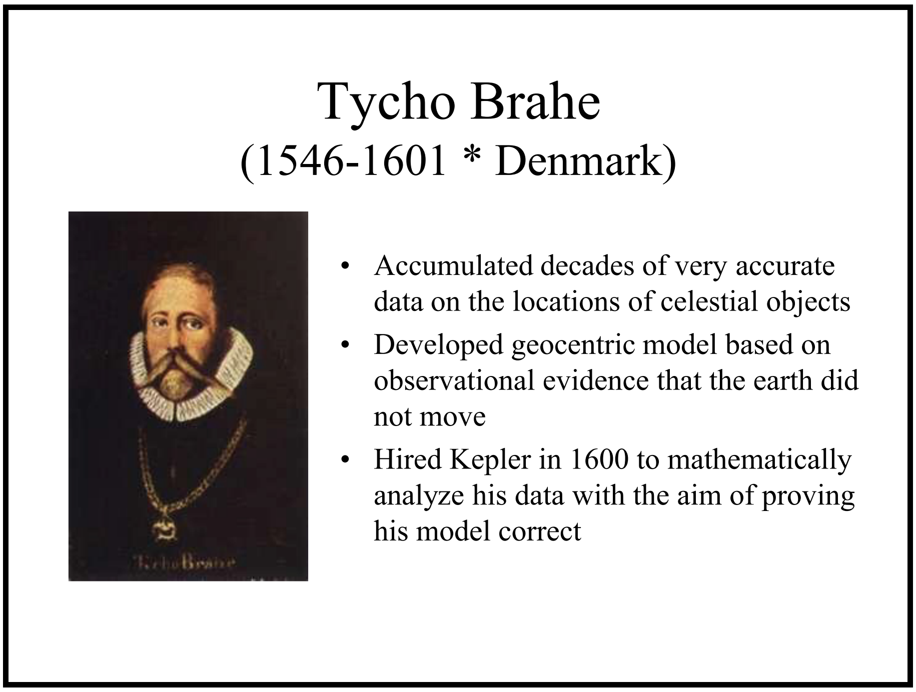


Figure 2 Tycho Brahe

Brahe planned to use his extensive data on the locations of the planets to demonstrate the correctness of his model. The idea was to use the model to calculate the position in the sky of a planet at some point in the past. He would then go back to his data to show that the calculation produced the actual observed location. If it could consistently do this, as Ptolemy's and Copernicus' could not, then his model would be shown to be correct. However, these calculations were extremely difficult. Brahe could not do them himself, so, in 1600 he hired the German mathematician, Johannes Kepler to do them for him.

3.3 Johannes Kepler

Although he was hired to make the calculations necessary to demonstrate the correctness of Brahe's geocentric model, Johannes Kepler had for some time been a convinced Copernican. Not that he believed that Copernicus' model was correct in all its details; he knew that its slight inaccuracies meant that it was wrong. But the aesthetic superiority of a heliocentric view was compelling to him.

Brahe died shortly after Kepler was hired. Kepler succeeded him as imperial mathematician and astrologer to the Holy Roman Emperor, with the responsibility of completing Brahe's unfinished work. Kepler now had access to Brahe's decades of unprecedentedly accurate astronomical observations.

Kepler had entered Brahe's employment with a specific heliocentric model of his own. Now he had the opportunity to check his model against the data, and soon found that his model was wrong. He did not give up, however. For four years he repeatedly devised new models, checked them against the data, and found out that they were wrong. In these attempts he focused on the planet Mars. He reasoned that the Creator would not have created a different orbit for each planet; that would be unaesthetic, something incompatible with his concept of God. If he could figure out the orbit of Mars, he was sure it would be the orbit of all the other planets as well.

After years of unsuccessful attempts using various combinations of circles, he finally gave up on that approach. Finally, in 1605 he hit upon the correct combination of path and speed that would match the calculations to Brahe's observations. Mars moves in an elliptical path with varying speeds depending of the distance between it and the sun. Mars speeds up as it approaches the sun and slows as it recedes. It does this in such a way that an imaginary line drawn between Mars and the sun sweeps out equal areas in equal time intervals. As he had suspected, this orbit worked for the other planets as well. Although the manuscript presenting this information was completed in 1605, it was not published until 1609 due to legal disputes over the use of Tycho's observations, the property of his heirs.

This correct orbit was arrived at strictly by trial and error. Kepler had no model in mind that allowed him to predict it and no clear explanation for why the planets moved in this way. The explanation would have to wait more than 50 years for Isaac Newton to figure it out. However, the accuracy with which the model was able to predict the past locations of the planets in the sky, as verified by Brahe's observations, left little doubt that the orbit was correct.

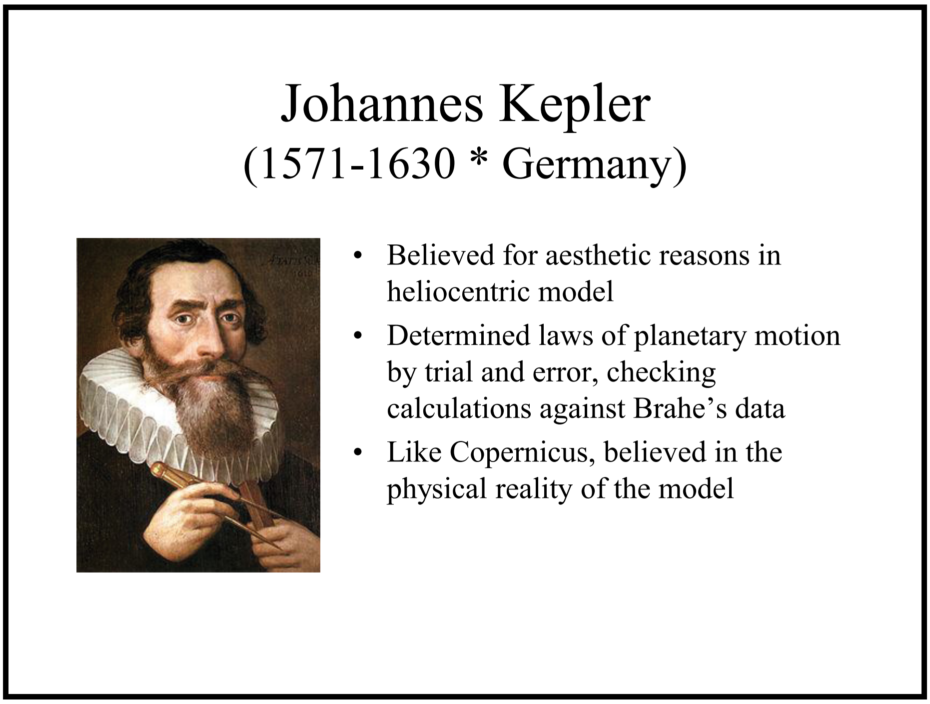


Figure 3: Johannes Kepler

3.4 Galileo Galilei

Galileo Galilei was a contemporary of Kepler, and, like Kepler was a convinced Copernican long before there was anything other than aesthetic reasons for supporting the heliocentric model. Other than that, the two men had little in common. Kepler was very mild mannered, somewhat sickly, and modest. Galileo was the opposite. Kepler was a Protestant and Galileo a Catholic, both strong in their faith. Galileo dismissed much of Kepler's work as useless fiction and refused to accept elliptical orbits for the planets, continuing to believe they had to be circular in some way.

Galileo is significant in science for two distinct reasons. First of all, he was the first, in 1609, to use a telescope to study the heavens and in this way made several important discoveries that undermined the Ptolemaic model accepted by most scholars and the Christian churches, both Catholic and Protestant. However, these discoveries did not prove that the earth itself orbited the sun, as Galileo liked to claim. Secondly, he is generally credited with inventing the scientific method as we understand it today, or at the very least, with being the first to apply it systematically.

Although Galileo did not invent the telescope, he was the first to use it to gain knowledge of the heavens. Among his discoveries were the mountains and craters on the moon. Because the moon was part of the celestial realm, Aristotle and religious dogma required it to be perfect. Well, almost perfect. It was clearly blemished, perhaps signifying that as the closest celestial object to the earth, it was a transitional object between the imperfect earth and the absolutely perfect heavens beyond. In any case, the scholars and churches taught that the moon is a perfectly smooth and spherical object. If you already believe this and look at the moon with the naked eye, it is easy to believe this is true.

Even through Galileo's relatively low power telescope, it clearly is not true. Galileo had trouble convincing others of this. They either refused to look through the telescope or claimed that the irregularities were an artifact of the telescope itself rather than a true image of the moon. The resemblance of the moon's features to those on the earth misled Galileo somewhat. He thought that the dark relatively smooth surfaces on the moon were oceans and named them seas. Today we call them maria, the Latin word for seas.

Galileo discovered that the planet Venus went through phases just as the moon does. This was important because it proved that Venus orbited the sun rather than the earth, thus proving the Ptolemaic model wrong. He was also able to demonstrate what some others had suspected, that the Milky Way, the band of diffuse light that arcs across the sky from horizon to horizon, is actually composed of hundreds of thousands of stars. He observed sunspots and used them to calculate the speed of rotation of the sun, about one revolution every 25 days.

Perhaps his most important discovery was the four (now called Galilean) moons of Jupiter. One of the strongest arguments in favor of the geocentric model was the fact that our moon orbits the earth. No one disputed this. The argument went that the earth could not possibly move because if it did, it would leave the moon behind. In the days before the discovery of gravity, this was a very powerful argument. However, whether one believed in a geocentric or a heliocentric universe, it was clear that Jupiter moved; it had to orbit something whether the earth or the sun. The fact that Jupiter was somehow able to move without leaving its moons behind destroyed the argument.

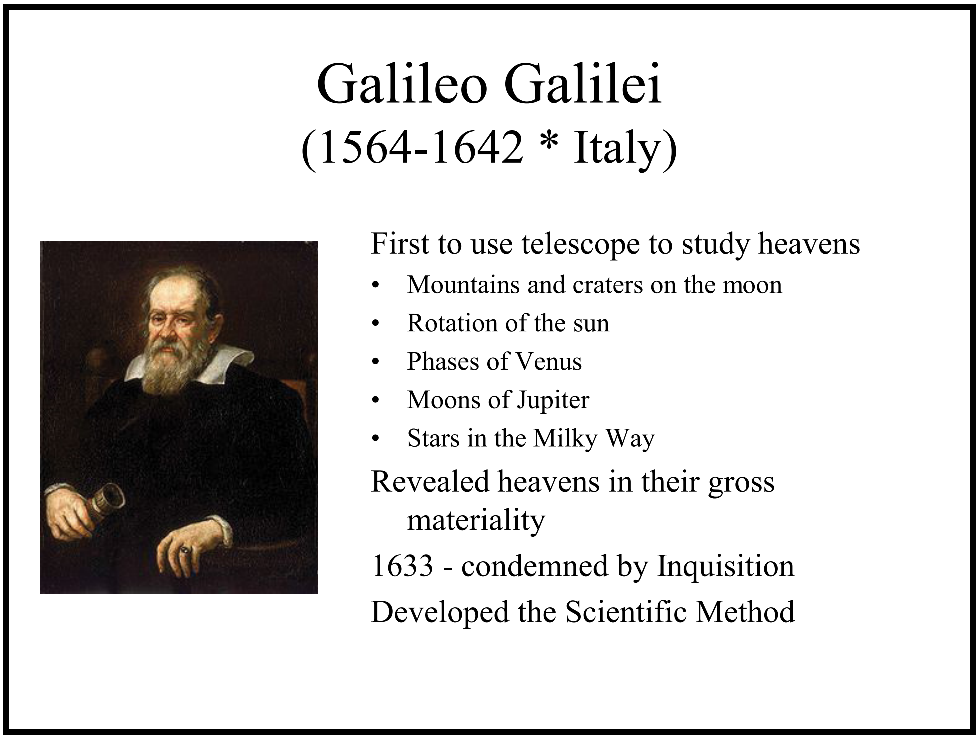


Figure 4: Galileo Galilei

As most everyone knows, Galileo got into serious trouble with the Church later in his life. He was called before the Inquisition in 1633. The root of his problem with the Church began in 1616. By then the Catholic Church, with the Counter Reformation well underway, had joined the Protestant churches in opposing the Copernican model. Galileo went to Rome to try to persuade the Church authorities not to ban Copernicus' ideas. They did not. However, Cardinal Bellarmine ordered Galileo not to "hold or defend" the idea that the earth moves and the sun stands still at the centre. The decree, however, did not prevent Galileo from discussing the heliocentric hypothesis as a hypothesis rather than a fact. In 1623 Cardinal Maffeo Barberini, a friend and admirer of Galileo, was elected Pope Urban VIII. Galileo felt it was now safe to take a stronger position with respect to the heliocentric model. His book, *Dialogue Concerning the Two Chief World Systems*, was published in 1632.

Before publishing, Galileo personally discussed the book with Urban. The Pope asked that Galileo give arguments both for and against both the heliocentric and the geocentric models and offered some of his own in favor of the geocentric over the heliocentric model. In the Dialogue, the arguments for the geocentric model and against the heliocentric model were made by Simplicio, which in Italian has the connotation of simpleton, and, in fact, Simplicio frequently came across as a fool. The book was clearly not a balanced discussion of the two models but rather a polemic for the heliocentric model, something Galileo, in 1616, had been forbidden to do. To make matters worse, Galileo put the exact words of Urban into the mouth of Simplicio. The Pope was not amused. Galileo was called to Rome to face the Inquisition.

He was threatened with torture if he did not publicly recant, which he did. He avoided torture, but was found "vehemently suspect of heresy" and sentenced to house arrest, where he remained for the rest of his life. In spite of his troubles with the Church, he remained a devout Catholic throughout his life. His justification for proposing theories contrary to the Bible is summarized in his statement "The Bible tells you how to go to heaven, not how the heavens go."

His trial before the Inquisition ended Galileo's work as an astronomer. Fortunately for science, it did not end his work as a physicist. During his almost decade of house arrest, Galileo made original contributions to the science of motion through an innovative combination of experiment and mathematics. Galileo is perhaps the first to clearly state that the laws of nature are mathematical. His studies of motion laid the groundwork for Isaac Newton's formulation of his three laws of motion. The first of these laws, logically just a special case of the second, is simply a restatement of work done by Galileo, and was included specifically to recognize Galileo's contribution. Galileo's empirical approach in his studies of motion is what we now know as the scientific method.

The Scientific Revolution is not yet complete. Kepler has shown the heliocentric model to be correct by determining the orbits of the planets around sun, but he has no explanation for why they move in those particular orbits. Galileo has made important discoveries in mechanics, completely destroying Aristotle's theory of motion, but was not able to replace it with a similarly comprehensive theory. The unfinished work of Kepler and Galileo will have to wait another 25 years for the genius of Isaac Newton to complete their work.

3.5 Isaac Newton

It is no exaggeration to say that Isaac Newton is the single most important contributor to the development of modern science. The Latin inscription on Newton's tomb, despite its bombastic language, is thus justified in proclaiming, "Mortals! Rejoice at so great an ornament to the human race!" It is perhaps a slight exaggeration to say, as Alexander Pope did as an epitaph for Newton:

> "Nature and Nature's laws lay hid in night; God said, Let Newton be! and all was light."

Newton entered Trinity College of Cambridge University in 1661. The Cambridge curriculum at that time was still strongly classical, but Newton preferred to read the more advanced ideas of modern philosophers such as Descartes and astronomers such as Copernicus, Galileo, and Kepler. In private studies he had begun to master the field of mathematics as shown by notebooks he kept at the time. No one at Cambridge apparently recognized his genius. Newton obtained his degree from Cambridge in August 1665 without honors or distinction.

The university temporarily closed for the next two years as a precaution against the Great Plague, and Newton returned to his mother's farm. During these two years (his 23rd and 24th) Newton filled notebook after notebook with ideas and experimental observations. These may be the two most productive years in the entire history of science.

In that relatively short period of time, Newton made brilliant and important discoveries regarding light and color. He continued his studies of mathematics and invented the calculus, which he then used to describe the motion of objects. Finally, and perhaps most significant of all, he developed a mathematical equation describing gravity. Thus Newton not only knew how the planets moved, he knew why they moved that way. What Kepler had laboriously determined through trial and error, Newton using his laws of motion and the law of gravity could calculate on the back of an envelope.

Although Galileo never dropped two balls from the Leaning Tower of Pisa, the story that Newton arrived at his theory of gravity after watching an apple fall from a tree in his mother's orchard appears to be true. At least, Newton said it was.

Newton returned to Cambridge in 1667, and was elected a minor fellow at Trinity. Finally, his talents were beginning to be recognized. The next year he became a senior fellow upon taking his master of arts degree, and in 1669, before he had reached his 27th birthday, he became Lucasian Professor of Mathematics. The duties of this appointment offered Newton the opportunity to organize the results of his earlier optical researches, and in 1672 he published his work on light and color. This work established his reputation as a scientist of the first magnitude.

Newton was, however, a highly secretive and suspicious person who found it extremely difficult to submit his ideas to the scrutiny of others. Many of the great discoveries made during the two years on the farm remained unpublished for decades. It was not until 1684 that Edmond Halley finally persuaded Newton to make his work on motion and gravity known.

Halley was intensely interested in planetary orbits, and also those of comets. He and fellow scientist Robert Hooke suspected that an inverse-square relationship produced the orbits, but were not able to deduce from this hypothesis a theoretical orbit that would match the observed planetary motions. Halley traveled to Cambridge to seek the advice of Newton. What would be the orbit of a body subjected to such a force? Newton told him he had already solved the problem – the orbit would be an ellipse – but that he had mislaid his calculations to prove it.

Shortly afterwards Newton sent Halley a copy of his demonstration. Realizing the significance of what Newton had done, Halley, utilizing great skill and tact, persuaded the reluctant Newton to develop and publish his ideas on celestial mechanics. Newton's *Mathematical Principles of Natural Philosophy* (commonly known as the *Principia* from its Latin title), containing Newton's three laws of motion and his law of gravity, was published in 1687. Halley read the manuscript, corrected the proofs, and paid the publication costs out of his own pocket.

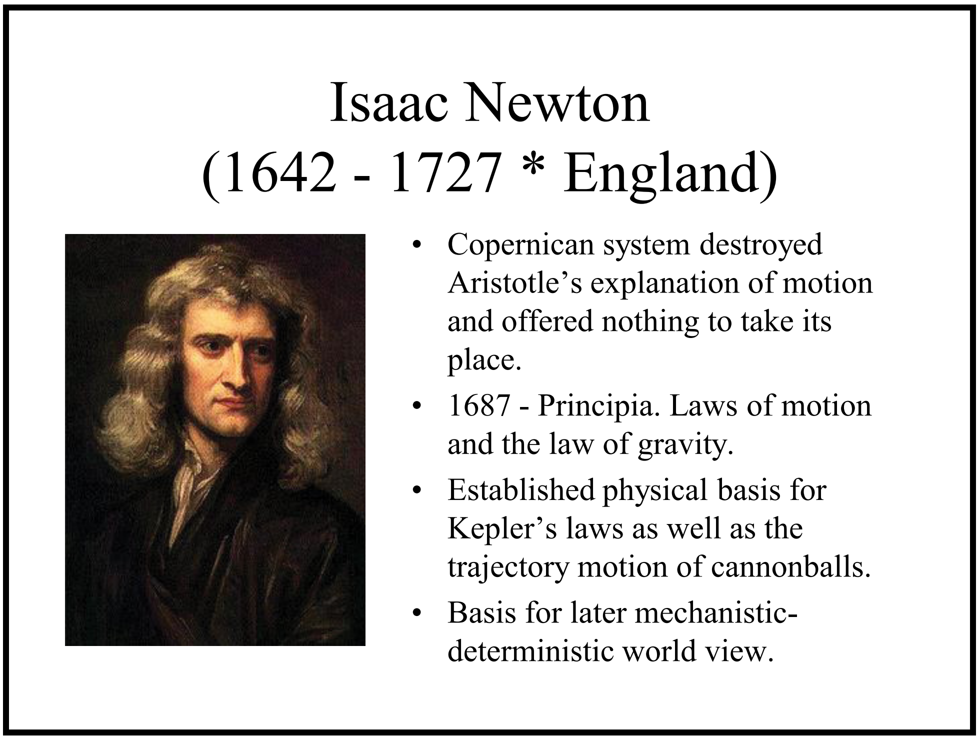


Figure 5: Isaac Newton

With the publication of the *Principia*, it seemed as if the science of mechanics was complete. The relationship between applied force and subsequent motion was now firmly established. The single cosmological force, gravity, had been completely described. Objects moved in accordance with strict natural laws that could be understood mathematically. Some continental scientists and philosophers were at first skeptical. They felt that Newton's concept of gravity as a force acting through a distance was insufficiently mechanical to be correct. Newton was also bothered by this. In a letter to a fellow scientist, he said,

> "That gravity should be innate, inherent, and essential to matter, so that one body may act on another body at a distance through a vacuum,…is to me so great an absurdity that I believe no man who has in philosophical matters a competent faculty of thinking can ever fall into it."

However the spectacular success of Newton's mathematical description of the motion of both earthly and heavenly objects soon overcame the philosophical objections and Newton was celebrated as the greatest scientist who had ever lived. We will return to the problem of action-at-a-distance in a later chapter.

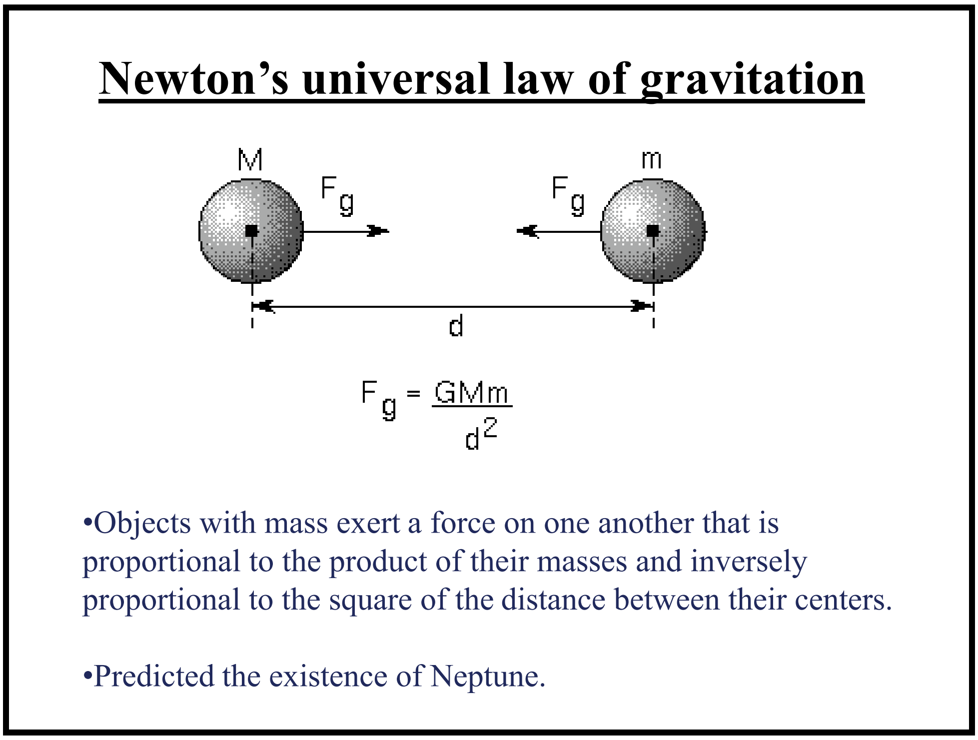


Figure 6: The law of gravity

As with many geniuses, Newton was a complex, and in many ways, highly neurotic individual. He was paranoid regarding his discoveries and often delayed publication for fear that someone would steal his ideas. He was constantly embroiled in extensive and bitter disputes with other scientists. One of the most infamous of these was with the philosopher and scientist Gottfried Leibniz over who should receive credit for the invention of the calculus. Although many of the great scientists have appreciated, and in some cases even contributed to, the arts, Newton was not one of them. He was not the least interested in music, characterized great works of sculpture as "stone dolls," and described poetry as "a kind of ingenious nonsense."

Newton was an unorthodox Christian, a monotheist who rejected the divinity of Jesus. For this reason, he constantly delayed his ordination as an Anglican minister, as was required by the various academic and official positions he held over his lifetime. Though heretical in his views, Newton was none the less highly religious. He wrote a number of tracts dealing with the literal interpretation of the Bible. Newton was a strong believer in prophetic interpretation of the Bible. He was also an astrologer and alchemist at a time when most scientists had long ago abandoned these ideas. During his lifetime he actually wrote more about the Bible and occult studies than science and mathematics.

After his death, Newton's body was discovered to contain massive amounts of mercury, probably resulting from his alchemical pursuits. Mercury poisoning could explain some of Newton's eccentricities. He was buried in Westminster Abby with the greatest of honors.

Newton viewed the material universe as consisting of atoms whose motion is determined by precise mathematical laws. Newton and virtually all of his contemporaries took the existence of a Creator as an obvious fact. However, this mechanical universe brought into question the role of the Creator with respect to the universe. Does the Creator interfere with the mechanical cause and effect from time to time? Or, did the Creator create the universe and the laws governing it and then allow the universe to evolve in accordance with those laws?

Newton believed the former. He felt that divine intervention was necessary for the creation of the solar system and was also necessary to keep it operating smoothly. Most of the scientists and philosophers that succeeded Newton rejected his theistic arguments. Newton's rival Leibniz, for example, thought that God created the universal machine, set it in motion, and then had no need to intervene further in its operation. The universe unfolded according to mathematical laws with all the precision and inevitability of a well-made clock. This religious perspective is known as Deism.

Later, the French physicist and mathematician, Pierre Simon Laplace, developed this idea further. In a famous quote, he said:

> We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.

Laplace's perspective is known as the mechanistic-deterministic worldview.

3.6 The Enlightenment

Richard Tarnas in his book *The Passion of the Western Mind*, summarizes the change in worldview resulting from the Renaissance and the Scientific Revolution in the following way:

> "And so between the fifteenth and seventeenth centuries, the West saw the emergence of a newly self-conscious and autonomous human being – curious about the world, confident in his own judgments, skeptical of orthodoxies, rebellious against authority, responsible for his own beliefs and actions, enamored of the classical past but even more committed to a greater future, proud of his humanity, conscious of his distinctness from nature, aware of his artistic powers as individual creator, assured of his intellectual capacity to comprehend and control nature, and altogether less dependent on an omnipotent God."

These changes were the foundation for the transition from the medieval worldview to the modern worldview, a worldview fully articulated during the Enlightenment.

The Enlightenment refers to a particular historic period that occurred primarily in Western Europe and the United States in roughly the 18th and 19th centuries. It is the time when the accomplishments of the Scientific Revolution were incorporated into a new worldview now called the modern worldview. Enlightenment also refers to the ideal of putting the whole of human life under the rule of reason. What is accepted as truth to be believed or as a principle to be put into practice should not be based on authority, but on reasons that are judged to be sufficient in and of themselves. The Enlightenment was also a time of increased awareness of human rights, a time when more liberal political systems were established. The founding fathers of the American Revolution were acting in the spirit of the Enlightenment.

4 Charles Darwin and Evolution

Your goals for this chapter are to know the following:

- Briefly discuss Darwin's theory of the origin of the species? What mechanism did Darwin propose to explain evolution? How does this mechanism work?
- What role did the voyage of the Beagle, and the Galapagos Islands in particular, play in the development of Darwin's theory?

The period of history known as the Scientific Revolution essentially ended with Isaac Newton and the publication of the *Principia*. But this was just a revolution in the physical sciences. The revolution in the biological sciences did not occur until 1859 with the publication of Charles Darwin's *Origin of the Species*, which proposed decent from a common ancestor as the way in which species originate and natural selection as the mechanism by which this occurs.

4.1 The Theory of Evolution

Darwin was not the first to suggest that species evolve. The idea is an ancient one going back at least to the presocratics. Anaximander, is frequently cited as a proto-evolutionist because he felt life arose from moisture and progressed over time from a fish-like ancestor to modern humans.

A more modern pre-darwinian evolutionist was Darwin's own grandfather, Erasmus Darwin. He was one of the leading intellectuals of 18th century England, a man with a remarkable array of interests and pursuits. Erasmus Darwin was a respected physician, a well known poet, philosopher, botanist, and naturalist. As a naturalist, he formulated one of the first formal theories on evolution in 1795. Although he did not come up with natural selection, he did discuss ideas that his grandson elaborated on sixty years later, such as how life evolved from a single common ancestor. He also discussed how competition and sexual selection could cause changes in species.

Erasmus Darwin also presented his evolutionary ideas in verse, in particular in the posthumously published poem *The Temple of Nature*.

> Organic life beneath the shoreless waves
> Was born and nurs'd in ocean's pearly caves;
> First forms minute, unseen by spheric glass,
> Move on the mud, or pierce the watery mass;
> These, as successive generations bloom,
> New powers acquire and larger limbs assume;
> Whence countless groups of vegetation spring,
> And breathing realms of fin and feet and wing.
>
> Erasmus Darwin. *The Temple of Nature.* 1802.

Perhaps the best known pre-darwinian theory of evolution is that of Jean-Baptiste Lamarck who was the first to develop a truly coherent evolutionary theory. He published his ideas in the early 1800s. Lamarck believed that the environment gives rise to changes in animals and that under certain conditions, these changes could be passed on to offspring. This idea is generally referred to as 'the inheritance of acquired characteristics.' In Lamarck's view, as the lower branches of trees became depleted of leaves, the leaf eating ancestors of giraffes would have to stretch their necks to reach leaves higher in the trees. This would strengthen and gradually lengthen their necks. These giraffes would then have offspring with slightly longer necks.

Darwin was aware of these evolutionary ideas. He read his grandfather's books and admired them, not so much for the content as for the fact that they were notorious. He was also familiar with Lamarck. At Edinburgh, one of his teachers, Robert Grant, befriended Darwin, possibly because, as an ardent evolutionist, he was an admirer of Erasmus Darwin. One day while they were walking, Grant spoke with enthusiasm of Lamarck's theory. Later Darwin recalled, "I listened in silent astonishment, and as far as I can judge, without any effect on my mind."

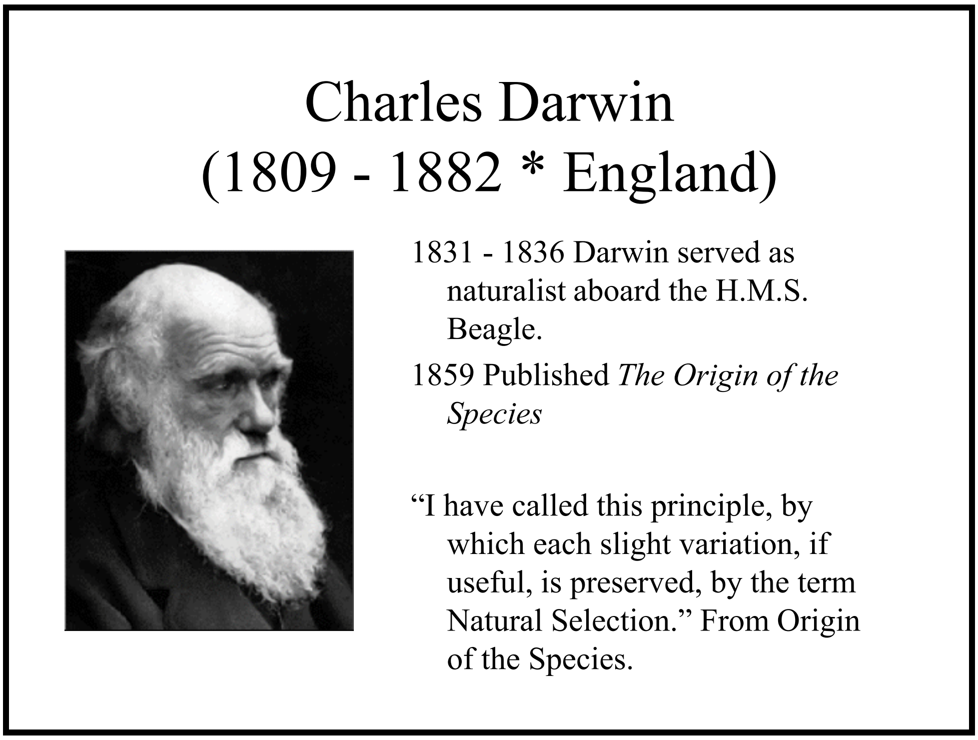


Figure 1 Charles Darwin

In 1831, Charles Darwin, an excited 22-year old adventurer, climbed aboard the *H.M.S. Beagle* for a 5-year voyage. He left with rather conventional views on the origins of species, not having been convinced by either his grandfather's books or his knowledge of Larmarckism. Little did he realize that this voyage would set him upon a collision course with the prevailing view of human nature.

Darwin served as naturalist on the *Beagle* and as companion to the captain, Robert Fitzroy. The *Beagle's* objective was to chart the coastline of South America. As the ship circled the continent, Darwin's curiosity was piqued by the changes he noted in the flora and fauna. He became particularly fascinated by the variety of animal species on the Galapagos Islands, an isolated chain off the coast of Ecuador. The islands contained many unique species, but species that resembled those on the nearby mainland.

These observations caused Darwin to question the prevailing belief that all species were created distinct and immutable in a single creation event. Would all of these similar, but distinct, species have been designed at one time only to be selectively deposited on these tiny, remote islands? Darwin did not think so.

The outline for an alternative scenario came to him. Darwin recognized the Galapagos as newly formed volcanic islands, and concluded that as plants and animals from the nearest landmass colonized a particular island, they became isolated from their parent population and evolved to fit the island's environmental niches. The differences in environment from island to island resulted in similar, but distinct, species. The variety of finch species is perhaps the most striking example. Maybe, over the generations, life forms adapted their structure and functions to the survival demands of their particular environments. Darwin was not sure how this process of adaption took place, but he returned from his arduous voyage convinced that it did.

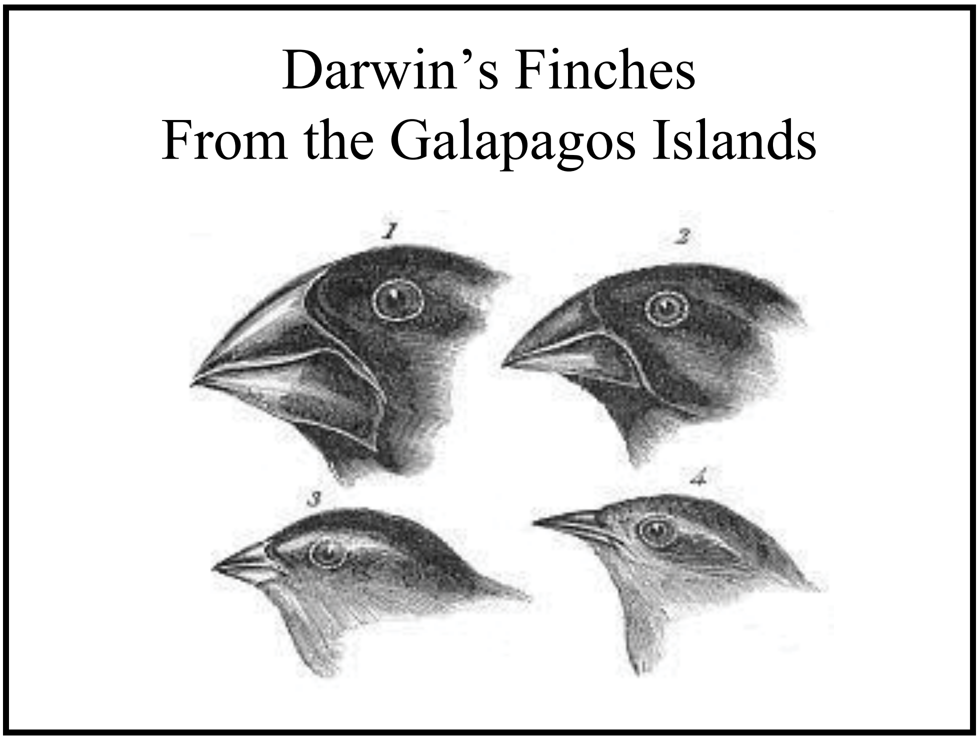


Figure 2 Darwin's Finches

Darwin's thinking on the process of adaption was clarified by reading Thomas Malthus' theory that the expanding world population would eventually exceed its food supply. Darwin saw that a similar proliferation of thriving organisms in the natural realm might provide the mechanism for evolution. As organisms thrive and multiply within a given environment, they eventually outstrip the ability of the environs to meet their needs. At this point, competition begins. Those individuals particularly suited to the habitat survive and have offspring who share and pass on the successful traits of their parents. Less successful individuals die off without reproducing. Thus, by this process of natural selection, they become adapted to their environment. They change as a species in the direction most suited for survival and reproduction.

Darwin's convictions along these lines were bolstered by one of his hobbies. He raised pigeons and had seen the effects of breeding based on selected attributes. Why couldn't a natural environment operate like a breeder, allowing the survival and proliferation of organisms with "desirable" traits?

To return to the example of giraffes, as with any characteristic of a species, there would be variations in neck length within the population of the leaf eating ancestors of giraffes. As the lower branches of trees became depleted of leaves, natural selection would favor those with the longest necks. In contrast to Lamarck, it is not the acquired characteristics, but the innate ones that result in evolution.

Darwin resisted publication of his theory of evolution for 20 years. Finally, in 1859, fearful of losing his claim to the discovery, Darwin published. His fear was caused by a letter he received in June of 1858, from an acquaintance, Alfred Wallace, a fellow naturalist. Like others he had received from Wallace, it came from the Malay Archipelago. This envelop was bulkier than others, containing a manuscript as well as a letter. The cover letter said, "Here is a hypothesis I've hit upon to explain the origin of species." Wallace had independently come up with the concept of evolution by natural selection. (Although Darwin considered the ideas in Wallace's paper to be essentially the same as his own, there was an important difference that in his panic Darwin failed to notice. Darwin saw natural selection in terms of competition between individuals of the same species, whereas Wallace saw natural selection in terms of species in competition with one another.)

The letter from Wallace requested that Darwin forward the manuscript to Charles Lyell, the renowned geologist. Darwin dutifully did so, but with a heavy heart, assuming that he had lost his chance to claim discovery. Lyell advised him to calm down. Maybe something could be salvaged.

Although Darwin had never published his theory, he had sent summaries of his ideas to several other naturalist he knew to be supportive. Among them were Lyell, and Joseph Hooker, a botanist. Lyell and Hooker, were sympathetic to Darwin's predicament, and devised an arrangement that would recognize Wallace's contribution, but would give, justly, Darwin priority for the discovery. They arranged and sponsored a joint presentation of Wallace's manuscript and Darwin's unpublished work at the next meeting of the Linnean Society, one of London's better scientific associations. The arrangement was made without Wallace's permission or even knowledge.

Darwin reluctantly consented, worrying whether it was an honorable thing to do. He sent Hooker his 1844 essay and the six-paragraph summary he'd sent the previous year to the American botanist Asa Gray. The few weeks it took for all this to transpire, were an incredibly difficult time for Darwin. He felt trapped between the demands of honor on behalf of Wallace and his own self-interest. Worse of all, this was taking place while his youngest child, Charles, lay dying of scarlet fever

On the evening of July 1, 1858, the Darwin-Wallace material was read to an audience of about 30 people. Neither author was in attendance. Wallace was in New Guinea, unaware of the event in London. Darwin was at home with his wife mourning the death of their 19-month-old son just three days earlier.

Darwin published *The Origin of Species* in 1859. In this book, through copious examples, he established the strong inference that species evolve one from another. He also outlined the concept of natural selection, the mechanism by which he thought evolution occurred. Dealing only with plants and nonhuman animals, Darwin did not directly address the question of human origins. In fact, the book's only mention of human evolution was the prediction that "light will be thrown on the origin of man and his history."

However, the implications of the theory for human origins were clear. Darwin met with the feared reactions. When she heard of Darwin's work in 1860, the wife of the Bishop of Dorcester expressed her shock this way: "Descended from apes! Let us hope it is not true, but if it is, let us pray that it will not become generally known."

Darwin died on 19 April 1882. By the time of his death, he was one of the most respected scientists in the world, so much so that it was decided that, although he was a well-known agnostic, he should be buried in Westminster Abbey, along with Isaac Newton and other great British scientists. He was buried to music commissioned specifically for the funeral. The anthem's composer drew from the Book of Proverbs to write '*Happy is the man that findith wisdom and getteth understanding*.'

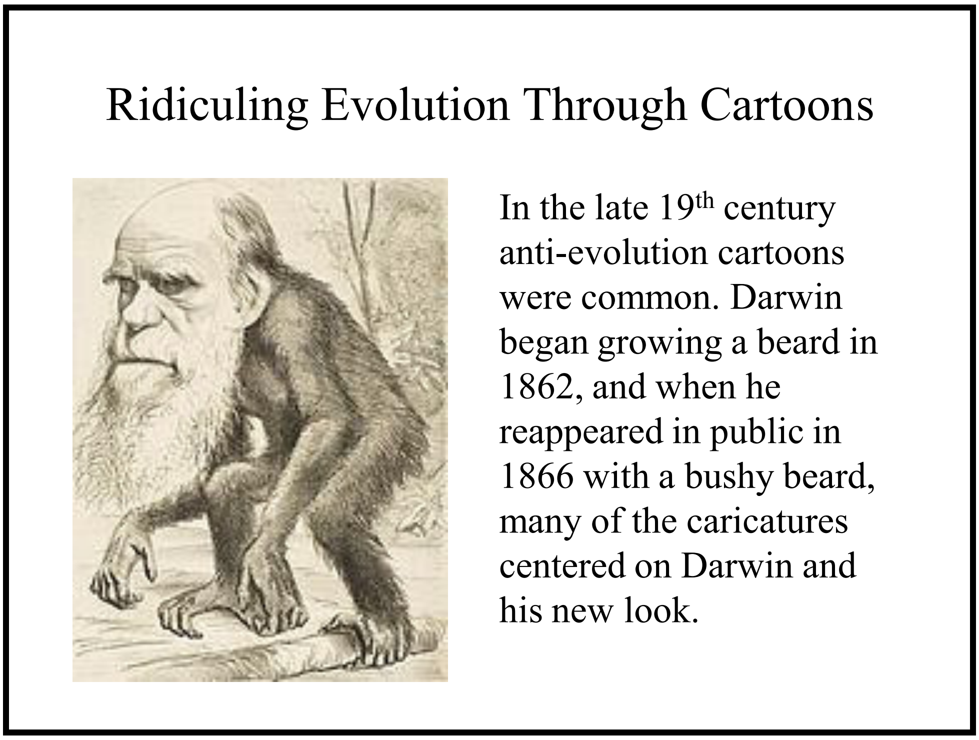


Figure 3 Darwin cartoon

Natural selection has sometimes been referred to as 'survival of the fittest.' This can be a little misleading if fittest is thought of as strongest. Actually, fittest refers to the ability to reproduce. Thus a part of natural selection is what Darwin called sexual selection. Many traits that are sexually selected for can actually hinder survival of the organism, but this can be compensated for if they increase reproductive opportunities. Consider the classic example of the peacock's tail. It is metabolically costly, cumbersome, and essentially a 'predator magnet.' Reproductively, this is more than compensated for by the ability of the tail to attract mates.

4.2 Fossils

Steven Stanley, an American paleontologist and evolutionary biologist, has pointed out that "it is doubtful whether, in the absence of fossils, the idea of evolution would represent anything more than an outrageous hypothesis." Yet we do have fossils and our ability to date them and their surrounding geological layers has become more and more precise. Though the mechanisms for evolutionary change are still debated, such gradual changes themselves are extremely well documented by the fossil record.

The word 'fossil' comes from the Latin word meaning 'having been dug up.' They are the remains of all or parts of animals and plants from the remote past. Fossils are formed in a number of different ways, but most are formed when a plant or animal dies in a watery environment and is buried in mud and silt. Soft tissues quickly decompose leaving the hard bones or shells behind. Over time sediment builds over the top and hardens into rock. As the encased bones decay, minerals seep in replacing the organic material cell by cell in a process called 'petrification.' Alternatively the bones may completely decay leaving a cast of the organism. The void left behind may then fill with minerals making a stone replica of the organism.

A fossil normally preserves only a portion of the deceased organism, such as bones, teeth, or exoskeletons. Preservation of soft tissue is rare. Fossils may also consist of the marks left behind by the organism while it was alive, such as footprints. Some fossils are very striking and have been collected at least as far back as recorded history.

By the 19th century, it was recognized that certain types of fossils were associated with certain geological strata. Thus rocks from distant locations could be correlated based on the fossils they contained.

Various explanations have been put forth throughout history to explain what fossils are and how they came to be where they were found. Many of these explanations relied on folktales or mythologies. In China the fossil bones of ancient mammals including *Homo erectus* were often mistaken for "dragon bones" and used as medicine and aphrodisiacs. In the West the presence of fossilized sea creatures high up on mountainsides was seen as proof of the biblical deluge.

Aristotle realized that fossil seashells from rocks were similar to those found on the beach, indicating the fossils were once living animals. Leonardo Da Vinci also recognized them as once living animals and commented on their interpretation, including an argument that they could not have resulted from the biblical flood. At the time Darwin wrote *The Origin of Species*, the fossil record was sparse, and Darwin pretty much discounted its significance as evidence of biological evolution. In *The Origin of Species*, Darwin constantly refers to the geological record as 'imperfect.' However, that is no longer the case.

Toward the end of the 19^{th} century, and especially in the 20^{th} century there was an almost exponential increase in the fossil record. Also, we now have the ability to accurately date most fossils. The fossil record now constitutes strong evidence in favor of biological evolution.

One of the most important features of a good scientific theory is its ability to make testable predictions. The fossil record allows this. One example involves marsupials. Today most marsupials are found in Australia. But 40 million years ago, early marsupials were located in the southern part of South America. The earliest marsupials found in Australia date from about 30 million years ago. How could they have gotten from South America to Australia?

Forty million years ago, South America and Australia were connected by what is now Antarctica in the southern supercontinent, Gondwana. Since marsupials had to go overland from South America to Australia, they must have passed through Antarctica. Thus it was predicted that there would be fossil marsupials in Antarctica. The prediction was later confirmed by the discovery there of more than a dozen marsupial fossil species dating between 35 and 40 million years ago.

There have been many similar predictions based on the theory of evolution. All have turned out to be correct. The evidence for evolution is overwhelming. It is often claimed that the theory of evolution is not scientific because it cannot be falsified. This is not true. The genuine falsification of any prediction based on the theory would falsify the theory. For example the prediction that the morphological changes that lead from the most primitive species of a particular type of organism to the present must occur in chronological order. A true counterexample would falsify the theory.

The claim that evolution cannot be falsified is usually based on the fact that evolutionary theory cannot predict the future evolution of a species. The factors causing evolution are extremely complex and are not completely understood. This precludes the ability to test the theory by observing its results, and therefore evolution cannot be falsified in this particular way.

Another common claim is that we have never directly observed speciation. For the most part this is true, because speciation generally involves time spans greater than human history. However, it is, in fact, not true. Bacteria can reproduce in as little as twenty minutes. In laboratory experiments lasting decades and extending over tens of thousands of generations, the origin of new species by natural selection from a common ancestor has been observed.

It is unfortunate that the word 'theory' has two very distinct meanings. An example of the first is the heliocentric theory of the solar system or the theory of relativity. In this sense, theory refers to a well established, coherent explanation of a vast body of data. In the second meaning, theory is just another word for hypothesis or guess. With the tremendous increase in the fossil record, the ability to accurately date fossils, and, perhaps most importantly, genetic evidence, biological evolution is now a theory in the first sense of the word.

4.3 Punctuated Equilibrium

When Darwin proposed his theory on the origin of species, he believed that evolution took place as a result of the accumulation of small variations over long periods of time – that species arise gradually. He did not assume that the pace of change was constant, but did believe that it was slow and continuous. This view of evolution is called gradualism. However, if evolution is gradual, there should be a fossil record of small, incremental changes along the way. But the fossil record at the time showed no indication of this. Darwin was concerned, but assumed that the absence of intermediate forms was due to the incompleteness of the fossil record. He assumed that as the fossil record became more complete, the transitional forms would be there.

With a few well established exceptions, this has not turned out to be the case. The fossil record clearly shows that most species are stable, remaining essentially unchanged over millions of years. In 1972, Steven Jay Gould and Niles Eldredge proposed an alternative theory to explain this. Species are in generally stable, but the stability can be 'punctuated' by a relatively rapid burst of change that results in a new species and that leaves few transitional fossils behind. This concept is known as punctuated equilibrium.

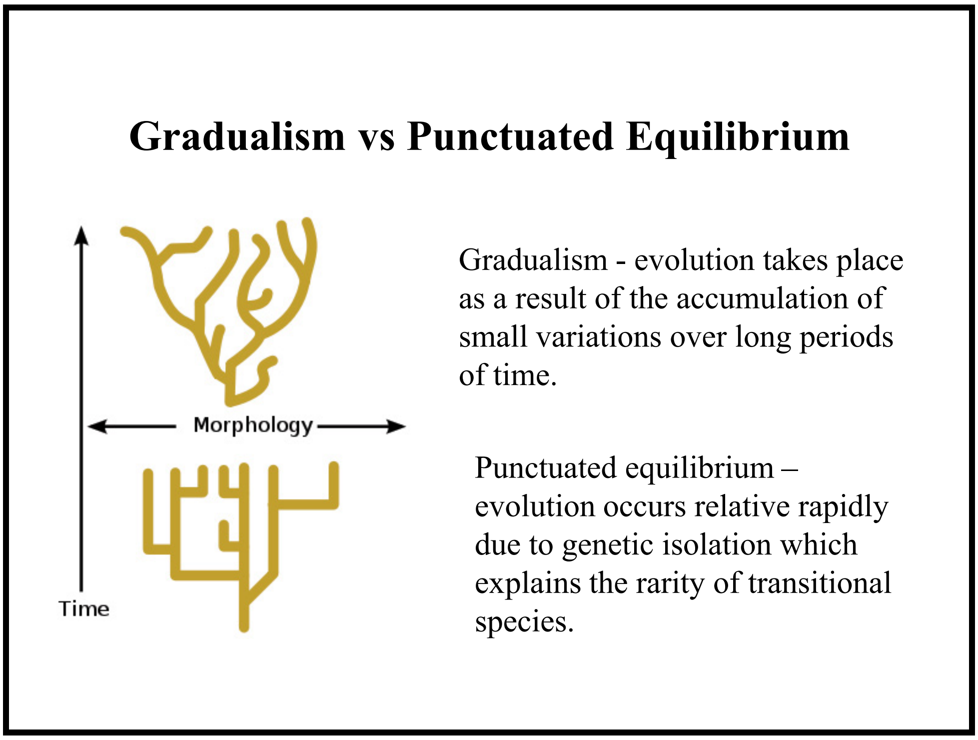


Figure 4 Punctuated Equilibrium

As opposed to gradualism, changes leading to a new species don't occur in the mainstream population where interbreeding would tend to blend them out. Rather speciation tends to occur when a small group becomes physically and genetically isolated from the mainstream population. In the small group, any changes that would create a survival advantage in the new environment would not be blended out by interbreeding with the mainstream population. Thus, in a relatively short time, maybe only thousands of years, a new species will evolve. The relative importance of punctuated and gradual patterns of evolution has not yet been determined. Research to address this question is ongoing.

4.4 The Modern Synthesis

In addition to punctuated equilibrium, Darwin's original ideas on evolution have been modified in another way. The Austrian biologist and monk Gregor Mendel was a contemporary of Darwin. Mendel's meticulous experiments cross-breeding pea plants indicated a previously unknown mechanism for heredity. Mendel's paper, *Experiments on Plant Hybridization,* published in 1865, was criticized, but mostly ignored. Darwin was unaware of it. Darwin struggled with the problem of how traits were inherited and was never able to come up with a mechanism whereby organisms pass traits on to their offspring. Had he been aware of Mendel's work, he would have known.

Mendel had read the abbey's German translation of *Origin* and was in general supportive of Darwin's ideas. A minor mystery is why Mendel apparently never contacted Darwin himself. Unfortunately, when Mendel died in 1884 the next Abbot burned all his papers to end disputes over taxation, so we do not know if Mendel ever considered writing to Darwin.

Mendel's work was rediscovered in 1900, sixteen years after his death. It was then realized that his experiments provided the description of inheritance that Darwin's theory lacked. The extreme importance of Mendel's work was recognized and he finally received the credit for one of the great discoveries in the history of science.

The current theory of evolution, the modern evolutionary synthesis (or neo-darwinism), explains that the evolution of species occurs through a combination of Darwin's mechanism of natural selection and Gregor Mendel's theory of genetics. This synthesis, was produced between 1936 and 1947. The previous development of population genetics, between 1918 and 1932, was a stimulus, as it showed that Mendelian genetics was consistent with natural selection and gradual evolution. The neo-darwinian synthesis is the current paradigm in evolutionary biology.

> There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being evolved.

The last words of *The Origin of Species*, 1st Edition, Charles Darwin (1859)

5 Matter and Light

Your goals for this chapter are to know the following:

- What is meant by wave-particle dualism?
- List the major divisions of the electromagnetic spectrum in order of decreasing wavelength. Do the same for the visible portion of the spectrum.
- How is each of the three types of spectra–continuous, bright-line, and dark-line–produced?
- What type of spectrum is associated with our sun and other stars?
- Why do different elements have different patterns of lines in their spectra?

The word 'universe' is normally taken to mean everything that exists. The most obvious components of the universe are mater and light. By the end of the 19th century, is was known that light is simply one form of electromagnetic radiation that also includes radio waves, infrared and ultraviolet light, X-rays, and gamma rays. The matter that we are most familiar with consists of the chemical elements in the form of atoms composed of protons, neutrons, and electrons. Surprisingly, the chemical elements make up just a small part of the matter of the universe, most of which is in the form of a mysterious dark matter. More of that later.

Physical events are interactions between matter and energy, one form of which is electromagnetic radiation. These events unfold within the arena of space-time

5.1 Atoms

Empedocles was a Greek philosopher who sometime around 450 BCE introduced the concept that all matter is made up, in differing proportions, of four elemental substances, earth, air, fire and water, and that the ratios of these affect the properties of the matter. Empedocles' theory was an important development in scientific thinking because it was the first to suggest that some substances that looked like pure materials, like stone, were actually made up of a combination of different elements.

Democritus, some 30 years later, introduced the concept of atoms. Democritus taught that "nothing exists except atoms and the void: all else is mere opinion." He conceived of atoms as indivisible and eternal, and in fact the word atom is derived from the Greek word meaning "that which cannot be divided." The atoms of Democritus were small, discrete, and identical in composition, though they might differ in size and shape with the differences in size and shape determining the specific properties of the substance. According to Democritus, perceived changes in the world were produced by changes in the groupings of atoms.

Aristotle accepted the theory of Empedocles but rejected that of Democritus. He argued that logic ruled out the concept of discrete atoms. The atoms of Democritus had extension in space and were identical in composition, but these properties are not compatible with indivisibility: extension in space implies divisibility and composition implies yet smaller parts. Aristotle taught that matter is continuous rather than discrete in nature and that "everything continuous is divisible into divisibles that are infinitely divisible."

Despite philosophical difficulties, the concept of atoms was too persuasive to be dismissed and by the 17th century most scientists, including Galileo and Newton, were atomists. In the early 19th century, the English chemist John Dalton used the concept to account for chemical reactions, placing atomic theory in a scientifically respectable context for the first time. By the end of the 19th century, the concept of the atom was well established, though some influential scientists at the time did not accept the atom as a real constituent of nature. It was not until the early 20th century that an experiment suggested by Albert Einstein proved the existence of atoms.

Our current atomic theory is based on quantum mechanics and is a nonvisualizable model. Our best visualizable model is the planetary model with a tiny nucleus of protons and neutrons surrounded at a relatively great distance by electrons. The nucleus, though only a tiny fraction of the size of the atom, contains almost all of the mass of the atom. The atom is overwhelmingly empty space. Ernest Rutherford, who suggest this model in 1911, referred to the nucleus as "a fly in a cathedral," a dimensionally correct analogy.

Protons are positively charged particles, and, as the name implies, neutrons are neutral. These two types of particles have a tendency to bind together under the influence of the strong nuclear force. They have comparable masses, with the neutron being slightly more massive than the proton. The electron is a negatively charged particle, only about one-two-thousandth the mass of the proton. It is not affected by the strong nuclear force, but is attracted to the positively charged nucleus by the much weaker electromagnetic force. The combination of the strong nuclear and the electromagnetic force holds atoms together.

5.2 The Chemical Elements

A chemical element is any substance that cannot be broken down into simpler materials by chemical means. Hydrogen, oxygen, and iron are examples of common elements. An atom is the smallest possible quantity of an element. Such substances as salt and water, on the other hand, can be broken down chemically, and are called compounds. Compounds are composed of molecules, which are chemical combinations of atoms. For example, a water molecule is composed to two hydrogen atoms and one oxygen atom, and written H_2O.

The nucleus of an atom is composed of neutrons and protons. The composition of the nucleus is described by an atomic number, Z, defined as the number of protons in the nucleus, and an atomic mass number, A, defined as the number of neutrons plus protons in the nucleus. If the atom is neutral, then there must be Z electrons in orbit around the nucleus to balance the positive charge of the nucleus. In many ways, the neutron and the proton behave as if they are two different states of the same particle. For this reason, the word nucleon is used to include both particles. Thus, the atomic mass number is the number of nucleons in the nucleus.

The notation used to specify a particular nucleus is ${}^{A}{}_{Z}X$, where X represents the chemical symbol of the element. For example, the helium nucleus consisting or 2 protons and 2 neutrons is represented as ${}^{4}{}_{2}He$.

Each chemical element has a different number of protons in its nucleus. The number of protons determines the number of orbital electrons, which in turn determines the chemical properties of the element. The periodic table of the elements is arranged in order of increasing numbers of protons in the nucleus.

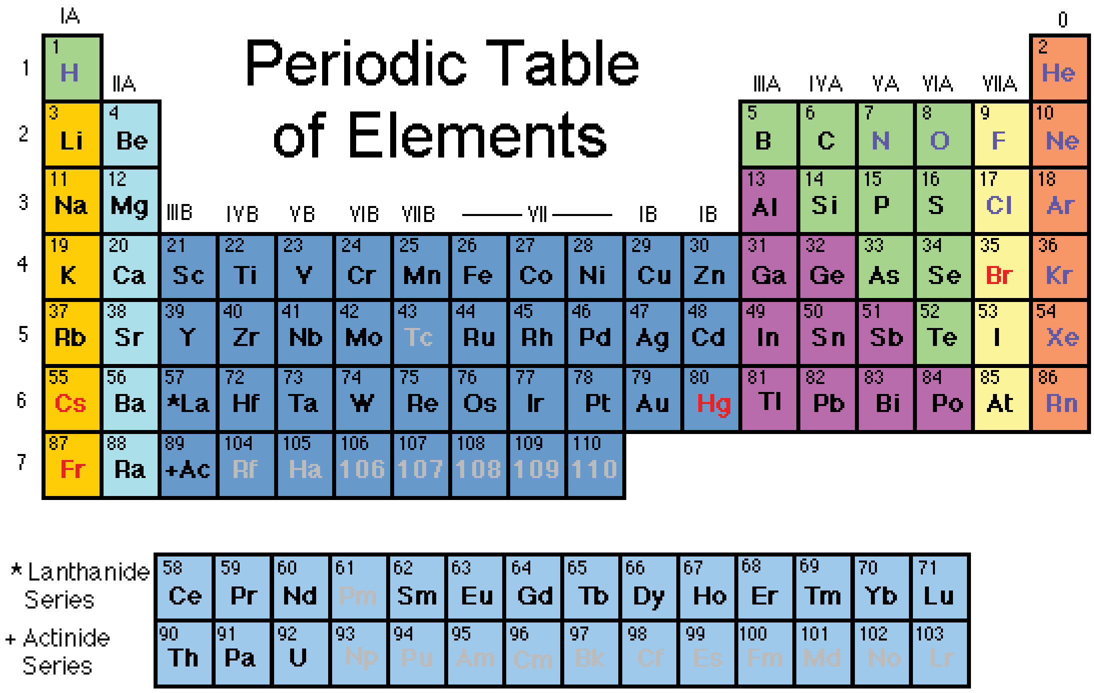


Figure 1 The Periodic Table

In 1913, the Danish physicist Niels Bohr modified the planetary model by suggesting that the electron orbits of the planetary model can have only certain allowed radii. That is, an electron can orbit the nucleus at certain allowed distances, but all radii between these allowed distances are forbidden. Each of the allowed orbits corresponded to a specific energy of the atom. Thus the energy of an atom, like the radius of the electron's orbit, is restricted. All intermediate energies are forbidden. The Bohr model of the atom was an extremely radical suggestion at the time. Not even Bohr himself could give any theoretical reason why the electron orbits would be restricted in this way. His primary justification of this assumption was simply that it seemed to fit observed patterns in the properties of atoms, namely that each type of atom emitted or absorbed only certain energies of electromagnetic radiation.

5.3 Radioactive Decay

Some nuclei are inherently unstable. Their ratio of protons to neutrons is incorrect for the mass of the nucleus. In general, the heavier the nucleus is, the more neutrons relative to the number of protons are required for stability. For instance, a helium nucleus can be stable with two protons and two neutrons, but a lead nucleus with 82 protons requires something on the order of 126 neutrons to be stable. Nuclei with the wrong ratio for stability will respond either by ejecting a helium nucleus (alpha decay), changing a neutron into a proton (beta negative decay), or changing a proton into a neutron (beta positive decay). Sometimes a nucleus can have an appropriate neutron to proton ratio, but too much energy. In this case, the excess energy can be emitted in the form of electromagnetic radiation (gamma decay).

Radioactive decay occurs on a time scale that depends on the particular nucleus – some almost immediately, others taking billions of years. The rate at which a particular radioactive nucleus decays is given by the half-life, the time it takes one-half of the nuclei to decay. For example, carbon-14, a nucleus with six protons and eight neutrons, decays to nitrogen-14, a nucleus with seven protons and seven neutrons. Its half-life is 5730 years. That is, in each interval of 5730 years, half the remaining carbon-14 nuclei decay. Thus after 11,460 years (two half-lives) only one-quarter of the original number of radioactive nuclei remain. After three half-lives, only one-eighth would be left, and so on. Mathematically, this is called exponential decay.

Half lives make it possible for radioactive substances to be used for dating. For example potassium-40 decays to argon-40 with a half-life of 1.3 billion years. Argon-40 is a gas, so if a rock is melted, the argon is released into the atmosphere. Once it has re-solidified, any additional argon produced by radioactive decay will be trapped in the rock. In this way, the ratio of argon to potassium-40 is a direct measure of the time since the rock has solidified. For instance, if the ratio of argon to potassium-40 is exactly 3 to 1, then only one-quarter of the original potassium-40 remains and the rock must be exactly two half-lives, or 2.6 billion years, old. If the ratio is 3.7 to 1, the equation of exponential decay gives an age of 3.2 billion years. This technique applied to solar system rocks indicates that the oldest ones (meteorite fragments of asteroids) are about 4.6 billion years old, and hence the age of the solar system is about 4.6 billion years.

5.4 Electromagnetic Radiation

Essentially everything we know about the content of the universe, and certainly about objects beyond our solar system, is based on deciphering the information contained in the electromagnetic radiation emitted by those objects. We know of the stars in the sky because we see them. As more powerful optical telescopes were built, we discovered additional kinds of objects in the sky. Finally, in the last half of the 20th century, we discovered still more of the contents of the universe through the use of telescopes that could "see" in other regions of the electromagnetic spectrum.

Questions about the nature of electromagnetic radiation began at least as far back as the ancient Greeks with their ideas about the nature of visible light. Various models were suggested, but it wasn't until the second half of the 17th century that the nature of light began to be formally debated among scientists. The one thing that was clear was that light could transfer energy from one place to another. For example, sunlight can heat water. Two schools of thought arose as to how this was accomplished.

Robert Hooke and Christian Huygens, contemporaries of Newton, proposed a wave model. According to this view, light was like sound. When one speaks, the vocal cords cause the air in the throat to oscillate, which causes the air near it to oscillate, and so on until the air in the vicinity of the listener's eardrum oscillates. Each particle in the medium merely oscillates about some fixed position; there is no net transfer of air, just energy.

Isaac Newton objected to this model. He preferred the idea of particles moving through space from one material object to another. When one throws a baseball at a target, energy is transferred. In this same manner, Newton viewed light as a stream of tiny particles moving at high speeds. Like bullets fired from a gun.

At the time, there was no experimental way to distinguish clearly between the wave model and the particle model. This began to change in the early part of the 19th century when evidence began to accumulate in favor of the wave model. The apparent *coup de grace* to the particle theory was delivered in 1865 when James Clerk Maxwell, an English theoretical physicist, proposed his theory of electromagnetism. The theory suggested that electromagnetic disturbances could be propagated through space as waves. The theoretical speed of the waves was exactly that of the speed of light, which had been accurately measured some years earlier. The conclusion seemed unmistakable: light is not only a wave, it is an electromagnetic wave.

Maxwell's theory also made it clear that visible light was just part of an electromagnetic spectrum; that electromagnetic waves with wavelengths both longer and shorter than visible light should exist. Heinrich Hertz, the discoverer of one of these, radio waves, said in 1889,

> We know that light is a wave motion. We know the speed of the waves, we know their length. In a word, we know completely the geometric relationships of this motion. These things no longer permit of any doubt, and a refutation of this view is unthinkable to the physicist. In so far as human beings can know the truth, the wave theory is a certainty.

Perhaps unthinkable to most, but not to Albert Einstein. In 1905, the same year he published his theory of relativity, his paper suggesting an experiment to prove the existence of atoms, and a paper on thermodynamics that earned him a Ph.D. from the University of Zurich, the 23-year old Einstein, published a paper suggesting that in some situations electromagnetic radiation would behave as if it were a stream of particles. He even suggested an experiment that would show this to be the case.

According to Einstein, electromagnetic radiation interacts with matter as discrete packets of energy, later called photons. The energy of a specific photon is associated with the wavelength of the radiation. The longer wavelengths correspond to lower photon energies and the shorter wavelengths correspond to higher photon energies.

The experimental evidence soon indicated that Einstein was right. Electromagnetic radiation did, in some instances, behave as if it were a stream of particles. But the evidence of the 19th century that it behaved in other instances as if it were a wave could not be ignored. However the two models are mutually exclusive. There is no way to combine them into a single coherent model.

The resolution of this dilemma came by abandoning the effort to force electromagnetic radiation into one or the other model. Some phenomena have no analog in the world of everyday sensory experience. In that world, a clear distinction exists between waves and particles. Something is either a wave or a particle, but not both. However, at the submicroscopic level, the distinction no longer exists. Electromagnetic radiation is indisputably both wave-like and particle-like.

This conclusion is referred to a wave-particle duality. Because there is nothing in everyday experience that exhibits this dualism, it is impossible to visualize the true nature of electromagnetic radiation. Thinking about a phenomenon for which there is no concrete analog can be an unsettling experience. As Einstein put it, “It is only with reluctance that one’s desire for knowledge endures a dualism of this kind.”

Scientists who must cope with this dilemma in their work take a pragmatic approach. They make use of one or the other model, depending on the situation. In some situations, it is more appropriate to treat radiant energy as a wave. To explain how electromagnetic radiation travels through space, the wave model is necessary. To explain how electromagnetic radiation delivers energy to matter, the particle model is necessary. Although this is a workable approach, one should not confuse these useful models with reality itself. Electromagnetic radiation is more complex than depicted by either the wave or the particle model.

5.5 The Electromagnetic Spectrum

By the first decade of the 20th century, physicists had discovered electromagnetic radiation in a wide range of wavelengths. At one end of the spectrum are radio waves of long wavelength and low photon energy. At the other end are gamma rays of short wavelength and high photon energy. All types of electromagnetic radiation travel through space at the speed of light.

Electromagnetic radiation can be produced in a number of different ways. The various regions of the electromagnetic spectrum are usually classified according to the way in which the radiation is produced. Radio waves are produced by oscillating electric currents. This part of the spectrum is subdivided according to the primary use of the radiation. Infrared, visible, and ultraviolet radiation is mostly produced by transitions of the outermost electrons of atoms as the atoms move from a higher to a lower energy state. X-rays are produced by the slowing of very high-speed electrons or by transitions of the inner orbital electrons of high-mass atoms. Gamma rays are emitted by the nuclei of atoms.

The distinction between infrared (below red), visible light, and ultraviolet (beyond violet) is based entirely on how the human eye responds to the radiation. Infrared radiation has wavelengths that are too long (or photons with too little energy) to excite the receptor cells in our eyes. Ultraviolet radiation has wavelengths that are too short (or photons with too much energy). The range of wavelengths within visible light correspond the perceived colors. The longest wavelength visible light is red, next orange, yellow, green, blue, and finally, violet with the shortest wavelength that can produce a visual response.

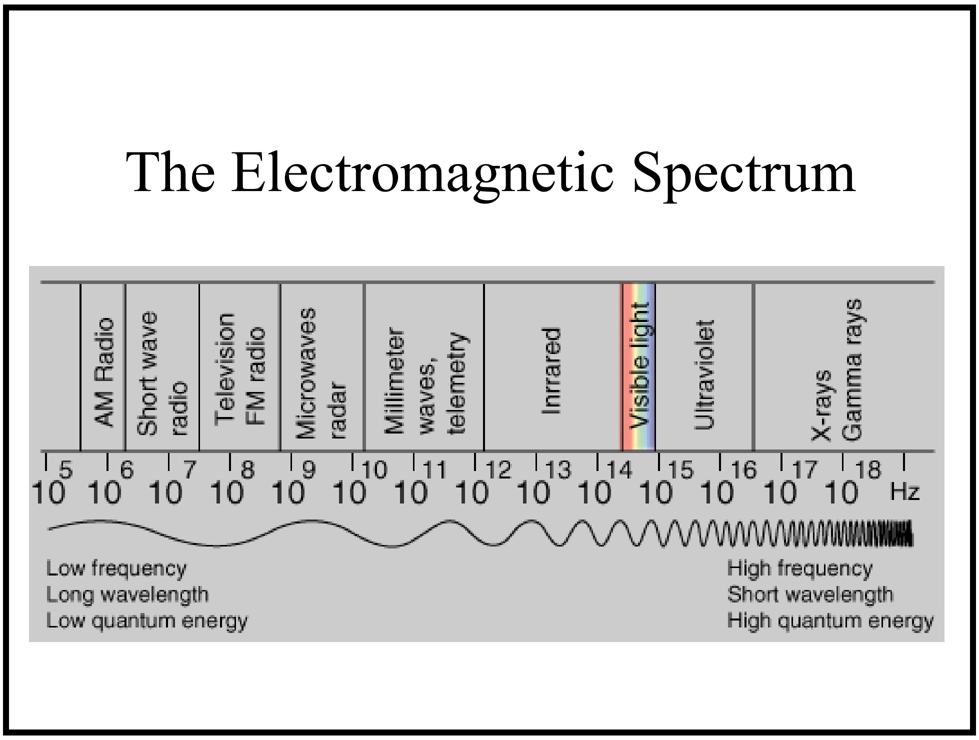


Figure 2 Electromagnetic Spectrum

5.6 Spectroscopy

Spectroscopy refers to the study of electromagnetic radiation based on the distribution of energy with respect to wavelength. A spectrometer is any device that will take incoming radiation and separate it on the basis of wavelength. A prism is an example. If sunlight falls on a prism, the light emerging from the other side will appear as a rainbow. The longer wavelength red light will come out at a slightly different angle than the shorter wavelength violet light. When electromagnetic radiation from various sources is studied using a spectrometer, three distinct types of spectra are found. If all wavelengths within a certain range are present, the spectrum is a continuous spectrum. When only certain wavelengths are present, the spectrum is a bright-line spectrum. Finally, a dark-line spectrum is one in which almost all of the wavelengths are present, but certain ones are missing.

A continuous spectrum is produced by a solid, a liquid, or a high-density gas. The total energy emitted and the distribution of this energy with respect to wavelength, are determined by the temperature of the emitting object. The spectrum is independent of the chemical composition of the object. An example of this is the radiation emitted by an incandescent light bulb.

A bright-line spectrum is emitted by a hot, low-density gas. Unlike the continuous spectrum, it is completely determined by the chemical composition of the gas. Each chemical has its own distinctive bright-line spectrum. The spectrum identifies the element in the same way a fingerprint identifies an individual. The explanation for this is the Bohr model. In this model each type of atom has its own distinct set of allowed energy levels, so only certain photon energies can be emitted as the atom moves from a higher energy state to a lower state. A bright-line spectrum is also known as an emission spectrum. Examples of this type of spectrum include neon lights and the mercury vapor lamps that are commonly used to illuminate parking lots.

A dark-line spectrum is produced when a continuous spectrum passes through a low-density gas. The wavelengths that are missing are, like the bright-line spectrum, and for the same reason, completely determined by the chemical composition of the low-density gas. In this case, however, the photons are being absorbed by the gas as the atoms move from a lower energy state to a higher state. An example of a dark-line spectrum is the radiation emitted by our sun and by other stars. At a level near the surface of our sun, a continuous spectrum is emitted. However, before exiting the surface, the radiation must pass through the sun's low density atmosphere. There wavelengths are absorbed by the various chemical elements that make up our sun's composition.

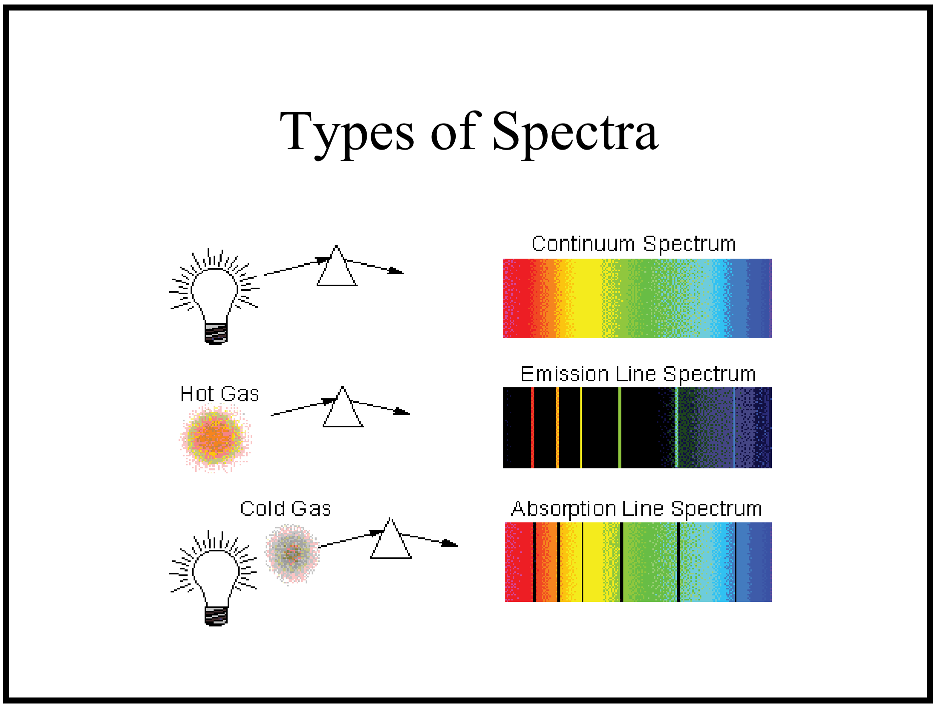


Figure 3 Types of Spectra

In the 1830s the positivist philosopher Auguste Comte was asked if there was anything science would never be able to determine. He gave what is almost certainly the right answer, yes, but then proceeded to give a wrong example. He famously said we could never know the chemical composition of the stars. Through an analysis of their absorption spectra, we now know their composition quite well.

6 The Present Universe

Your goals for this chapter are to know the following:

- Describe the Milky Way Galaxy? What are the four main components of our galaxy? Where is our Solar System located within the Galaxy? Draw a picture.
- When and how was it discovered that there were other galaxies in the Universe in addition to our own Milky Way Galaxy? What is the nearest large galaxy to our own? What is the scale of distances between galaxies? How does this compare to the sizes of galaxies?
- What is meant by the redshift in the light from distant galaxies? What is Hubble's Law and what is the theoretical interpretation of this observational fact?
- What does it mean to say space is expanding? Does our Solar System expand with time? Our galaxy? What in the universe is actually getting bigger as space expands? What in the universe is not affected by the expansion of space?
- Outline the history of our understanding of expanding space. What role does Einstein's General Theory of Relativity play in this? What role did Hubble and Eddington play?
- Describe in some detail the balloon model for expanding space. In what ways is the model a good representation of our universe? In what major way is it an incorrect representation?

The dome of the night time sky is studded with stars. There are approximately 5000 that are bright enough to be seen by human eyes. For most of the history of our species, that was all the stars that existed. After all, why would a creator make stars that could not be seen by us, the very purpose of creation?

6.1 The Milky Way

In addition to the stars, moon, planets, and sun, there is also an arc of diffuse light that stretches across the sky from horizon to horizon. It is prominent; obviously something of importance. Virtually all cultures had myths to explain its origin. Ancient Armenian mythology called the arc The Straw Thief's Way. According to legend, the god Vahagn stole some straw from the Assyrian king Barsham and brought it to Armenia during a cold winter. When he fled across the heavens, he spilled some of the straw along the way. The Cherokee called it The Way the Dog Ran Away explaining it as cornmeal spilled by a dog who stole it and spilled it as he was fleeing.

We call it the Milky Way. The name comes from the Greek myth of its origin. Zeus had a son, Heracles, who was born of a mortal woman. Zeus was fond of him and decided to let the infant suckle on his divine wife Hera's milk when she was asleep, an act which would endow the baby with godlike qualities. When Hera woke up and realized she was breastfeeding an unknown infant, she pushed him away and the spurting milk became the Milky Way. Our word galaxy also comes from this story, as the Greek for milk is *gala*.

Although some presocratics speculated that the Milky Way may consist of stars too close together and faint to be seen as individual stars, the idea was not accepted by Aristotle and was therefore not part of the Late Medieval and Renaissance worldviews. Like so much of Aristotle's work, his view of the Milky Way did not survive Galileo's empirical approach to knowledge. With his telescope, Galileo was able to see the individual stars within the Milky Way. Thus the realm of the stars was no longer seen as spherical but rather disk shaped. Galileo, not given to modesty, bragged that, "The universe which I with my astonishing observations and clear demonstrations had enlarged a hundred, nay, a thousand-fold beyond the limits commonly seen by wise men of all centuries past, is now for me so diminished and reduced, it has shrunk to the meager confines of my body."

This disk model of the Milky Way survived into the 20th century, but it presented a serious philosophical problem. The Copernican Principle, the notion that we are not the center of the universe, seemed to be contradicted by the model. As astronomers peered into the disk, it appeared to end at about the same distance from the solar system in all directions. Are we in the center after all? Observationally it appeared that we are.

Well, we are not. The problem is that the disk of our galaxy, in addition to stars, unexpectedly contains gas and dust. The dust is very effective at absorbing and scattering visible light, thus blocking our view of the more distant stars in the disk. The actual disk is many times bigger than the part we can see, and, rather than being in the center, we are actually out in the suburbs. By 1920 this was a well established fact, primarily due to the work of the American astronomer Harlow Shapley.

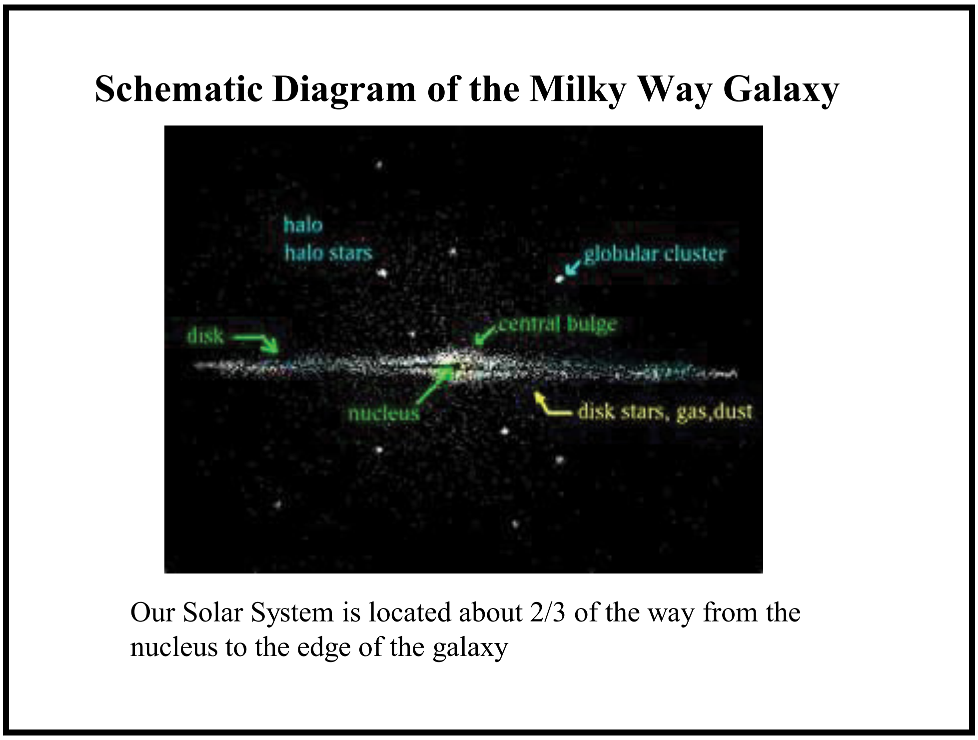


Figure 6.1 Schematic Milky Way

6.2 Other Galaxies

What had not been well established by 1920, was whether our Milky Way Galaxy was unique in the universe or simply one of many galaxies separated by vast distances of space. The debate centered on the nature of spiral nebulae, fuzzy, spiral-shaped structures in the night sky. Some argued that they were other galaxies like our own, but at extreme distances from us, making them appear small. Others argued that they could not possibly be that far away. Each side supported their position with complex, but, in the end, inconclusive evidence. The fundamental problem was that in 1920 there was no way to directly measure the distances to the spiral nebulae.

A variable star is one whose luminosity changes with time. It had been established earlier that certain variable stars, Cepheid variables, have luminosities that are proportional to their periods of variability. For example, a Cepheid variable that completed its periodic variability in 30 days was more luminous (emitted more energy per second) than one with a period of 20 days. The period and luminosity are related by a mathematical equation. Because the period of variability is easy to measure, the period-luminosity relationship can be used to establish the luminosity of the star. Then by measuring the amount of energy we actually receive from the star (the apparent brightness of the star), the distance to the star can be determined. The apparent brightness of an object diminishes with the distance from the object. For example, the wattage of an incandescent light bulb is a measure of its luminosity, its intrinsic brightness. If you move farther and farther from the bulb, it appears dimmer and dimmer. The inverse-square law is an exact mathematical relationship between luminosity, apparent brightness and distance. However this method could not be utilized to measure the distances to spiral nebulae until there was a telescope powerful enough to see individual stars in the spiral nebulae.

In 1919, the 100-inch Hooker telescope of the Mount Wilson Observatory was completed and Edwin Hubble was put in charge of it. With this powerful instrument, Hubble was able to identify Cepheid variables in several spiral nebulae. He determined that they were at great distances from us, distances consistent with their being galaxies. His historic 1924 paper ended the debate, finally establishing the fact that we live in an Island Universe. The spiral nebulae became spiral galaxies, more or less like our own spiral galaxy. In addition to spiral galaxies, Hubble also observed elliptical and irregular galaxies.

Figure 6.2 Andromeda Galaxy

Hubble realized that the galaxies are not uniformly, or even randomly, distributed throughout space. Rather they are clustered. Our own galaxy is part of a cluster of about 30 galaxies called the Local Group. At distances many times greater than the distance between galaxies in a cluster, are other clusters of galaxies. We now know that in addition to clusters of galaxies, there are clusters of clusters, superclusters as they are called. This seems to be the last level of clustering. However, the superclusters are not uniformly or randomly distributed either, but rather are distributed around the edges of huge volumes of seemingly empty space known as voids. As best we can determine, on the scale of voids the universe is at last homogeneous. As far as our most powerful telescopes can see, this distrubutional model appears to hold. The universe is like foam, with the voids being the empty space within bubbles formed by the superclusters.

6.3 Hubble's Law

Hubble continued to use the 100-inch telescope to study galaxies. In 1929, he published a paper that would later be recognized as the most important observational fact about the universe ever discovered. It is a simple fact that on the surface did not seem particularly revolutionary. It was only after it was properly interpreted that it became so.

The light we receive from galaxies has many wavelike properties. It had been known for some time that the light we receive from galaxies beyond the Local Group is redshifted. That is the wavelengths of the light are stretched out relative to the light we receive from nearby objects. In the visible spectrum, red is the longest wavelength and blue is toward the shorter end of the spectrum. If the wave is longer than expected it is called redshifted. Conversely, if it is shorter than expected, it is called blueshifted. The light from distant galaxies is always redshifted. By 1929 Hubble had gathered enough data to establish that the amount of the redshift is direct proportion to the distance to the galaxy from which it was emitted. This observational fact is now known as Hubble's Law.

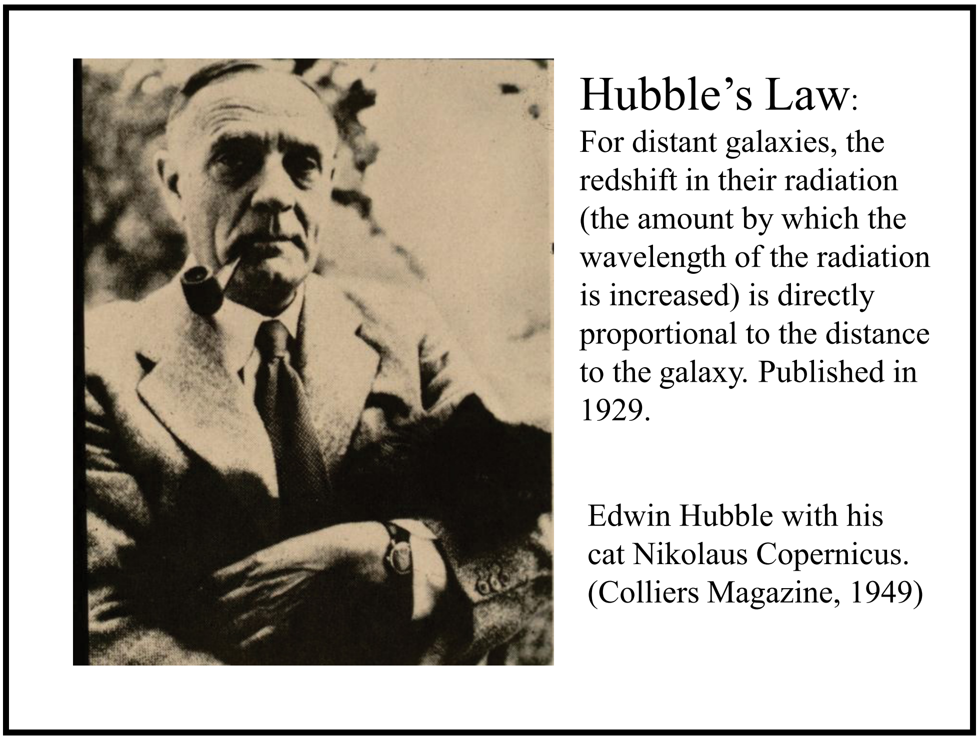


Figure 6.3 Hubble's Law

Hubble recognized that he had established an important piece of information about the nature of the universe, but he did not know exactly what that information was telling us. He was an observational astronomer, and what was needed was a theoretical explanation. That explanation was given in 1930 by the British theoretical physicist, Arthur Eddington.

6.4 The Expansion of Space

Ironically, the explanation had been available well before the existence of other galaxies had been established. In 1905 Albert Einstein published his Theory of Relativity, a description of the nature of space and time that united them in a way that predicted the existence of some seemingly bizarre phenomena. Among them were time dilation and space contraction, which were soon empirically verified. It also predicted the equivalence of mass and energy, that is, that mass is just another form of energy, the two related by the famous equation, $E = mc^2$. He realized at the time that this 1905 theory was incomplete. He was certain that gravity was a property of space-time but it was not included in his original theory. He immediately set about to fix this.

By 1915 he had revised his theory to include gravity. The new version is now known as the General Theory of Relativity, as opposed to the 1905 version, now called the Special Theory of Relativity. As a first application of the new theory, Einstein calculated the orbit of Mercury. It was well known that calculations based on Newton's theory of gravity gave slightly the wrong answer. As with the discrepancy in the calculation of Uranus' orbit, the initial attempt to fix this involved an undiscovered planet that was influencing the orbit. The planet was even given a name – Vulcan. It soon became clear that such a planet did not exist and therefore there must be a problem with Newton's theory of gravity. Calculations based on Einstein's theory of gravity, however, gave exactly Mercury's observed orbit. Clearly Einstein's theory was an improvement over Newton's.

In addition to giving the correct orbit of Mercury, there was another feature of Einstein's theory that made it a significant improvement – it did away with action-at-a-distance. In Einstein's theory the sun does not pull on the planets. Rather it warps the space-time in its vicinity in such a way that the inertial motion of the planets through space-time is an elliptical orbit around the sun.

By 1916 however, Einstein began to have doubts about the theory. When he applied it to the universe as a whole, it clearly predicted that space must be either expanding or contracting. The nature of gravity in the theory would not allow for a static universe. In the theory, gravity was only attractive and this would cause an initially static universe to collapse in on itself. He, and the other scientists he consulted, believed that the universe was in fact static. Because of the theory's ability to account for Mercury's orbit, Einstein knew that he was on the right track, so rather than discard the theory, he tried to tweak it to eliminate the offending prediction. He introduced a repulsive force into his equations, the cosmological constant, whose sole purpose was to balance the attraction of gravity and thus allow for a static universe.

Einstein, however, had made an error. His introduced repulsive force could not properly balance gravity. The equilibrium between the two was unstable, like balancing a pencil on its point. Theoretically possible, but any slight disturbance unbalances it, in this case resulting in either the expansion or contraction of the universe. Exactly what Einstein was trying to avoid. This was pointed out in the 1920s by at least two theoreticians, one a Belgian priest, George Lemaitre, about whom we'll have more to say later.

Einstein lecturing on the GTR in Pasadena, California, 1932.

Einstein developed the general theory of relativity (GTR) in 1915.
It predicted that space had to be either expanding or contracting.
Einstein believed this to be incorrect and changed his theory.

Figure 6.4 Einstein Pasadena

These papers were mostly ignored by other physicist and it is unclear that Einstein himself was aware of them. One person who was aware on them however, was Arthur Eddington, considered at the time to be one of the few physicists who actually understood Einstein's theory. When he heard about Hubble's Law, he immediately saw it as observational evidence that space is expanding. When Einstein learned of this, he called his inclusion of the cosmological constant in his theory "the greatest blunder of my scientific career." We will learn in the next chapter that it may not have been a blunder after all.

The expansion of space does not mean that everything in the universe is getting bigger with time. Some structures are held together by forces sufficiently strong to prevent the expansion of space from increasing their size. This is true for galaxies (and everything in them) as well as for clusters of galaxies. However, gravitational attraction between the clusters and superclusters is not sufficiently strong, and the expansion of space results in the increase in the sizes of the superclusters and the voids.

The concept of expanding space is a complex and confusing one. The first question most people ask is, "What is it expanding into?" Unfortunately this question doesn't have an answer, at least not one that can be visualized. Generally when we think of something expanding, we visualize it getting bigger. That is, we visualize its edges as moving outward. But the universe has no edges. Aside from the fact that an edge to the universe doesn't make any sense (what would happen if you stepped over the edge?), the General Theory of Relativity, the theory we use to describe the space-time of the universe, requires an unbounded universe. This is true whether the universe is infinite (clearly no edges) or finite (seemingly not possible), either of which is allowed by the theory.

Our universe has three spatial dimensions (at least, three that have extension). A finite but unbounded space can only occur if the space is curved. To visualize curved space, one dimension greater than the dimensions of the space is required. Thus, we cannot visualize a curved three-dimensional space because we cannot visualize in four dimensions. Since concrete, visualizable models are helpful in understanding abstract concepts, astronomers use a curved two-dimensional space to help us understand an expanding, finite but unbounded universe.

Imagine a two dimensional object, something like the surface of a piece of paper. If it is not infinite, it is bounded by edges. But if we bend it into the shape of a sphere, curve it into the third dimension, it no longer has edges. The surface of a sphere is a finite, but unbounded, two-dimensional surface. We can visualize a finite but unbounded two-dimensional surface because we can visualize in three dimensions. We cannot visualize a finite but unbounded three-dimensional surface because we cannot visualize in four dimensions.

Now imagine the sphere to be the surface of a balloon and let that surface represent the space of a fictitious two-dimensional universe. Obviously, expansion will be represented in this model by blowing up the balloon. If we attach small pieces of paper to the surface of the balloon and let each represent a cluster of galaxies, we have a pretty good representation of the consequences of expanding space for our universe. As the balloon is blown up, the clusters of galaxies get further apart (i.e. the sizes of superclusters and voids increases) while the size of the clusters is unaffected. The distance between clusters is increasing, but the clusters are not moving. That is, they are not moving through space, as they are fixed in space (stuck to the surface of the balloon in the model).

As three-dimensional beings, it is easy for us to answer the question, what is the space of the two-dimensional universe expanding into? It is expanding into the third dimension of space. However, this answer would make no sense to two-dimensional beings. In their universe there is no third dimension to space. In a similar way, we could say that our three-dimensional universe is expanding into the fourth dimension. But that makes no sense to us. It is best to just stop asking the question in that way. We are seeking to visualize something that is inherently non-visulaizable.

The balloon model also represents another important fact about our universe. All places on the surface of the balloon are equivalent. There is no center in this model universe, at least not one that is in the space of the universe itself – not on the surface of the balloon. Just as our universe has no edges, it also has no center. This is an important point to remember when we talk about the Big Bang in a later chapter.

The Balloon Model of Expanding Space

Clusters of galaxies are represented by pieces of paper on the balloon.

As the balloon is blown up its surface area (space) increases with time.

The clusters of galaxies do not increase in size. They get further apart but do not move through space.

Figure 6.5 Balloon Model

Expansion of Space

- 1916 - Einstein’s general theory of relativity predicts that space must be either expanding or contracting. Einstein does not believe this and tries to “fix” the theory.
- 1920s - Other astronomers and physicists show that all versions of the GTR require either the expansion or contraction of space.
- 1929 - Hubble’s Law.
- 1930 - Arthur Eddington explains Hubble’s Law as the expansion of space as described by the GTR.
- 1930 - Einstein calls his not accepting his original theory “the greatest blunder of my scientific career.”

Figure 6.6 Expansion of Space

The General Theory of Relative can be used to describe a universe that is either finite or infinite, just as it can be use to describe a universe that is either expanding or contracting. Observational data is required to determine which of these theoretical possibilities matches our actual universe. Hubble's Law is observational evidence that we live in an expanding, rather than a contracting, universe. In our balloon model, we represented a finite universe. However, recent evidence seems to indicate that we live in an infinite, rather than a finite, universe.

If the universe is infinite, how can it be expanding? This is perhaps even more difficult to imagine than a finite expanding universe. How can there be more space later if there is already an infinite amount of it now? One idea that might help with this is to think of numbers. How many whole numbers are there? There are an infinite number of them. If we divide each odd whole number by two we get another infinite set of numbers. Combining these new numbers with the whole numbers, we have half again as many numbers. There are still an infinite number of them, but, in a sense, the infinite set of whole numbers has expanded to an even larger infinity.

Maybe that doesn't help much, but it's not important. We simply cannot conceive of the universe as a whole and it's not helpful to try. When you think of the universe, just think of a finite part of it, say a part big enough to contain many voids and superclusters, and know that the rest of the universe is just more of the same.

6.5 The Contents of Galaxies

When we think about galaxies, we visualize them as our telescopes see them: stars with some interstellar gas and dust. By the 1970s, astronomers had come to realize that these components of galaxies make up only a small fraction of their mass. The additional mass is called dark matter and is inferred to exist because of its gravitational effect on visible matter. This is the only way it can be detected because it does not emit, absorb, or scatter electromagnetic radiation. Thus, it must consist of uncharged particles. At the present time we do not know the nature of this matter, only that it is not made up of ordinary matter: protons, neutrons, and electrons. Dark matter will be discussed in more detail in Chapter 9.

In addition to dark matter, most galaxies contain supermassive, but extremely compact, objects at their center. The evidence indicates that these objects are black holes with masses on the order of hundreds of thousands to billions of times the mass of our sun. The object at the center of our Milky Way Galaxy has a mass of about four million solar masses. Although his is a very large mass for a single object, it is small relative to the mass of our galaxy, whose stars alone have an estimated mass of hundreds of billions of solar masses.

6.6 The Universe vs The Visible Universe

There is consistently confusion in TV programs and also in some textbooks about what is meant by the word 'universe.' Generally speaking it means everything that exists. But there are times when what is really meant by the term is the visible universe; that is, the part of the universe that we have observational access to. The evidence at the present strongly indicates that the universe is infinite. However, because the age of the universe (at least that part of the universe for which observational evidence is possible) is finite (13.7 billion years), the visible universe is finite.

As we look out into space we are also looking back in time. Light travels at 186,000 miles per second. At this speed it takes 2.2 million years for light from the Andromeda galaxy to reach us, so another way to quantify the distance to Andromeda is the say that it is 2.2 million light-years away. Some galaxies we are able to observe are more than 10 billion light-years away. Obviously we cannot observe anything that is more than 13.7 billion light-years away, and, thus, the observable or visible universe is finite, extending out 13.7 billion light-years from us in all directions. This is our so-called 'horizon.' The Big Bang model does not attempt to describe that region of space significantly beyond our horizon – space-time could well be quite different out there.

It is not uncommon to hear a commentator say that at a certain time after the Big Bang, the universe was the size of a pea. Often they will even hold up a pea to show you how small it was. In fact the universe probable is, and always has been, infinite. What they really mean is that, at a certain time after the Big Bang, the visible universe was the size of a pea. This is ok because the visible universe is finite in size and getting bigger with time. However, it is very misleading. A pea has a center and an edge in space. The visible universe does not. This further leads to the idea that the Big Bang occurred at the center of this pea and matter has spread out to the size of a pea since then. This is not the case. The center of our visible universe is us. This would be equally true of any observer anywhere in the universe.

The distinction between the universe and the visible universe is not easy to keep in mind, and in most cases it is not an important one. There are times in the following chapters where the word universe is used when visible universe might be more appropriate. However, the distinction is important when talking about the size of the universe.

7 The Big Bang

Your goals for this chapter are to know the following:

- What is the big bang theory of the early universe? What were the two most significant predictions made by this theory?
- Compare and contrast the big bang model with the steady state theory of the universe.
- Discuss the cosmic background radiation is some detail. What was the origin of the cosmic background radiation? Briefly discuss the discovery of the cosmic background radiation.
- Compare and contrast Lemaitre's model of the early universe with Gamow's.
- Identify the single most important evidence in favor of the big band.

Hubble's Law is an observational fact and is beyond dispute. Anyone who doubts it can make the measurements his or herself and will get the same result. Expanding space is not a fact. It is the theoretical explanation of a fact. It can be doubted. Perhaps there is an alternative, completely different explanation for Hubble's Law.

The ability to explain known facts is an important criterion a theory must pass to be accepted. But there is an even more important criterion – the ability to make predictions of facts that are not known at the time. If these predictions turn out to be true, the theory is not proven true, for it is always possible that a later prediction of the theory will fail. However, with each new successful prediction, confidence in the theory is strengthened. The concept of expanding space has made many such successful predictions and as a result is tentatively accepted by scientists as true. It is such an integral part of any explanation of the nature of the universe, that we will assume, for the purposes of discussion, that it is true.

7.1 George Lemaitre

If the universe is expanding, the question arises as to what the universe was like in the past. The first person to address this question seriously was a Belgian cosmologist-theologian Abbe George Lemaitre. In 1927 he had shown that Einstein's attempt to eliminate the prediction of an expanding or contracting space from his General Theory of Relativity did not work. His paper received little attention and, discouraged, he turned his research in other directions. However, with Eddington's explanation of Hubble's Law, Lemaitre returned to cosmology in the 1930s.

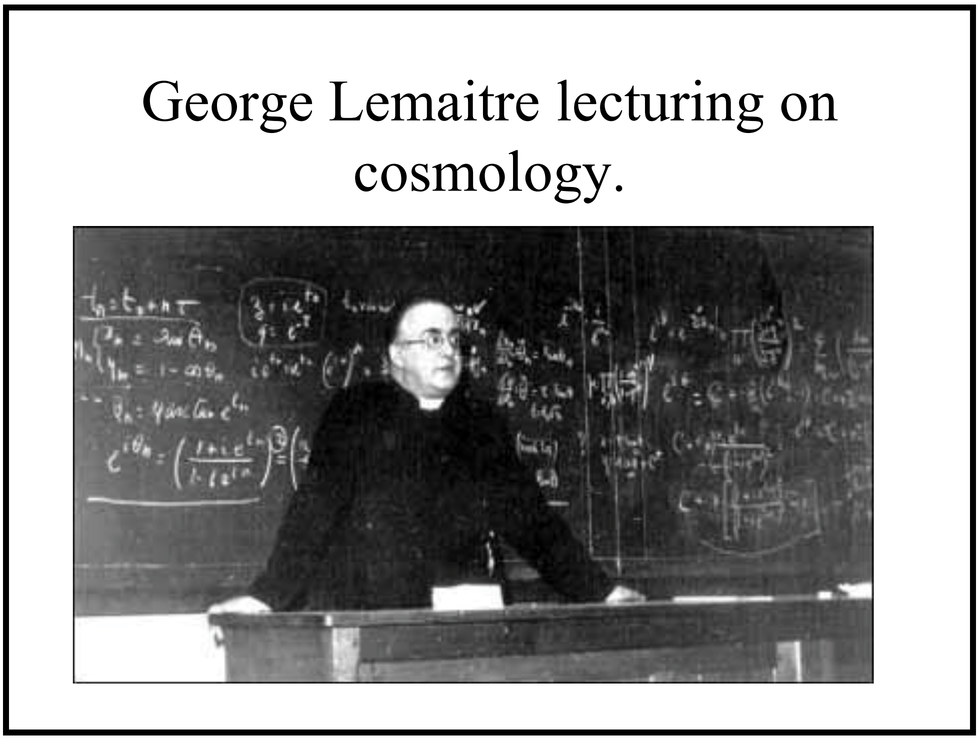


Figure 7.1 George Lemaitre

Lemaitre reasoned that if space is expanding, then the universe must have been denser in the past than it is now. Using the physics of the General Theory of Relative, he realized that there would have been a time in the finite past when the universe would have been infinitely dense, and therefore our universe must have had a beginning in time. This was contrary to the prevailing opinion among scientists.

Most assumed that the universe was infinitely old; that it always was and would always be. Although Lemaitre's model represented a significant philosophical and cosmological breakthrough, it was inadequate in terms of physics. There was little that could be done with the model to see if it could, in fact, evolve into the present universe. That would have required a knowledge of nuclear physics that did not exist at the time. There was another reason why scientists were skeptical of Lemaitre's idea of a universe that had a beginning in time. He was a Catholic priest, and thus possibly predisposed to believing that the universe had been created in the finite past. When science is used to predict something that the predicting scientist probably believed to be true to start with, one should be skeptical.

7.2 George Gamow

George Gamow was born in Odessa, Russia in 1904. He received his Ph.D. in physics from the University of Leningrad in 1928 and in 1931 accepted a position of professor there. This was a time of increasing oppression in the Soviet Union, and Gamow and his wife made several unsuccessful attempts to defect. After years of refusing to give them permission to leave, the government unexpectedly allowed them to attend the 1933 Solvay Congress in Belgium. They used this opportunity to leave the Soviet Union forever. Gamow spent the rest of the year at various scientific institutions in Europe and was appointed professor of physics at George Washington University in Washington, D.C., in 1934.

Gamow's first major contribution to physics was to explain the nature of alpha decay using the new theory of quantum mechanics. By 1942 he had turned his attention to astrophysics, developing a theory of the internal structures of red giant stars. He became interested in Lemaitre's ideas concerning the early universe, and as early as 1946 began to believe that the high temperatures of the early universe could provide the appropriate conditions for the creations of the chemical elements in their proper ratios. Since the time Lemaitre had first addressed the problem, significant advances in nuclear physics has occurred. In 1948, Gamow and Ralph Alpher (Gamow's PhD graduate student) developed a model of an early universe consisting of neutrons at a very high temperature.

When their work was published in 1948, the eminent astrophysicist Hans Bethe was surprised to find himself listed as a co-author. Though Bethe had nothing to do with the work, the practical joker Gamow thought it would be funny to have an important paper authored by Alpher, Bethe, and Gamow, a pun on the first three letters of the Greek alphabet: alpha, beta, and gamma. To Gamow's delight, the paper appeared in the April first issue of *The Physical Review*.

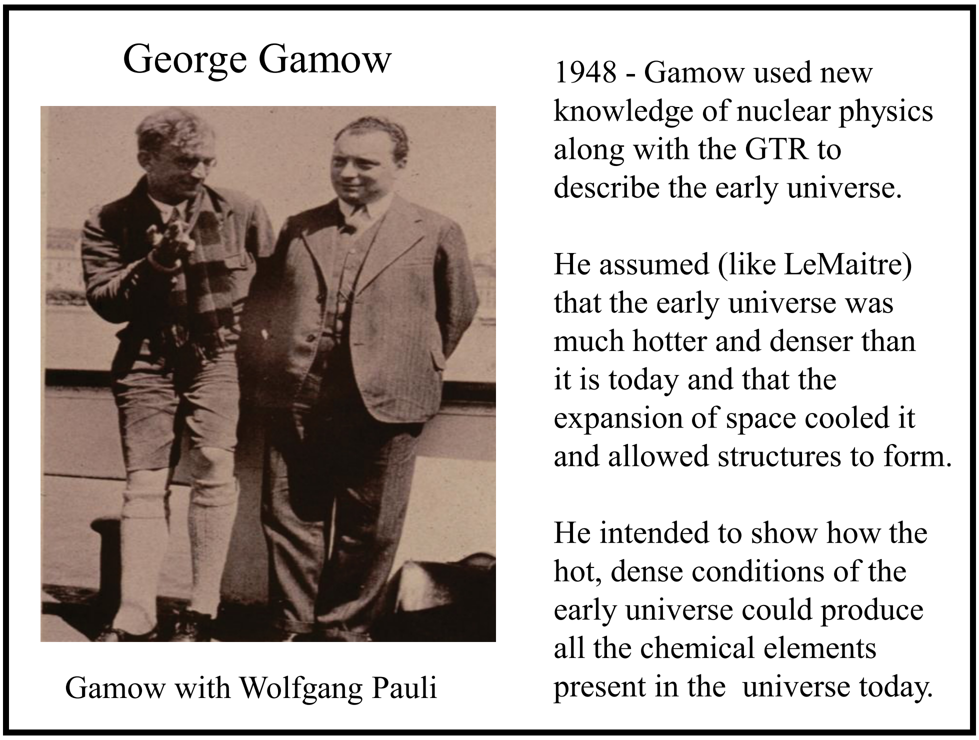


Figure 7.2 George Gamow

This theory (and its subsequent versions) is known as the Big Bang. Ironically the name was given to it by Fred Hoyle, in a BBC radio interview in 1949. Hoyle was co-author of an alternative cosmological theory known as the Steady State theory and a strong critic of Gamow's theory. It is commonly reported that Hoyle intended the term Big Bang to be pejorative, but he explicitly denied that this was the case. He said it was just a striking image meant to emphasize for radio listeners the difference between his and Gamow's theories.

The Big Bang is an unfortunate term for the theory. It suggests an explosion, that is, matter flying out in all directions from a central point into previously unoccupied space. The Big Bang was not an explosion of matter. If anything' it was an explosion of space. Matter has filled all of space more or less uniformly since the beginning of time.

In 1953, the model was modified so that it began with matter in the form of protons, neutrons, and electrons. Through a series of nuclear reactions the protons and neutrons fused to form nuclei. This process is called nucleosynthersis. Gamow's model readily explained the high percentages of hydrogen and helium, but completely failed to account properly for the heavier elements. No matter how many times Gamow redid the calculations, nucleosynthesis essentially stops after helium. This was considered at the time to be a fatal flaw in the theory.

George Gamow's Big Bang theory assumed that the early universe was extremely hot and dense, and that the temperature was continuously decreasing due to the expansion of space. These assumptions led to two unavoidable predictions. First, that the early universe essentially contained only the elements hydrogen and helium. This was a disappointment to Gamow. His motivation for beginning the calculations was to show that the relative abundances of chemical elements present in the universe now could be explained by the hot, dense conditions of the early universe. Also, if the heavier elements were not produced in the early universe neither he, nor anyone else at the time, knew where they could have come from.

Gamow also predicted that the universe should be filled with electromagnetic radiation left over from the events of the early universe, but significantly cooled by the subsequent expansion of space. Shortly after the abγ paper, Ralph Alpher and Robert Herman calculated the present temperature of the radiation at about 5 K (5 degrees above absolute zero). At the time this was an untestable prediction. For this and other reasons, the work of Gamow and his students, Alpher and Herman, was not taken very seriously.

Predictions of the Big Bang model

- The early universe contained only hydrogen and helium. Because of the expansion of space and its cooling effect, nucleosynthesis only occurred between 3 to 4 minutes after the big bang (A.B.B.) and essentially stopped after helium.
- The universe is filled with a background radiation whose temperature is a few degrees above absolute zero. When neutral atoms formed (about 370,000 yrs A.B.B.), the electromagnetic radiation essentially stopped interacting with matter. The expansion of space cooled the radiation from its initial value of about 3000 K to its present low value.

Figure 7.3 Predictions of the Big Bang

It is important to note that the Big Bang theory is not a theory of the origin of the universe as it is often said to be. Neither when first proposed nor now does it say anything of the origin. The earliest temperatures and densities are far too great to be described using our current understanding of physics. The Big Bang theory is rather a description of the very early universe.

Over his career, Gamow switched his interests from nuclear physics to astrophysics to cosmology. In the 1950s he turned his attention of the origin of life. He quickly recognized the significance of the DNA model proposed by Crick and Watson in 1953, and realized that the bases contained a code for amino acids. Gamow is best known to the public for his excellent books popularizing abstract physical theories, several of which are still in print.

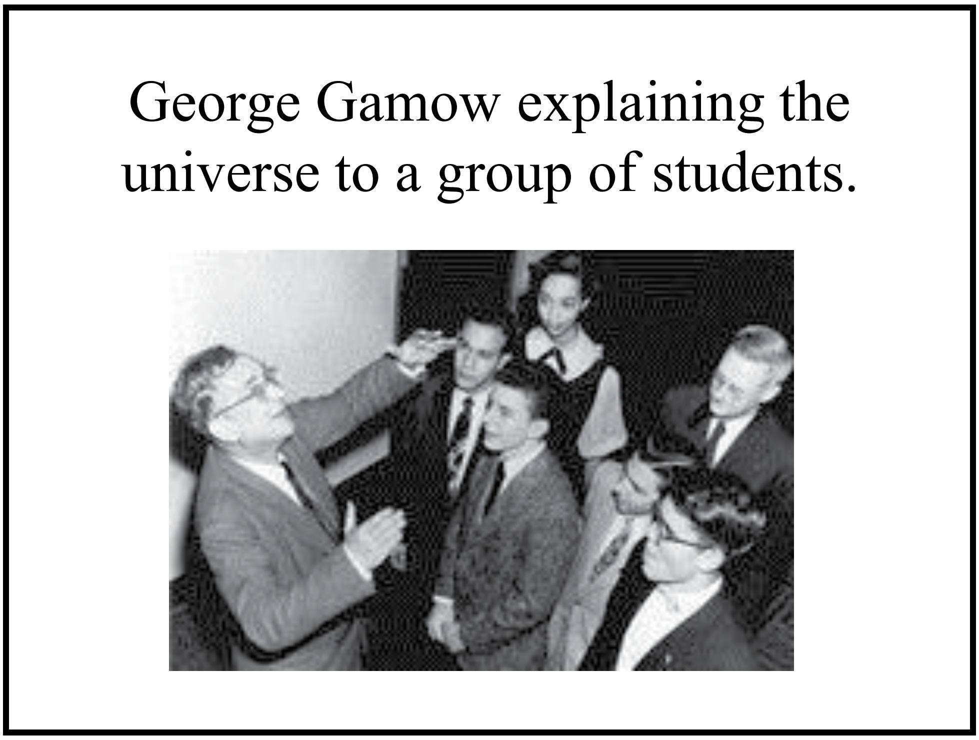


Figure 7.4 George Gamow and students

7.3 Robert Dicke

The first step toward confirming the first prediction was solving the problem of the origins of the heavier elements. This happened in 1957, ironically due in part to Fred Hoyle, the most prominent opponent of the Big Bang. He and his colleagues were able to show that heavier elements could be produced in the interiors of massive stars. Since then observations of the oldest stars in our galaxy and of very distant galaxies early in their evolution have completely confirmed the prediction that the early universe was essentially pure hydrogen and helium.

The confirmation of the second prediction, the existence of what is now called the cosmic background radiation, constitutes one of the most important discoveries in the history of science. The story starts in the early 1960s. Robert Dicke, and his colleagues at Princeton University, began to rethink the consequences of a Big Bang-like beginning to the universe. Working through the physics, Dicke, like Gamow before him, realized that the early universe had to be nearly pure hydrogen and helium. This was no longer a problem as it was by then known that heavier elements are produced in stars.

Dicke realized that if the universe had passed through a uniformly hot stage in its compressed past, then the present universe would necessarily be filled with electromagnetic radiation, cooled from its initial high temperature by the expansion of space. This had also been predicted by Gamow. However when Gamow predicted it, radio technology was not sufficiently advanced to detect radiation associated with such a low background temperature. As a result, the original prediction was soon forgotten. In 1975, Dicke wrote,

> "There is one unfortunate and embarrassing aspect of our work on the fireball radiation. We failed to make an adequate literature search and missed the more important papers of Gamow, Alpher, and Herman. I must take the major blame for this, for the others in our group were too young to know these old papers. In ancient times I had heard Gamow talk at Princeton, but I had remembered his model universe as cold and initially filled only with neutrons."

The technological situation had improved by the time the Princeton group reached Gamow's earlier conclusion. It was now clear that such a background radiation should be observable, and the Princeton experimentalists, unaware of the earlier prediction, immediately set about building an appropriate radio receiver. If something like the Big Bang had occurred, the background radiation must exist, and although it would not be an easy experiment, Dicke and his colleagues were confident they would be able to detect it.

The Princeton physicists were in a situation all scientists dream of. They were addressing a question of fundamental scientific importance. They would either discover the background radiation and provide extremely strong evidenced in favor of the Big Bang, or they would show that the background radiation does not exist and thereby disprove the Big Bang theory. In either case, a contribution to science worthy of its highest honor, the Nobel Prize, would result.

7.4 Penzias and Wilson

At the same time the Princeton group was setting up their experiment to search for the background radiation, two radio astronomers were working on Bell Laboratories' horn-shaped radio telescope in nearby Homedale, New Jersey. The instrument was intended for use in trans-Atlantic telephone communication, and Arno Penzias and Robert Wilson were assigned to evaluate the performance of the telescope. In return for this rather mundane work, they would be allowed to use the telescope for their scientific work.

Penzias and Wilson ran the tests at a wavelength of 7.35 centimeters, in the microwave region of the radio spectrum. They quickly found that no matter what direction in the sky the antenna was pointed, some signal was received. This was totally unexpected. Because there should be nothing out there producing radiation at this wavelength, they at first assumed that the signal was being generated within the instrument itself.

Although the level of the noise was not sufficient to significantly interfere with telephone conversations, Penzias and Wilson were consciousness scientists and wanted to understand where the noise was coming from. They carefully inspected the electronics, they covered all the rivets in the telescope horn with aluminum tape, and they even took the horn itself apart and removed the "white dielectric material" left behind by pigeons – but there was still far too much unaccounted for background noise. Penzias and Wilson were completely stumped. They had eliminated every conceivable source of the radiation and yet there it was, always present no matter where the antenna was pointed.

In 1965, a friend of Penzias', Bernard Burke of MIT, told him about a preprint paper he had seen by Jim Peebles, an experimental colleague of Robert Dicke's at Princeton, on the possibility of finding radiation left over from the early universe. Penzias and Wilson began to realize the significance of their discovery and called Dicke. After Dicke hung up the phone, he said to his colleagues, "Well boys, we've been scooped." Penzias and Wilson had unknowingly discovered the predicted cosmic background radiation.

The two groups decided to publish their results jointly. Companion letters were rushed to the *Astrophysical Journal Letters*, with Penzias and Wilson announcing their observations and Dicke and his colleagues providing the cosmological interpretation. The Princeton group soon completed their measurements verifying the work of Penzias and Wilson. The detection of the cosmic background radiation is regarded as the second most significant (after Hubble's Law) observation ever made concerning the nature of the universe. In 1978, Penzias and Wilson received the Nobel Prize for this discovery. Unfortunately, and many think unfairly, Gamow was not included.

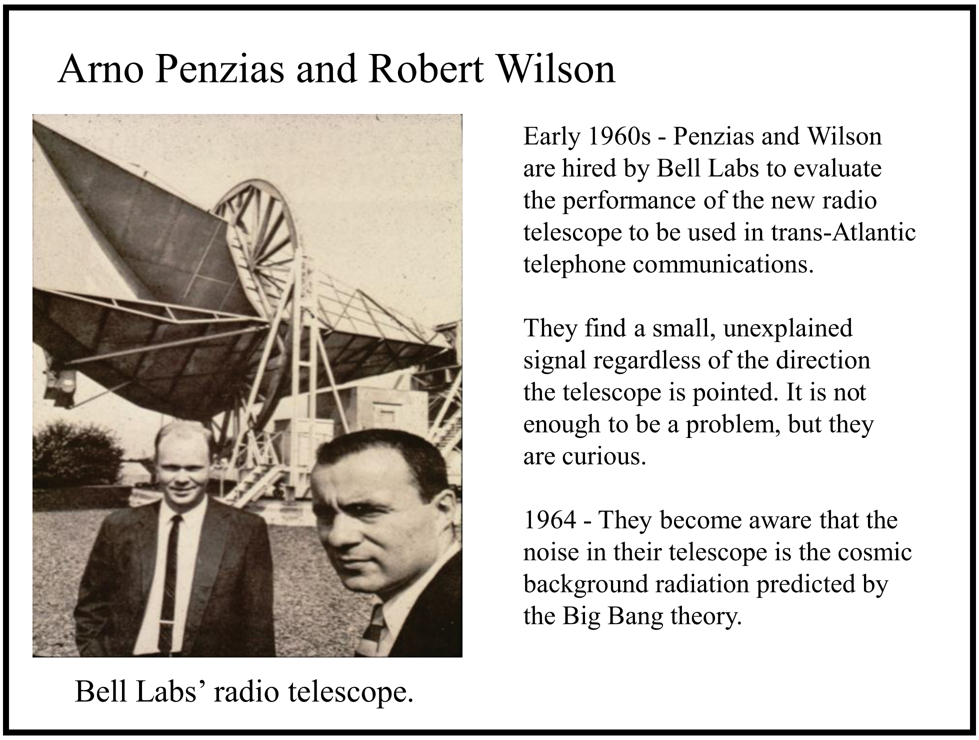


Figure 7.5 Penzias and Wilaon

8 The Early Universe

Your goals for this chapter are to know the following:

- List the sequence of formation of structures in the early history of the universe.
- Identify and explain the role of expanding space in the development of structures in the early universe.
- Compare and contrast the state of the universe before and after the formation of the cosmic background radiation.
- What is meant by the following terms: nucleosynthesis, quark, decoupling?

With the discovery of the cosmic background radiation, the pendulum of cosmological opinion swung abruptly from the Steady State model to the Big Bang model. The Big Bang model had predicted the existence of the background radiation, and the Steady State model not only did not predict it but could not even explain it. It now appears as if the universe must have had a beginning – a moment in the past beyond which it is impossible to trace any chain of cause and effect. Recent calculations indicate that this beginning occurred about 13.7 billion years ago.

To return in your mind to the Big Bang, you would have to travel in time but not in space. The Big Bang was here, in this very location, just far back in time. This planet, this solar system, this galaxy are at the center of the universe. This claim is, however, equally true for every other location in the universe. As with the balloon model, all places in the universe are equivalent. The Big Bang did not occur at a point in space; that is, the expansion is not an expansion of matter into what was previously unoccupied space. All of the space of the universe has been occupied, more or less uniformly, by matter and radiation from the beginning of time. The expansion is an expansion of space itself. Most matter stays essentially at rest in the space it occupies. Where we are now is where the material particles that would eventually form our bodies were located (give or take a tiny fraction of the size of the visible universe) when the universe began.

8.1 Quarks and Leptons

At the initial instant, the universe was in a state of such high density and temperature that our current physical theories cannot be applied. Some highly speculative theories can be used to discuss the universe 10^{-42} seconds after its beginnings. The physics used to describe the universe10^{-35} seconds after the Big Bang, though still unconfirmed in the laboratory, is on somewhat firmer ground. This is the era of Inflation, a time of extremely rapid expansion. Inflation explains several otherwise unexplainable facts about the universe. Also, at the time the theory was proposed in the early 1980s, it made a prediction about the geometry of the universe that everyone believed was false, but which by the 21st century had been shown to be true.

One millionth of a second after the beginning, expansion had reduced the temperature and density of the universe to well within the range covered by laboratory tested physical theories. To be on the conservative side, our discussion of the evolving universe will begin at this time. The temperature of the universe was about 10^{12} K. The ordinary matter of the universe, the matter that will eventually form atoms, stars, and us, is at this time in the form of elementary particles – quarks and leptons.

(Aside: Aristotle objected to the concept of elementary particles on the basis that, if they had extension in 3-dimensional space, they were, at least in principle, divisible and could not be elementary. It appears as if quarks and leptons do not, in fact, have extension in 3-dimensional space. After more than 2000 years, and long after scientists accepted the idea of elementary particles, Aristotle's legitimate objection has been addressed and answered.)

8.2 Protons and Neutrons

There are forces in nature that are capable of holding quarks and leptons together, but they are not infinitely strong. If the temperature is too high, the particles will collide so violently that any structure that has existed temporarily will be knocked apart by the impact. However, as the universe expands and cools, levels of structure appear depending on the strength of the force responsible for the structure. The structure that is formed by the very strongest of these forces appears first. That force is the force that quarks exert on one another. At a few millionths of a second after the beginning, when the temperature was about 10^{12} K, groups of three quarks came together to from nucleons, particles that will eventually form atomic nuclei. Depending on the type of quarks involved, the nucleon will either have a positive charge (a proton) or no charge (a neutron). This era in the history of the universe is known as quark confinement.

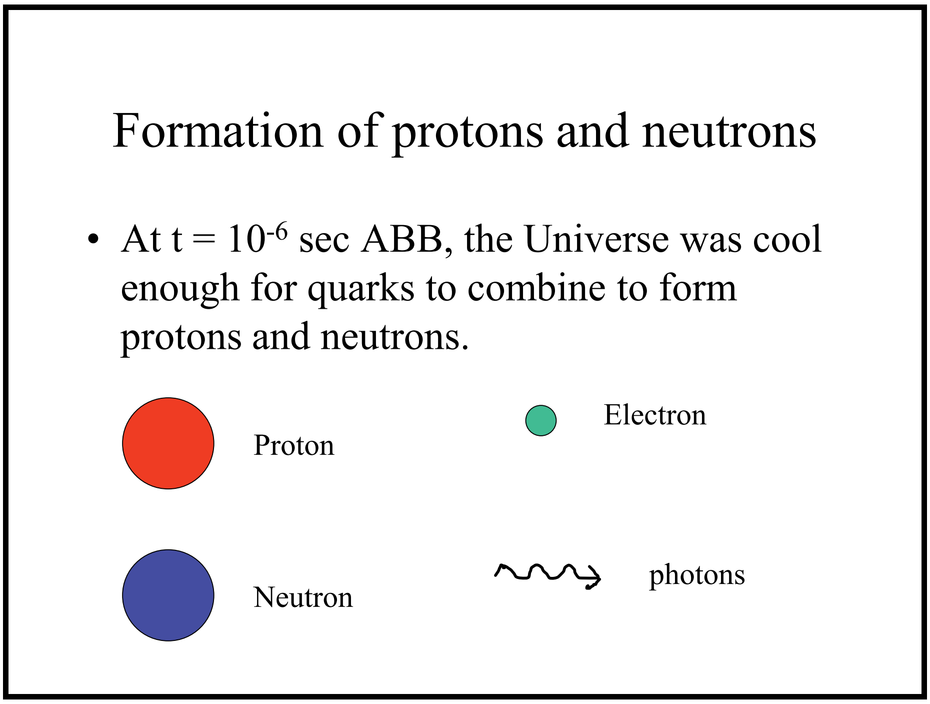


Figure 8.1 Protons and neutrons

The second level of structure forms about 3 to 4 minutes later when the temperature has been reduced to 10^9 K. At the time of quark confinement, there are an equal number of protons and neutrons. Because of their slightly greater mass, nuclear reactions favor the transformation of neutrons to protons over the transformation of protons to neutrons, so by this time protons outnumber neutrons about seven to one. It is now cool enough for the strong nuclear force to bind the protons and neutrons together to form atomic nuclei.

The strong nuclear force is actually a residual force resulting from the interactions between the quarks that form the protons and neutrons. This residual force is less strong than the force that binds quarks and therefore a lower temperature is required in order for it to exert an effect. Its nature also restricts it to very short ranges, unlike gravity or the electromagnetic force. The protons and neutrons must literally come in contact with one another for the force to have an effect.

8.3 Nuclei

Most of the neutrons are quickly bound with protons to form a helium nucleus consisting of two neutrons and two protons. A small fraction of the neutrons are bound into nuclei consisting of one proton and one neutron (hydrogen, sometimes called deuterium in this form), or two protons and one neutron (helium), Even rarer, nuclei consisting of three protons and four neutrons (lithium) can form. When all of the neutrons are bound into nuclei, left-over protons (hydrogen nuclei) constitute about 94% of the nuclei and almost all of the rest are helium nuclei consisting of two protons and two neutrons. The other types of nuclei constitute only a tiny fraction of a percent of the nuclear matter.

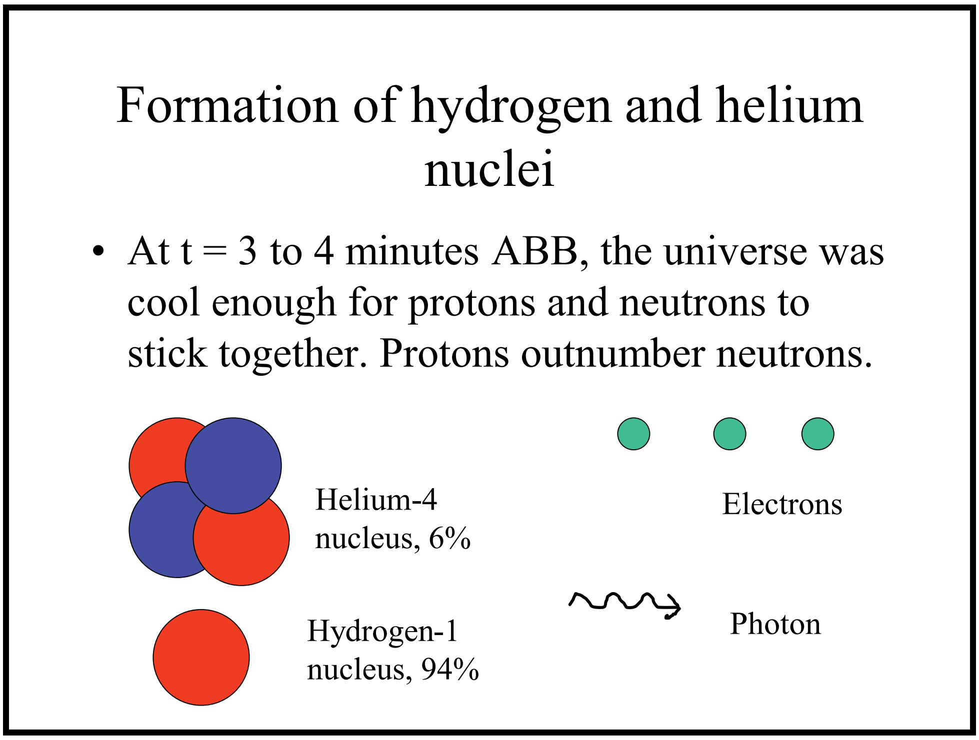


Figure 8.2 Hydrogen and helium

8.4 Atoms

The third level of structure does not form until about 370,000 years after the Big Bang. This level consists of neutral atoms formed when electrons combine with nuclei due to the electromagnetic force of attraction between the negatively charged electrons and the positively charged nuclei. The electromagnetic force is very much weaker than the force that binds protons and neutrons to form nuclei, so much lower temperatures must be reached before atoms can form, something on the order of 3000 K.

In this discussion of the early universe, we have focused on the development of structure in the universe. The particles involved are the quarks and leptons and the structures are, first of all, protons and neutrons, then nuclei, and finally, atoms. This matter, as we will learn later, makes up only a small fraction of the matter in the universe. This other matter, sometimes called exotic matter or dark matter, does not form structures and exists in the universe only in the form of elementary particles. Because our goal is to trace the cosmological history of human beings, and because we are composed of ordinary matter,our focus will be on the evolution of ordinary matter.

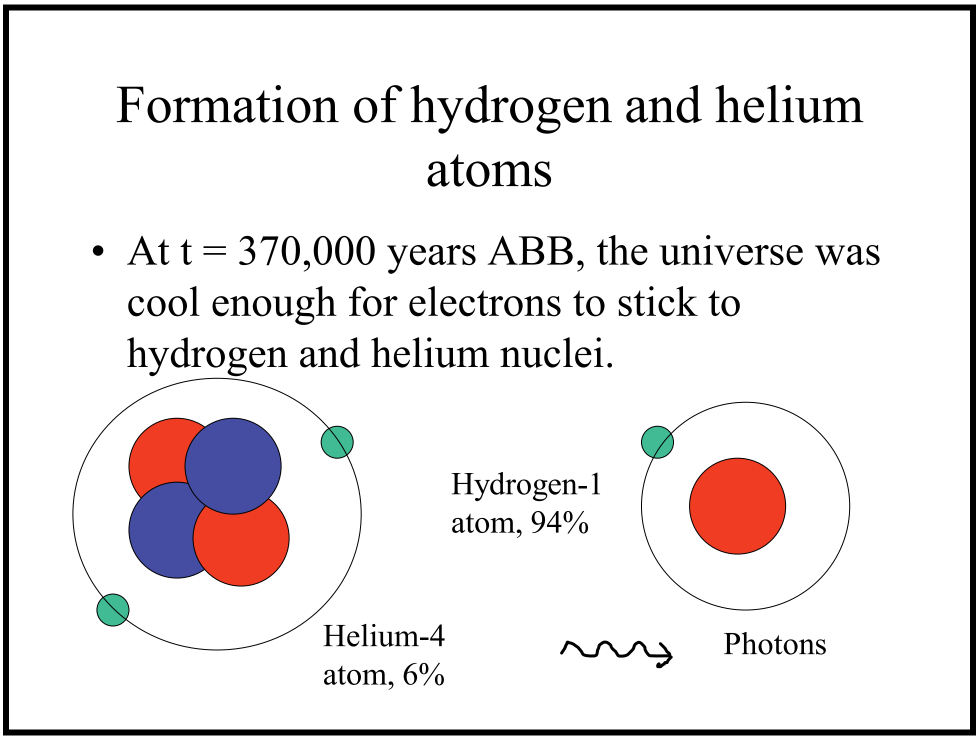


Figure 8.3 Atoms

Prior to the formation of neutral atoms, photons interacted very strongly with matter. This interaction was caused by the fact that the individual particles of matter (hydrogen nuclei, helium nuclei, and electrons) are electrically charged. Photons (electromagnetic radiation, after all) interact very strongly with charged matter, traveling only short distances before they are destroyed and new photons are created. Constantly interacting in this way, matter and electromagnetic radiation evolved together, both at a common temperature.

However, with the formation of neutral atoms, the probability of interaction between matter and photons is reduced to near zero. The average distance a photon travels between interactions with matter suddenly becomes so large that virtually all of the photons present in the universe at the time have not interacted with matter since, and thus must still be in the universe. Matter and radiation began to evolve independently, each occupying the same universe more or less uniformly, but without interacting. This phenomenon is referred to as decoupling. Matter has since gone on to form stars, planets, people, and so on, while the continuing expansion of space has caused the temperature of the electromagnetic radiation to drop. At the time of decoupling, the temperature of the universe was about 3000 K. Today it is 2.73 K.

The scenario just described is the same reasoning that both Gamow and Dicke went through in analyzing their assumption that the universe began a finite time ago, in an extremely hot and dense state. That reasoning convinced them that the early universe was essentially pure hydrogen and helium. It also resulted in their independently predicting the existence of the cosmic background radiation, the discovery of which is the single most significant evidence in favor of the Big Bang theory.

The Evolving Universe

- **The Planck era.** Time = 10^{-43} sec, temp = 10^{32} Kelvin. Current theories are inadequate. We can't get any closer to the Big Bang at t=0 and say anything with confidence (or even with informed speculation).
- **Inflation.** Time = 10^{-35} sec, temp = 10^{28} Kelvin. A temporary period of exponential expansion. A speculative theory, but one that has so far been consistent with observations.
- **Quark confinement.** Time = 10^{-6} sec, temp = 10^{12} Kelvin. Quarks become bound into the protons and neutrons we see today.
- **Primordial nucleosynthesis.** Time = 10 sec, temp = 10^{9} Kelvin. The universe cools to a point where protons and neutrons can combine to form light atomic nuclei, primarily helium, deuterium, and lithium.
- **Decoupling.** Time = 3.7×10^{5} yrs, temp = 3×10^{3} Kelvin. The universe cools to a point where electrons can combine with nuclei to form atoms, and matter becomes transparent. The cosmic background radiation is released.

Figure 8.4 Evolving universe

9 Modern Cosmology

Your goals for this chapter are to know the following:

- Identify the principle findings from the WMAP probe.
- What is the evidence for the existence of dark energy, and what are some of its consequences for the universe?
- Compare and contrast the contents of the universe at the time the cosmic background radiation formed and the present.
- Be familiar with the concept of curvature in the universe. Which geometry seems most likely at the present time?
- What are the assumptions of the ΛCDM model of the universe?

We are in a new era of our understanding of the universe known as precision cosmology. The term comes from the fact that strong new boundary conditions have been placed on cosmological models as a result of precise measurements of important cosmological parameters. For the most part, this new knowledge has resulted from increasingly detailed studies of the cosmic background radiation.

In 1989, NASA launched the Cosmic Background Explorer (COBE), which quickly produced a spectacular result: the temperature of the radiation is 2.725 K. Subsequently, COBE data revealed very slight irregularities in the overall temperature distribution – temperature fluxuations on the order of 0.0004 K. This was an extremely important result because studies to that point had found no fluxuations. In the absence of fluxuations, the existence of stars and galaxies made no sense at all. It was quite a relief to theorists that the fluxuations existed. For these discoveries, John C. Mather and George Smoot won the Nobel Prize in Physics in 2006. In a public statement, the Nobel Committee said "the COBE project can also be regarded as the starting point for cosmology as a precision science."

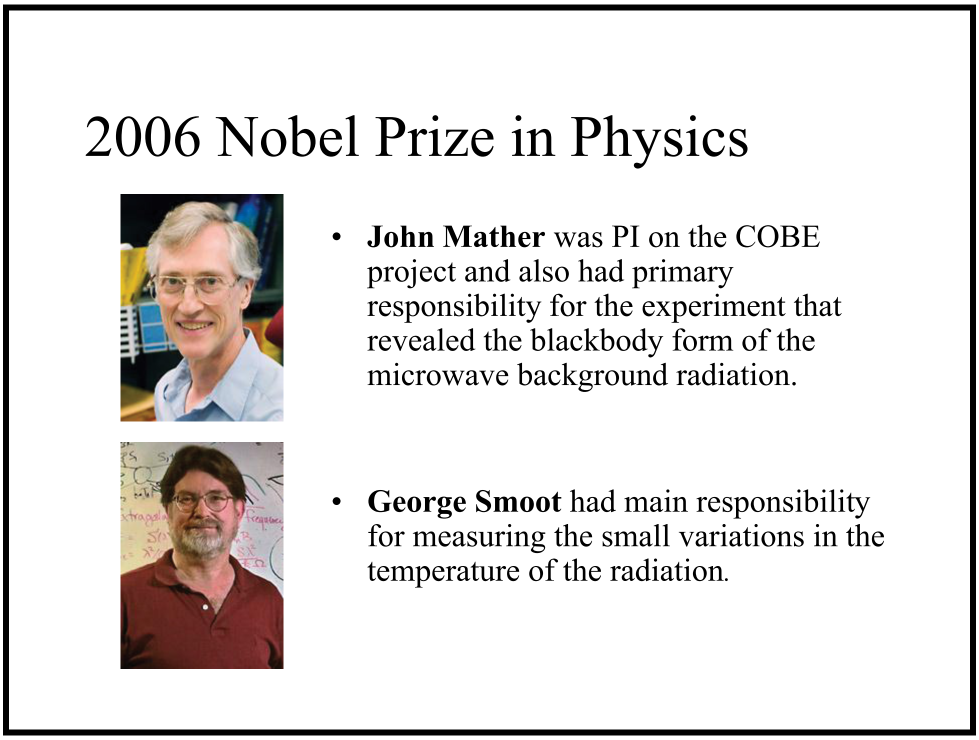


Figure 9.1 2006 Nobel Prize

9.1 WMAP

In 2003, the Wilkinson Microwave Anisotropy Probe (WMAP), a satellite instrument designed to map the sizes and strengths of the slight irregularities in the cosmic microwave background radiation, began relaying data back to earth. There have been periodic updates of the results, the last coming in 2010.

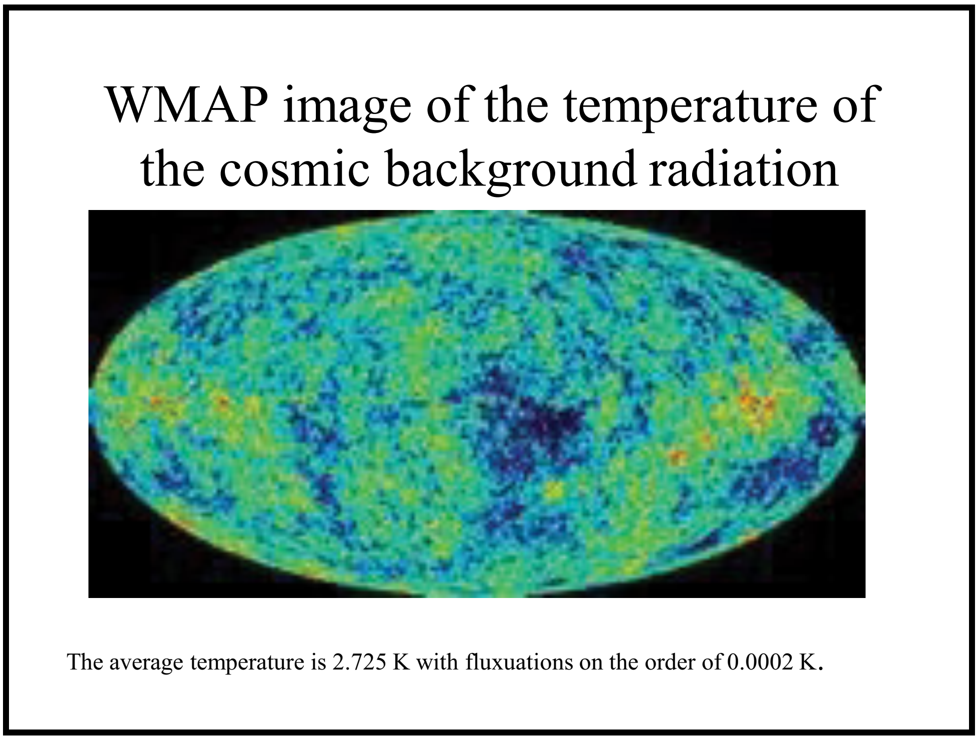


Figure 9.2 WMAP image

The whole sky image of the background radiation obtained by WMAP is one of the most spectacular images in the history of astronomy. The image is a high resolution map of the temperature of the radiation. The data represented in the map tells us the age of the universe (13.75 ± 0.11 billion years), the rate at which it is currently expanding (71.0 ± 2.5 km/sec per Mpc, where 1 Mpc = 3262 light years), the energy densities of the various components of the universe (which we will treat in more detail later), and many additional cosmological parameters.

The significance of the temperature fluxuations is that they correspond to density fluxuations of matter in the universe at the time the cosmic background radiation was produced, that is at the time neutral hydrogen and helium atoms formed in the early universe. This in turn, determines how gravity will clump the matter together to form the structures we see in the universe today, namely stars, galaxies, clusters, superclusters, and voids.

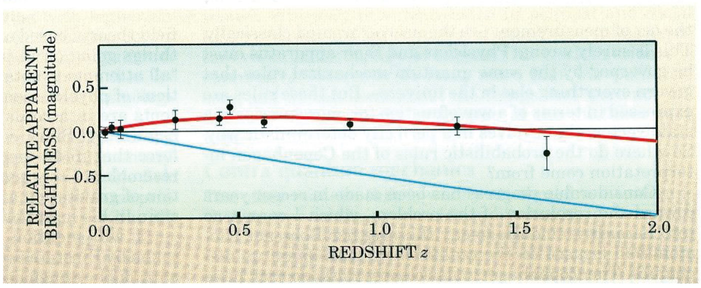
9.2 Dark Energy

In 1998, supernovae type Ia observations were carried out to measure the rate at which the expansion of space was slowing down. The fact that it was slowing down was expected due to the gravitational attraction of the clusters on one another. Instead, the data indicated that the rate of expansion was actually speeding up. This was so contrary to our understanding of the universe at that time that the scientists involved were reluctant to publish the result. They were sure that it must be wrong. It was only when the two independent teams conducting the research learned that the other team had gotten the same result that they dared to publish. Follow-up studies have unambiguously confirmed the result.

With the discovery that the expansion of space was accelerating rather than slowing down, new physics was required. Acceleration meant that there must be a cosmic repulsive force that is stronger than the attractive force of gravity. This was completely unexpected, but, as it turns out, not totally unprecedented. Back in 1916, Einstein had suggested just such a force to "fix" the problem that his General Theory of Relativity predicted that space must be either expanding or contracting. This repulsive force entered his theory in the form of the cosmological constant he added to his equations. After learning that space actually was expanding, Einstein called the inclusion of the cosmological constant the greatest blunder of his scientific career. Perhaps not.

By putting the cosmological constant back into the equations, we can explain the current acceleration of the expansion. Perhaps it is not the correct explanation (we still have a lot to learn about the nature of the force), but the calculations based on it are consistent with everything we know at this time. No matter what the exact nature of this repulsive force is, the effect it has on the universe goes by the name dark energy.

Supernovae brightness and the change in the rate at which space is expanding



The black line represents a constant expansion. The blue line represents the expected deceleration due to gravity. The red line represents the best fit to the supernovae brightness data (1998).

The data show that gravity controlled the rate of expansion of the early Universe (large redshifts). At more recent times (lower redshifts) the expansion is accelerating (positive slope). Thus, a repulsive force must now dominate gravitational attraction.

Figure 9.3 Supernova brightness

We believe that the force associated with the dark energy has been constant throughout the history of the universe. On the other hand, the effect of self-gravity has not been constant. Earlier in the history of the universe, the clusters of galaxies were closer together than they are today, and, thus, the attractive force between them was stronger. It was only after the universe had expanded for some time that the force of the dark energy exceeded that of gravity and the expansion of space began to accelerate. The data indicates that this occurred about 5 billion years ago.

The components of the universe that currently have significant consequences for the rate of the expansion of space are electromagnetic radiation, matter (both dark and ordinary) and dark energy. In order to compare the effects of self-gravity to the effects of the repulsive force, the two, must be express in the same units. The contributions of radiation and matter need to be expressed as energy densities rather than mass densities.

The following figure represents the relative values of these densities as the universe ages. The horizontal axis represents time, but is express in terms of a scale factor representing the distance between fixed coordinates in space as the universe expands. The value R = 1 is the present value and values less than one, represent the past when the fixed coordinates were closer together. The energy density of the radiation decreases more quickly than the matter density because both the number density and the energy per photon decrease with expansion, while for matter only the number density decreases, the energy (mass) is unaffected by expansion.

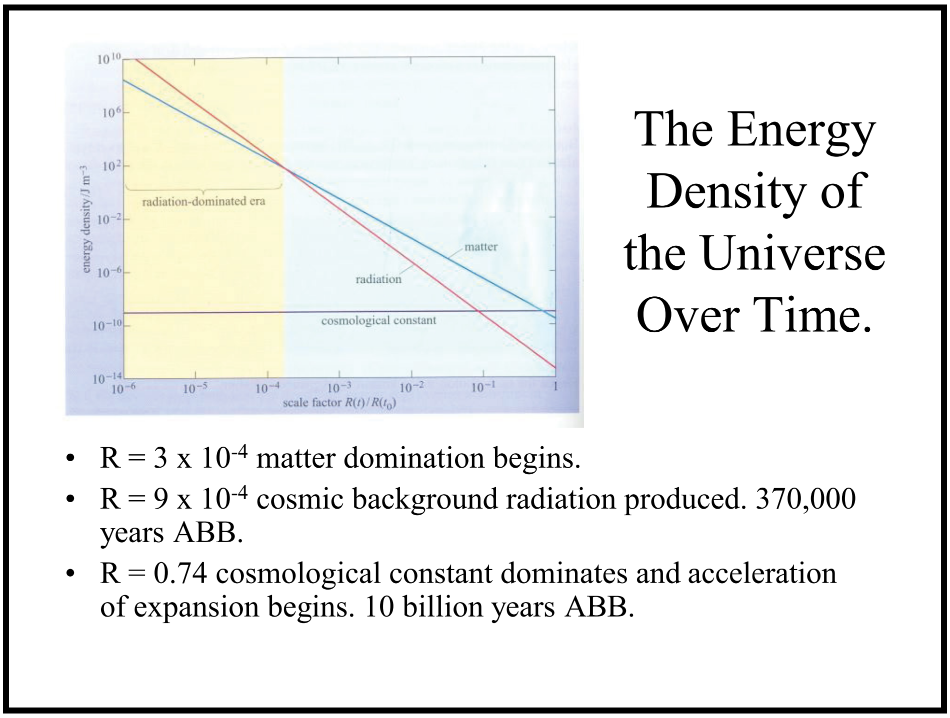


Figure 9.4 Energy density

The evolution of the universe over the past 13.7 billion years is represented in the following figure. The horizontal axis represents time while the vertical axis represents the size of the visible universe.

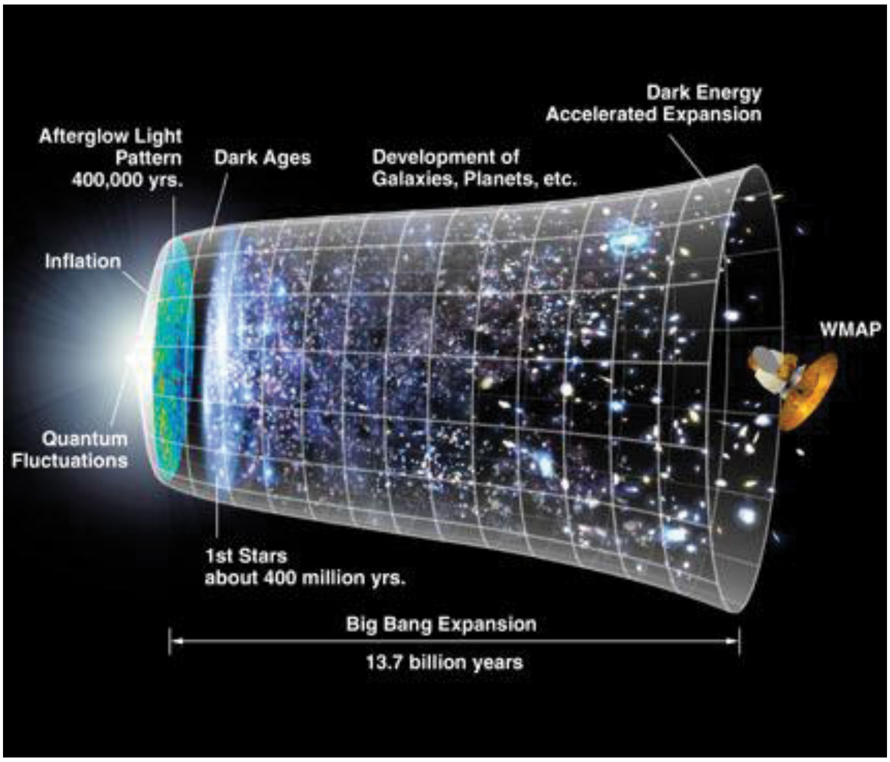


Figure 9.5 Expansion vs time

On the far left of the picture, inflation briefly produces exponential expansion. For the next several billion years, the expansion of the universe gradually slows, due to gravitational attraction. About 5 billion years ago, when the repulsive effect of dark energy overpowered the universe's self-gravity, the expansion began to speed up. Note the inflection point (the point where the curvature of the enclosing lines change) that represents this.

9.3 The Contents of the Universe

When we think of the contents of the universe, what comes to mind is ordinary matter and electromagnetic radiation. These are the components that are visualizable. Of course, it is much more complicated than that. There is dark energy responsible for the acceleration of expansion. There are ghostly particles know as neutrinos filling the universe, very small amounts of anti-matter, as well as very small amounts of matter consisting of particles made of quarks, but more massive than protons and neutrons. These particles have mass and contribute (though negligibly at the present time) to the self-gravity of the universe. There is also another kind of dark stuff, one with significant consequences for the universe. It is called dark matter. Dark matter is matter that is inferred to exist from gravitational effects on visible matter, but is undetectable by emitted or scattered electromagnetic radiation. Hence, the term dark.

Dark matter was postulated by Fritz Zwicky in 1934 to account for the fact that the galaxies in a cluster of galaxies could not have sufficient mass in the form of luminous stars to hold the clusters together. The speeds of the galaxies were simply too great for the gravitational forces to prevent them from flying off in all directions, dissolving the clusters. He realized that this additional mass must be many times greater than the mass from luminous matter.

For 40 years after Zwicky's initial observations, there were no corroborating observations indicating the presence of non-luminous matter. Then, in the late 1960s and early 1970s, a young astronomer at the Carnegie Institute, Vera Rubin, presented measurements of the star velocities in spiral galaxies. The velocity distribution of the stars, as distance increased from the centers of the galaxies, implied that the mass of the galaxies, rather than being concentrated in the center of the galaxies where most of the stars are located, was more uniformly distributed throughout the galaxies. Furthermore, these velocities were far too great for the galaxies to hold together due to the self-gravity of its stars. Whereas Zwicky used the term 'missing mass' for this non-luminous matter, Rubin preferred 'dark matter.'

Dark matter is not to be confused with dark energy. They are completely different. The most obvious difference is that dark energy is repulsive while dark matter is acted on by the attractive force of gravity. The exact nature of the dark matter is not known at the present time, but several possibilities have been suggested.

At first, theories focused on non-luminous objects made of ordinary matter (protons, neutrons, and electrons) such as neutron stars, brown dwarfs (low mass star-like objects), black holes, and/or faint white dwarfs. These objects were generally referred to as 'massive compact halo objects, or MACHOs. These objects exist, but it is now clear that they constitute only a very small fraction of the dark matter. Ironically, one of the leading contenders for dark matter is 'weakly interacting massive particles' or WIMPs. They are hypothetical particles predicted by theory, but not yet observed. The fact that the charge on these particles is zero would explain why they neither emit nor absorb electromagnetic radiation. The true nature of dark matter is one of the most important questions in astronomy today.

Prior to WMAP, we knew the categories of the contents of the universe, but did not know precisely the percentages of each. The WMAP data provided this, both as they exist at the present and as they existed at the time the cosmic background radiation was released.

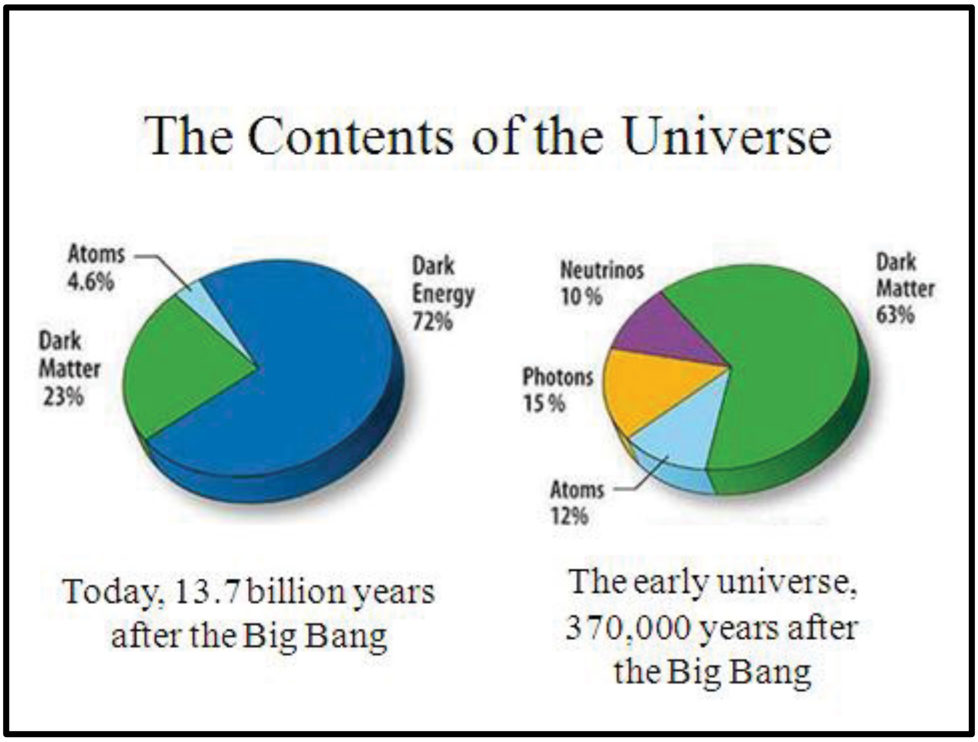


Figure 9.6 Contents of the Universe

More precisely, atoms constitute 4.56 ± 0.16 percent, dark mater 22.7 ± 1.4 percent, and dark energy 72.8 ± 0.5 percent, of the energy density of the universe. As expected, only matter and dark energy contribute significantly to the energy content at the present time. Also as expected, the dark matter contributes much more than the ordinary matter of protons, neutrons, and electrons.

Shortly after the Big Bang, dark energy made essentially no contribution, as can be seen in the earlier plot 'The Energy Density of the Universe Over Time.' Since decoupling occurred after the radiation-dominated era, the contribution of dark matter and atoms is greater than that of photons. Neutrinos, though their contribution is insignificant today, were sufficiently dense then to account for 10% of the energy of the universe. The following figure is an alternative representation of the present energy density pointing out that most of the ordinary matter in the universe is in the form of free hydrogen and helium, with most of the rest in the form of stars.

9.4 The Geometry of the Universe

One of the most profound insights of the General Theory of Relativity was the conclusion that the effect of gravitation could be reduced to a statement about the geometry of spacetime. In particular, Einstein showed that mass caused space to curve, and objects travelling in that curved space have their paths deflected, exactly as if a force had acted on them. This result eliminated the troublesome concept of gravity as an action-at-a-distance.

In the General Theory of Relativity, space and space-time are not rigid arenas in which events take place. They have form and structure which are influenced by the matter and energy content of the universe. Matter and energy determine the curvature of space (and space-time). The curvature of space tells matter how to move. In particular, small objects travel along the straightest possible paths in the curved space (space-time).

In curved space the rules of Euclidean geometry don't hold. Parallel lines can meet, and the sum of the angles in a triangle can be more, or less than 180 degrees, depending on how space is curved. In the General Theory of Relativity, there are three possibilities for the curvature of the universe: zero, positive, and negative. As we saw in discussing the balloon model, we cannot visualize curved three dimensional space, so we will again imagine a two dimensional space. Flat is like the surface of a flat piece of paper, positive curvature is like the surface of a sphere, and negative curvature is like a saddle, or to use a more up-to-day example, like a Pringle potato chip. One property of the curvature is the effect of the space on extended, locally parallel lines. In flat space, they remain parallel, in positively curved space, they converge, and in negatively curved space, they diverge.

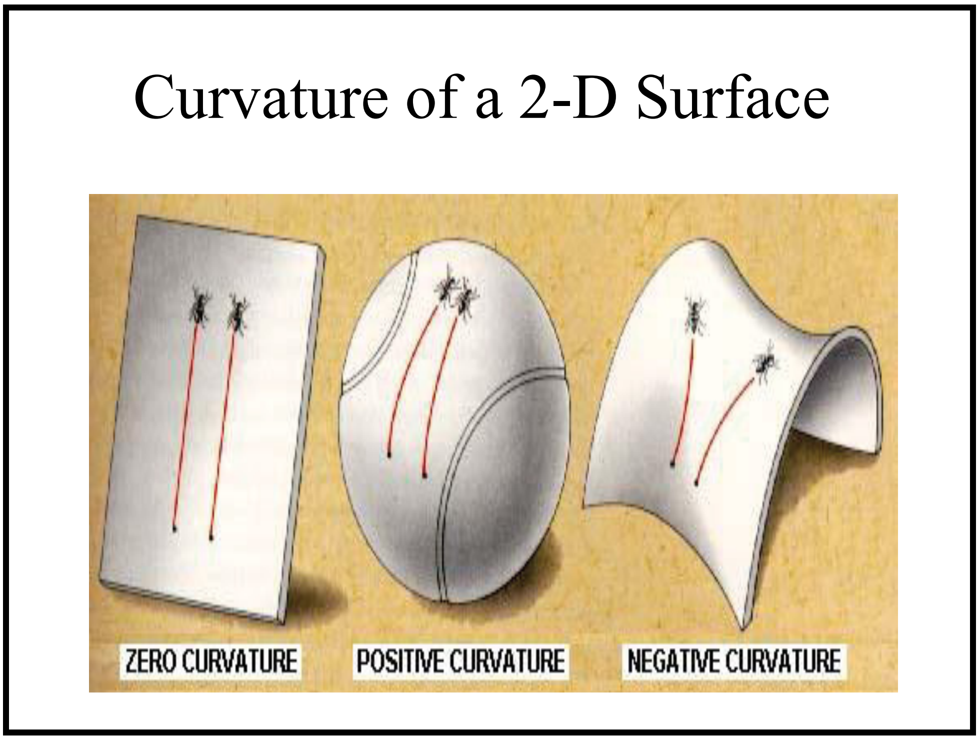


Figure 9.7 Curvature of a 2-D Surface

It is the energy density of the universe that determines whether the global curvature of space is flat, positive, or negative. The density that would make the universe flat is called the critical density. If the density is greater than the critical density, the curvature is positive, and the effect of gravity is sufficient to eventually halt the expansion and result in contraction. Such a universe is called a closed universe. If the density is less than the critical density, the curvature is negative, and gravity will not be strong enough to stop the expansion and the universe will expand forever. Such a universe is called an open universe.

The curvature of the universe is a constant property. If it started out open or closed, it will remain open or closed forever. Thus if we knew the geometry at any time in the history of the universe, we know the fate of the universe. It turns out that the properties of the cosmic background radiation depend on geometry and thus provides a test for the geometry of our universe. It has been clear for quite awhile that we do not live in a closed universe. The only question being whether our universe was flat or open. Prior to the discovery of dark energy, the evidence very strongly favored an open universe. Well before the WMAP data was available, theoretical considerations made it clear that measurements of the fluxuations in the temperature (and hence in the density of matter) at the time of the formation of the cosmic background radiation would determine whether we live in an open or a flat universe.

In the following figure, the red line represents the theoretical distribution of fluxuations in a flat universe, while the blue line represents those in an open universe.

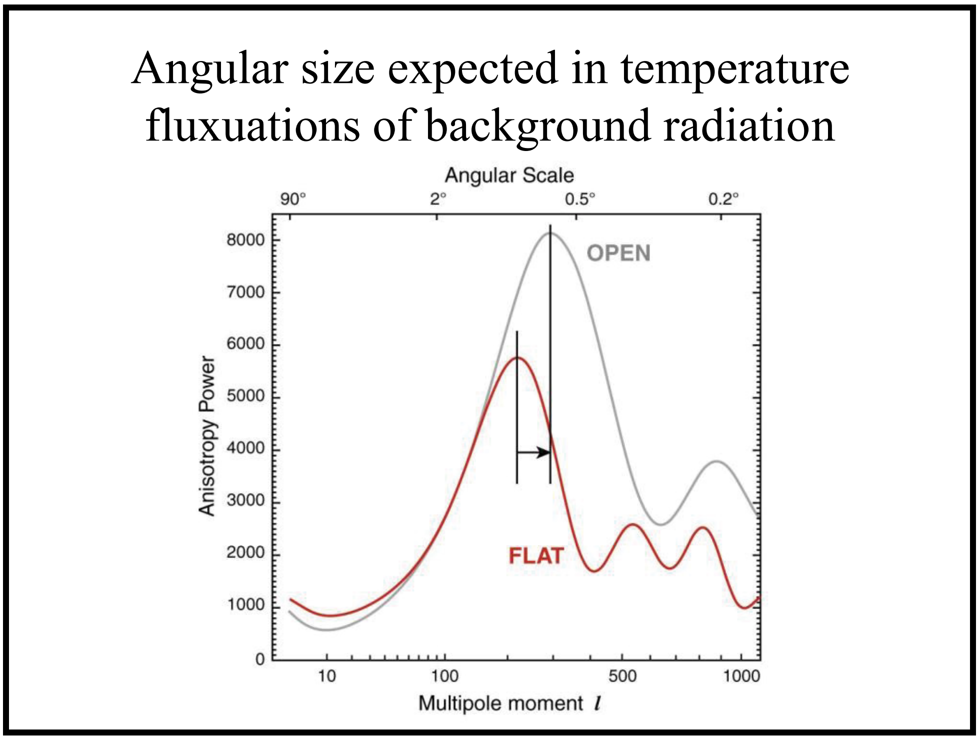


Figure 9.8 Angular size of temperature fluxuations

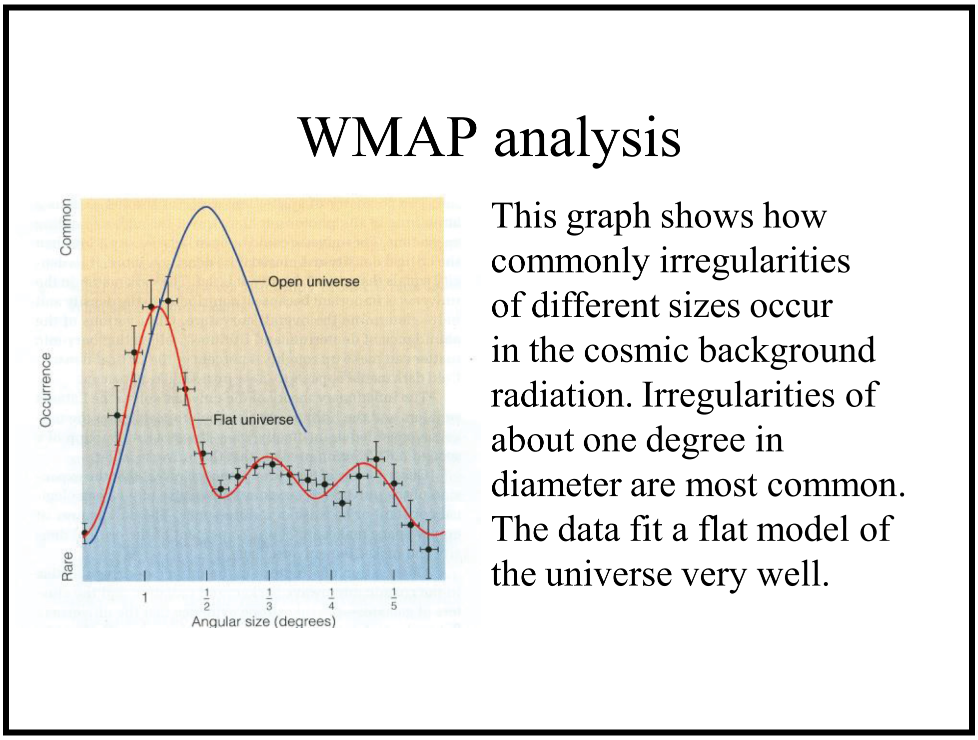


Figure 9.9 WMAP analysis

The density fluxuations were produced by sound waves traveling through the matter. (Sound waves are simply oscillating densities in some medium, whether it is air on earth or the matter of the universe itself.) The frequency of the sound waves and their harmonics is a function of, and therefore a measure or, the density of the matter. Thus by measuring the patterns of the temperature fluxuations, the angular size of the fluxuations, we are in effect measuring the density of the universe at the time of the formation of the cosmic background radiation. Calculations predict that in a flat universe the largest fluxuations will have an angular size of about one degree, while in a closed universe, it would be about 1.5 degrees and in an open universe, about 0.5 degrees. Compare the figure below with the earlier whole sky image obtained by WMAP. The image strongly favors a flat universe.

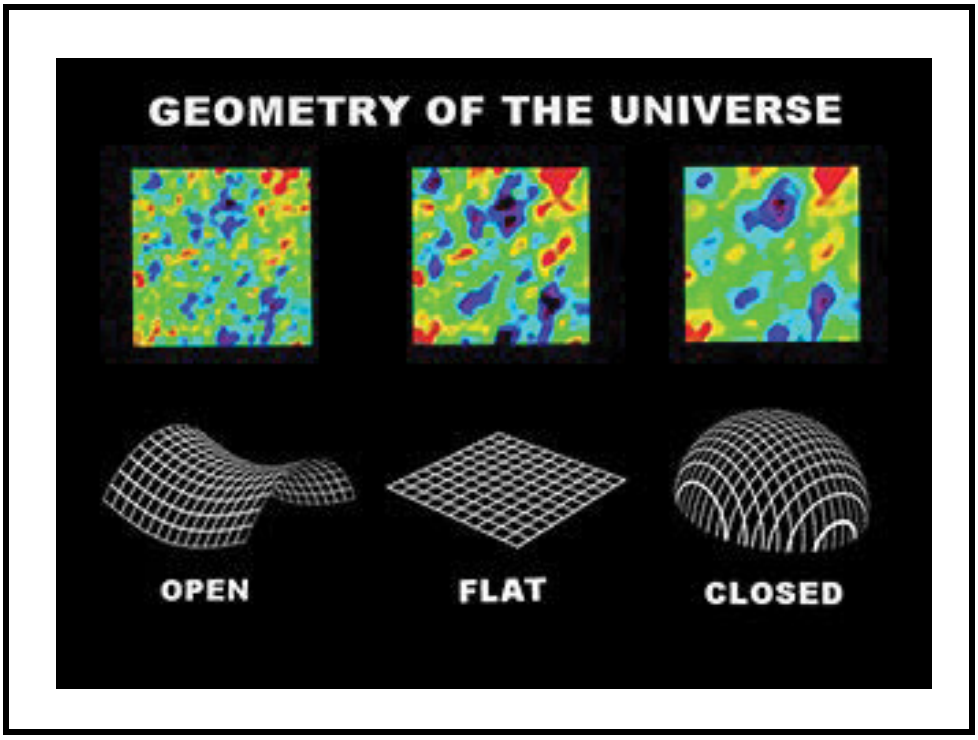


Figure 9.10 Geometry of the Universe

This result is important for a number of reasons. Back in the early 1980s when the inflationary model of the universe was first proposed, it was generally thought that we lived in an open universe. The estimated density of our universe was not even close to the critical density. Yet in order for the inflationary model to be correct, the universe had to be at least very close to flat. It now seems quite clear that the universe is flat. Remember that confirmation of a prediction in science is not proof of the theory, but it does constituted strong evidence in its favor.

More recent data on the cosmic background radiation have been obtained by the Planck probe launched by the European Space Agency. The Planck data confirm the WMAP results and its greater resolution reduces the uncertainty of some of the cosmological parameters.

9.5 Cosmological Models

The model that best explains the known data on the nature of our universe is the Lamda-Cold Dark Matter (ΛCDM) model. The lamda is the cosmological constant indicating the existence a repulsive force (dark energy) in the universe. There was initially some discussion on whether the dark matter in the early universe moved at speeds close to the speed of light (hot dark matter) or at non relativistic speeds (cold dark matter). The WMAP results are inconsistent with the hot dark matter hypothesis.

The ΛCDM model includes the following features:

- It assumes the cosmological principle that our observational location in the universe is in no way unusual or special; on a large enough scale, the universe looks the same in all directions (isotropy) and from every location (homogeneity).
- It assumes that the correct interpretation of Hubble's Law is that space is expanding.
- It assumes a flat spatial geometry.
- It assumes that the particles of dark matter had non-relativistic velocities (far below the speed of light) and consist of matter other than protons and neutrons, to be uncharged, thus not interacting with electromagnetic radiation, and to interact with each other and other particles only through gravity.
- The model includes a single originating event, the "Big Bang" or initial singularity, which was not an explosion but the abrupt appearance of expanding space-time.

The "Big Bang" scenario, with cosmic inflation and standard particle physics, is the only current cosmological model consistent with the observed continuing expansion of space, the observed distribution of matter, the observed abundances of lighter elements and their isotopes in the universe, and the spatial texture of minute irregularities (anisotropies) in the cosmic background radiation. Cosmic inflation is necessary to address certain problems that cannot be explained by the Big Bang model alone, such as why the cosmic background radiation is independent of direction in space. That is to say, radiation coming from opposite directions in space is identical. Without inflation, these regions could never have been in causal contact with one another and, therefore, could not have been in thermodynamic equilibrium as the observations require.

10 Stellar Evolution

Your goals for this chapter are to know the following:

- List the important properties of the sun along with the range of these properties in other stars.
- Describe the conditions at the center of the sun in terms of temperature, chemical composition, and state of ionization.
- Define the following terms: hydrogen fusion, iron core, supernova.
- Discuss the evolution of a 1-solar-mass star from birth to death.
- What is the most significant way in which the evolution of a high-mass star differs from a low-mass star?

We see the process of birth, ageing, and death at work in everything living. However, until recently, we viewed this drama as unfolding against an unchanging backdrop. The stars were seen as an immutable constant. Now we know that they too, in a very real sense, are mere mortals. Within our galaxy, stars are constantly born, live out their lifetimes, and die.

Our galaxy and others are collections of stars of all ages. The birth and death of stars is an ongoing process. Stars are formed by the gravitational collapse of interstellar gas and dust.

Figure 10.1 Galaxy

10.1 The Birth of a Star

When the first stars and galaxies formed some 13 or so billion years ago, not all of the available primordial hydrogen and helium was used up in the process of star formation. In galaxies such as ours, vast clouds of gas remained in the disk as interstellar matter. These first-generation stars evolved producing in their cores nuclei of elements heavier than helium. In their explosive deaths, these, along with even more massive nuclei produced in the explosion itself, were blasted into interstellar space, enriching the primordial hydrogen and helium with heavier elements. This enriched interstellar material is the raw materials from which later generations of stars have formed.

Stars are simply hot, dense concentrations of interstellar matter. In general, clouds of interstellar matter do not change much in size. They are in hydrostatic equilibrium, meaning that their tendency to contract due to gravity is balanced by their tendency to expand due to temperature and turbulence. Star formation begins when this balance is disrupted and gravity gets the upper hand. Observational and theoretical considerations suggest several different mechanisms that might cause this.

Interstellar Matter

- Chemical nature: mostly hydrogen, most of the rest is helium, and some heavier elements.
- Physical nature: mostly gas, some solid particles.
- Origin: hydrogen and helium from the Big Bang, the heavier elements were produced in the cores of massive stars and ejected into interstellar space by supernova explosions.

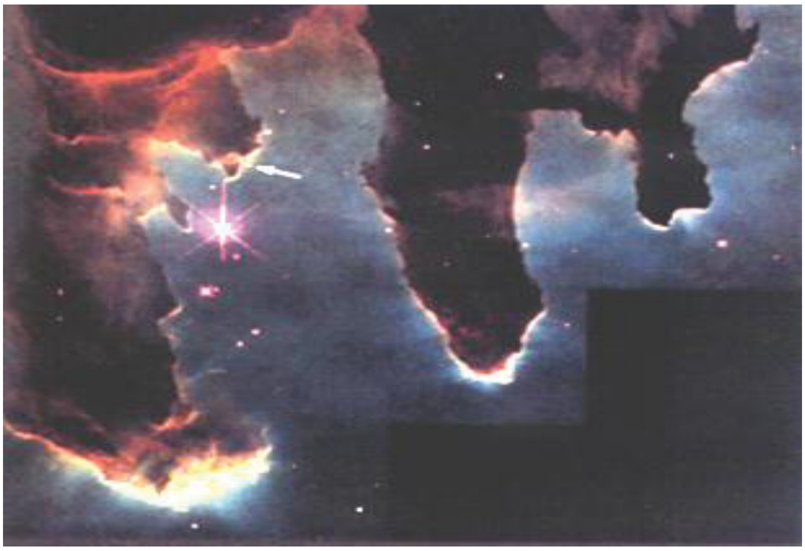
Figure 10.2 Interstellar Matter

The gravitational collapse of a cloud of gas and dust has two important consequences. First, while the cloud was still large, its rotational motion was not obvious. However, as the cloud collapses, it spins faster and faster, much as skaters spin faster and faster when they pull in their arms. The physical explanation for this is conservation of angular momentum. As the rotating mass becomes more and more compressed, its rotation speed increases, keeping the angular momentum constant. The rapid rotation has very important consequences for the formation of planets, which we will discuss in a later chapter on the solar system.

A second consequence of gravitational collapse is that the particles of the cloud speed up and collide more frequently as they fall inward. Temperature is simply a measure of the average kinetic energy (energy of motion) of the particles, so the collapse produces an increase in temperature. This effect is greatest at the center of the cloud. As the density of the central region increases, particles collide more often and pressure builds up. Eventually the density increases enough to trap electromagnetic radiation. This dramatically slows the collapse of the central region. Meanwhile, material from the rest of the cloud continues to fall toward the center, increasing still further its temperature, density, and pressure.

The central region is now a protostar. Although, as the name implies, a protostar is not quite a star, it has many star properties. It has a visible surface with a temperature of several thousand kelvins. If you could see this object, it would look like a large red star. Because of the high internal temperature of the protostar, the atoms collide violently enough to knock the electrons out of their orbits producing a state of ionized matter known as a plasma.

As the protostar continues to contract gravitationally, the core temperature continues to increase. When the internal temperature becomes high enough to initiate thermonuclear reactions converting hydrogen into helium, the protostar becomes a star.



Hubble photo of the star forming region in the Eagle Nebula

Figure 10.3 Eagle Nebula

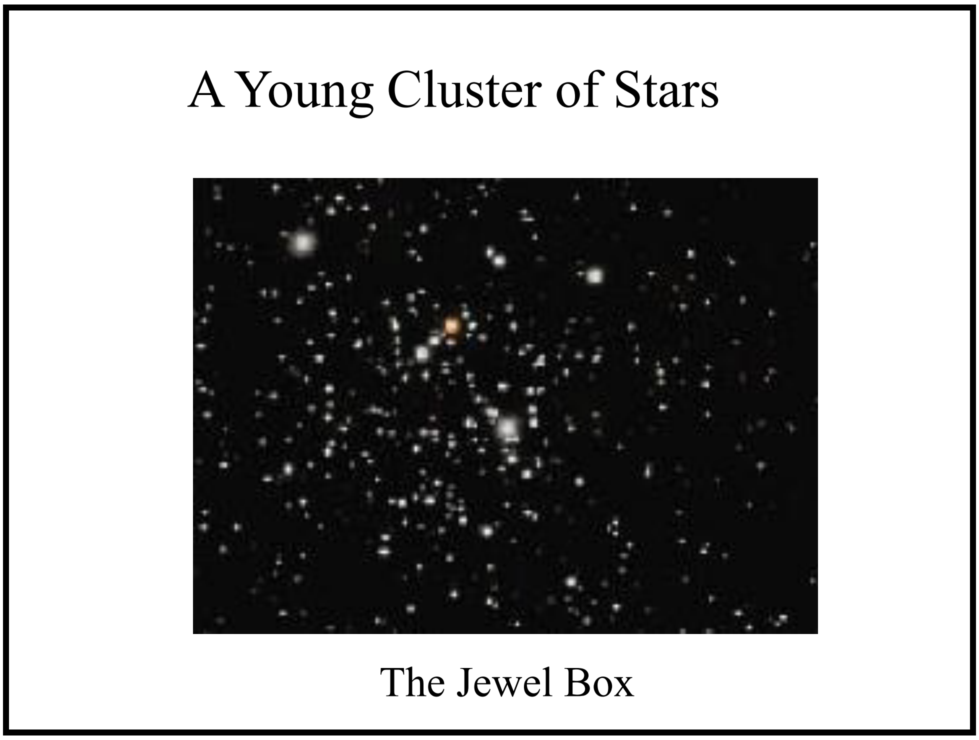


Figure 10.4 The Jewel Box

10.2 Hydrogen Fusion

In many prehistoric and ancient cultures, the sun was thought to be a supernatural phenomenon, usually personified as a god. As we might expect, it was an ancient Greek who, as far as we know, was the first to offer a natural explanation. Anaxagoras proposed that the sun was a giant flaming ball of stone rather than the chariot of the god Helios. Prior to the 19th century, speculation about the source of the energy the sun radiated centered on combustion because of the similarity between the sun and fire.

By the 19th century, the concept of conservation of energy was beginning to be understood and it was realized that if all of the sun's mass were in the form of some combustible material such as coal, it would burn out in less than 10,000 years. Clearly not enough time. Combustion, or any other chemical reaction, cannot be the sun's energy source. In the mid-1800s, steady gravitational contraction was suggested. However, this would extend the sun's life to only a few million years, and soon it was clear even that wasn't long enough.

The early part of the 20th century saw the beginning of our understanding of nuclear physics. Nuclear reactions often release relatively large quantities of energy, millions of times that released in chemical reactions. It was natural, therefore, to suspect that nuclear reactions might be involved in the production of stellar energy. In 1928, a specific nuclear reaction was suggested – the conversion of hydrogen nuclei into helium.

In order for a nuclear reaction to occur, nuclei must come close enough to one another for the strong nuclear force to pull them together. They must almost bump into one another. However, the protons (hydrogen nuclei) that initiate the fusion reaction repel each due to their positive charge (the electromagnetic force). To overcome this repulsion sufficiently to allow them to touch, they must be moving very fast, i.e. the temperature must be very high. The minimum temperature for any significant fraction of the protons to bump into one another is 10 million K. If the protostar has sufficient mass to develop this temperature in its core, a star is born. From this point until it exhausts the core hydrogen, the star is called a main-sequence star. Our sun is a main-sequence star.

If two protons collide, the strong nuclear force will bind them together temporarily. However, the very strong repulsive force that the two positive charges in such close proximity exert on each other is sufficient to break the bond that the strong nuclear force has established between them, bringing the fusion reaction to a halt. The only thing that can prevent this is for one of the protons to be transformed into a neutron by a beta positive nuclear reaction. This is an extremely unlikely reaction, and combined with the small fraction of a second the two protons are in contact, makes this reaction is very improbable. That's good, however. The improbability of the reaction allows stars to slowly consume their hydrogen fuel. Otherwise our sun would not have been the constant source of life-sustaining energy that it has been for the last 4.6 billion years.

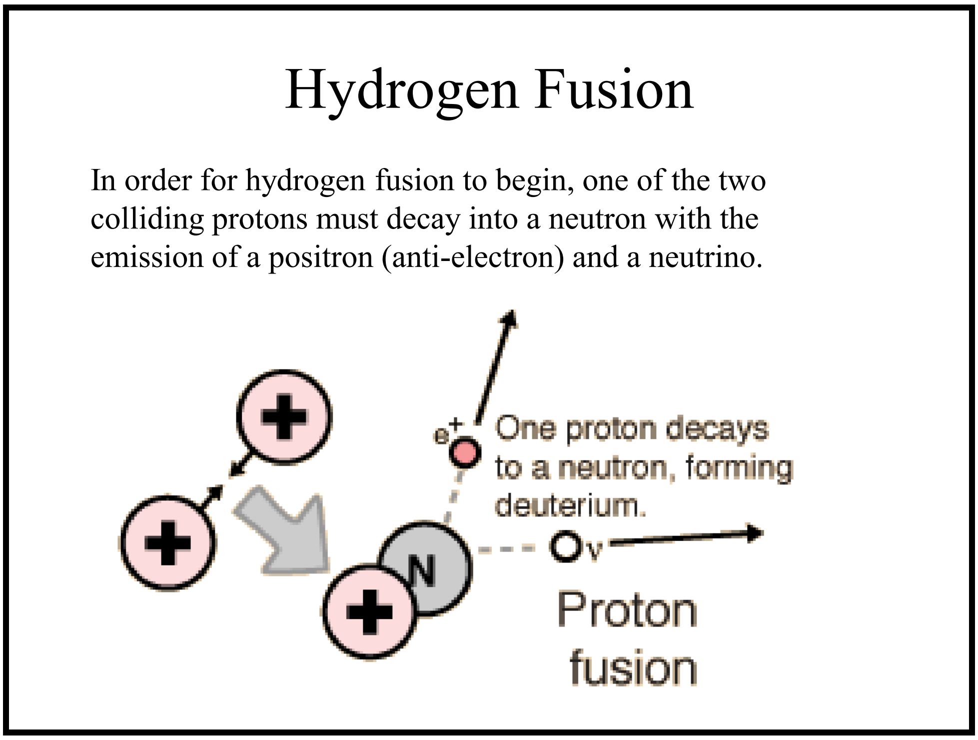


Figure 10.5 Hydrogen Fusion

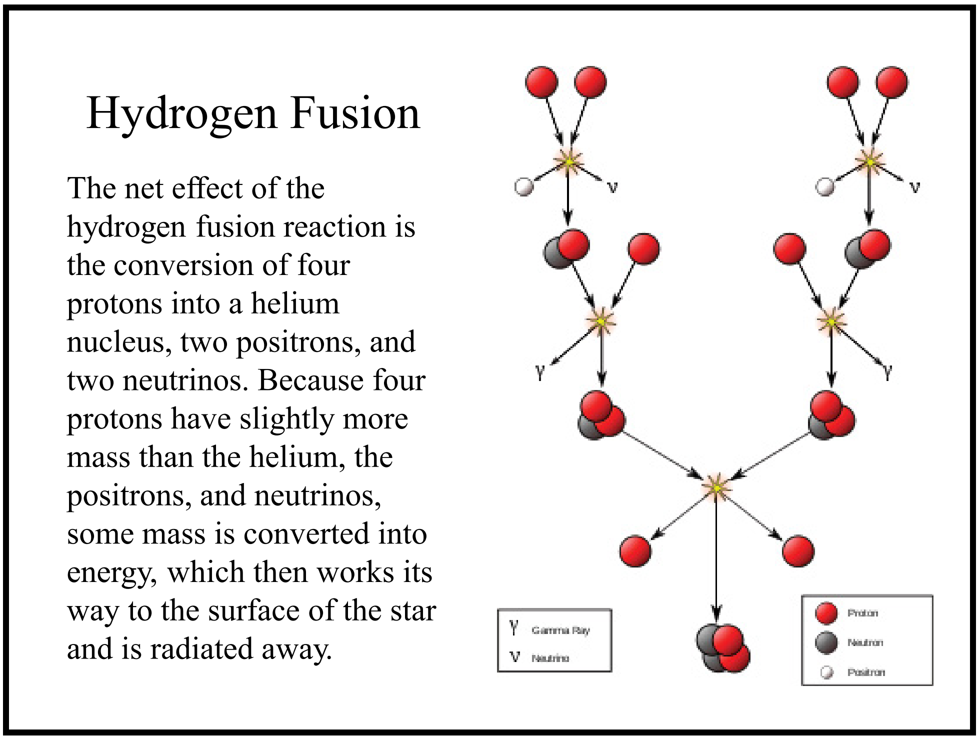


Figure 10.6 Hydrogen fusion reaction

If the beta positive reaction does occur, the resulting nucleus contains one proton and one neutron. This is a form of hydrogen known as deuterium, or heavy hydrogen. After the bottleneck of the beta decay has passed, the remainder of the reaction goes quickly. First another proton is added making helium with two protons and one neutron, known as helium-3 (for the three nucleons in the nucleus). The final step in the hydrogen fusion chain is when two helium-3 nuclei combine to form helium-4 (two protons and two neutrons), releasing two protons in the process. Overall, 4 protons have ceased to exist, and in their place is a helium-4 nucleus, two positrons, two neutrinos, and energy. Using nuclear notation, this reaction is written:

$$4({}^{1}_{1}H) \rightarrow {}^{4}_{2}He + 2({}^{0}_{+1}e) + 2({}^{0}_{0}\nu) + \text{energy}$$

The positrons are represented by the symbol 'e' with the zero superscript indicating that it is not a nuclear particle and the +1 indicating a positive charge as opposed to the negative charge on an electron, also represented by the symbol 'e'. The neutrinos are represented by the symbol 'ν', the Greek letter nu with the zeros represented a non-nuclear particle with no charge. The energy comes from the decrease in mass associated with the reaction according to Einstein's equation $E = mc^2$.

The mass of a main-sequence star determines its properties. These are usually expressed in units based on the value of the property for our sun. The subscript 'o' stands for our sun. Thus, if a star emits 25 times more energy per second than our sun, its luminosity, $L = 25\ L_o$.

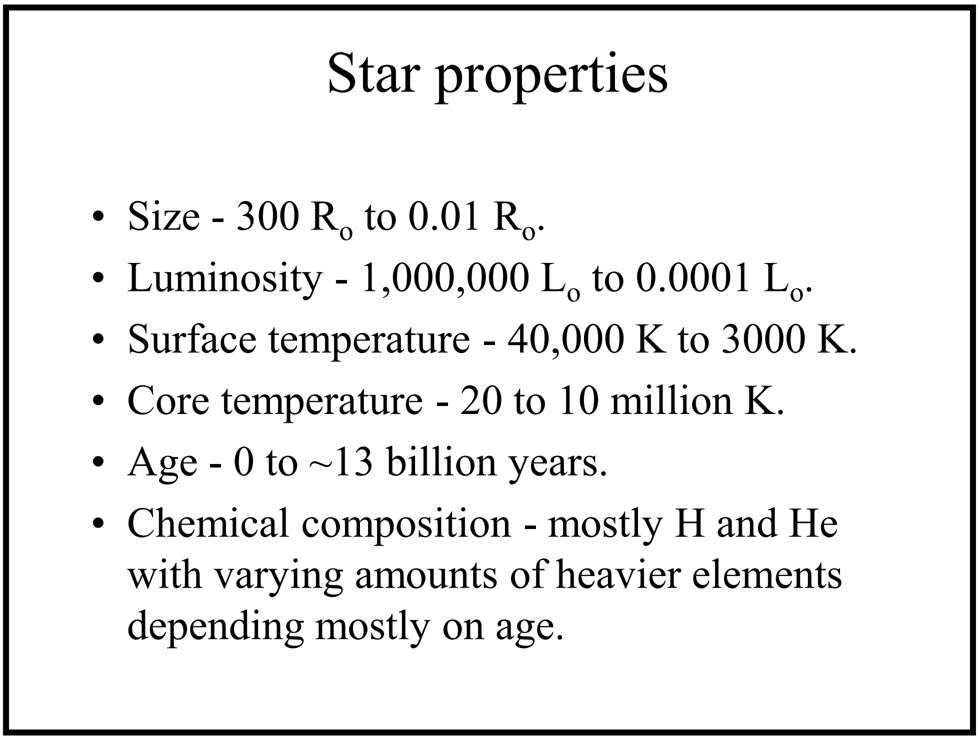


Figure 10.7 Star properties

10.3 The Deaths of Stars

Mass is the single most important property in determining how a star evolves, and particularly how it dies. With respect to their deaths, stars can be divided into three distinct mass categories: very low mass (0.08 to 0.4 solar masses), low mass (0.4 to 9.0 solar masses), and high mass (9.0 to 60 solar masses). The physics of the deaths of very-low-mass stars is complicated. It is also hypothetical in the sense that their lifetimes exceed the age of the universe. None have ever died, and they won't be discussed further.

For low-mass stars, such as our sun, as the hydrogen in the core is depleted, surface temperature and luminosity slowly change. Even though the hydrogen is being depleted, the energy produced by hydrogen fusion increases, exerting greater outward pressure, causing the outer regions of the star to expand and cool. They eventually become red giants with cores completely depleted of hydrogen, the luminosity being provided by hydrogen fusion in a shell surround an inert helium core.

Because energy is not being generated in the core, gravity contracts it causing the core temperature to increase. Eventually the core temperature becomes sufficient to fuse helium into carbon and some oxygen. Higher temperatures are required than is the case for hydrogen fusion because the helium nuclei have two protons and thus repel each other more strongly than do hydrogen nuclei.

When the helium is depleted, the core will again contract and heat up, but will not reach sufficient temperatures to fuse carbon. Low-mass stars die with carbon and oxygen cores. In the dying process, the star will begin to pulsate, and eventually the outer region will be ejected, leaving the small hot central region behind. This dead star is known as a white dwarf.

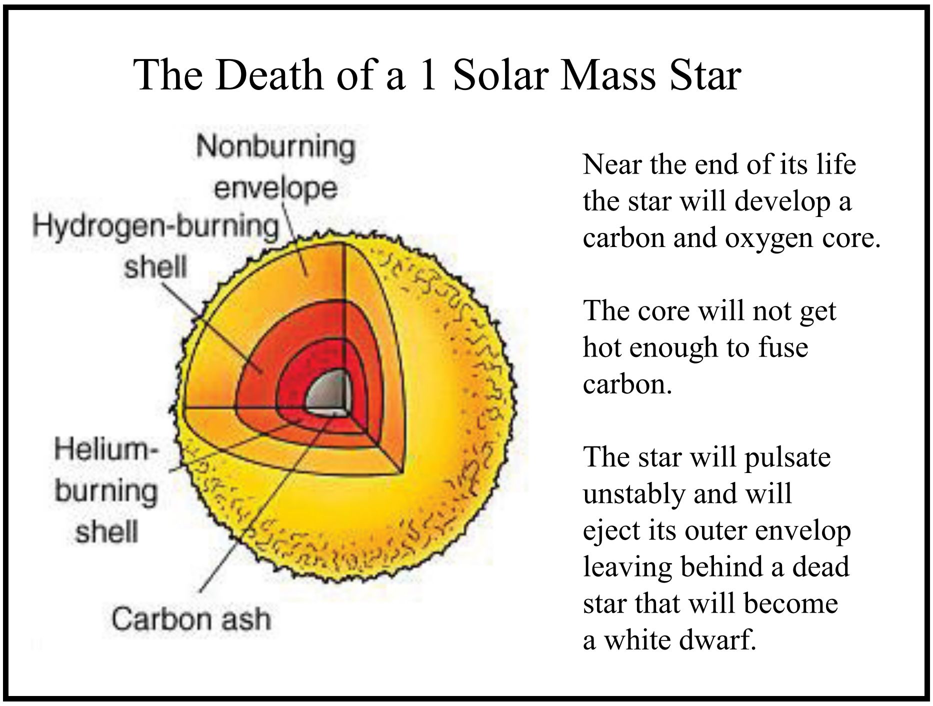
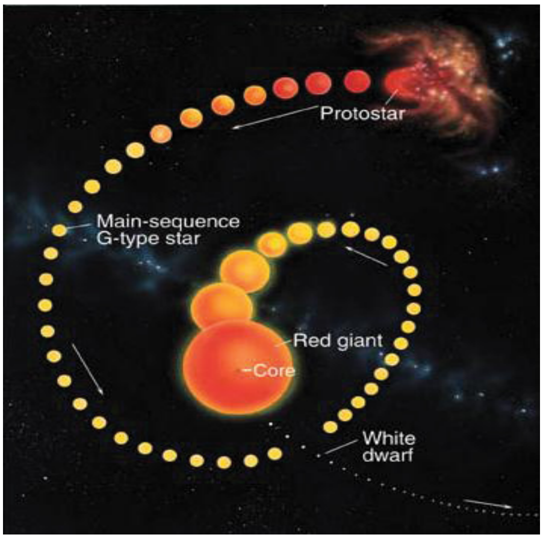


Figure 10.8 Death of a 1 Solar Mass Star

Evolution of a 1 solar mass star



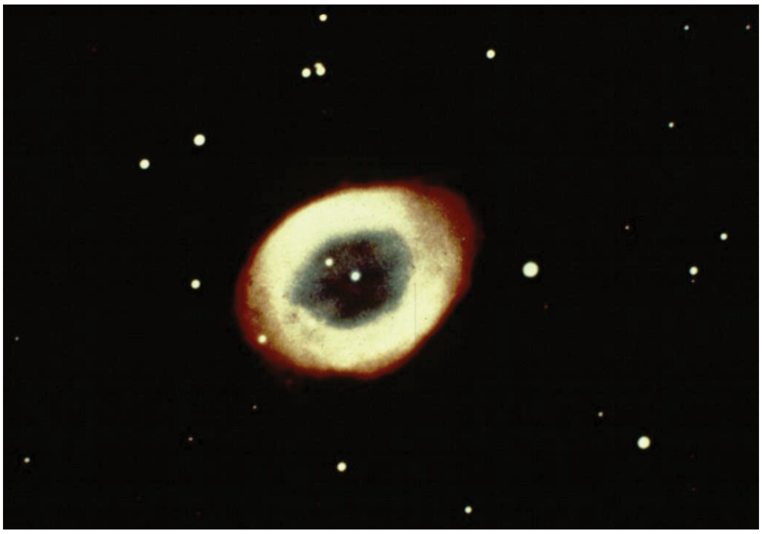
Protostar - core not hot enough for fusion.

Main sequence - hydrogen fuses to helium in the core.

Red giant - fusion in a shell surrounding the core. More energy is being generated. Star swells up and cools. Ejects outer regions.

White dwarf -dead star, no fusion. It is just cooling off

Figure 10.9 Evolution of a 1 Solar mass star



The Ring Nebula - when a star like our sun dies it ejects its outer regions. A dead white dwarf star is left behind.

Figure 10.10 The Ring Nebula

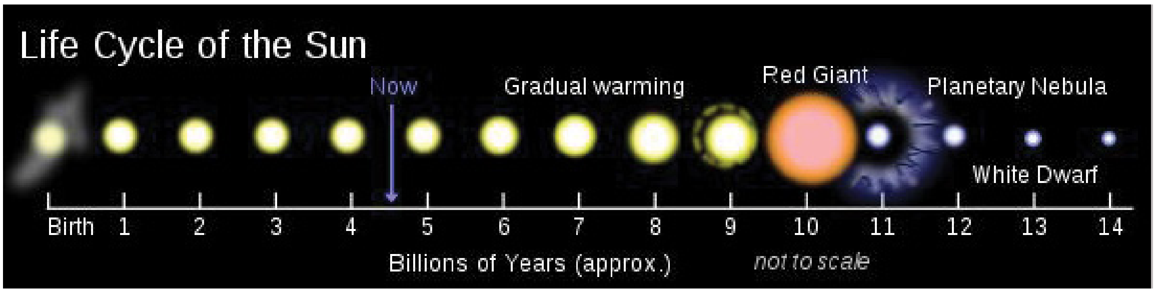


Figure 10.11 Life Cycle of the Sun

10.4 Supernovae

The dividing line between low mass and high mass stars is based on whether or not their cores reach temperatures sufficient to fuse carbon. The mass at which this occurs is not known exactly, but is generally thought to be about 9 solar masses. Once carbon fusion begins, the star will go through a sequence of fuels, contracting, heating, and fusing the next heaviest element left in the core. The star becomes layered like an onion.

The succession of fusions continues until the star develops an iron core. Up to this point, each successive fusion reaction has resulted in the liberation of energy. That is, in each reaction, the products of the reaction have slightly less mass than the nuclei initiating the reaction, and an amount of energy given by $E = mc^2$, where m is the mass lost, is released. This helps to support the star against gravity. However, iron is the most stable nucleus that can be produced in a star. As a result, when the temperature becomes sufficient for the iron to react, rather than releasing energy, energy is absorbed, producing more contraction and still higher temperatures. The result is a run-away reaction. The higher temperature causes the iron reactions to speed up, absorbing even more energy, and so on.

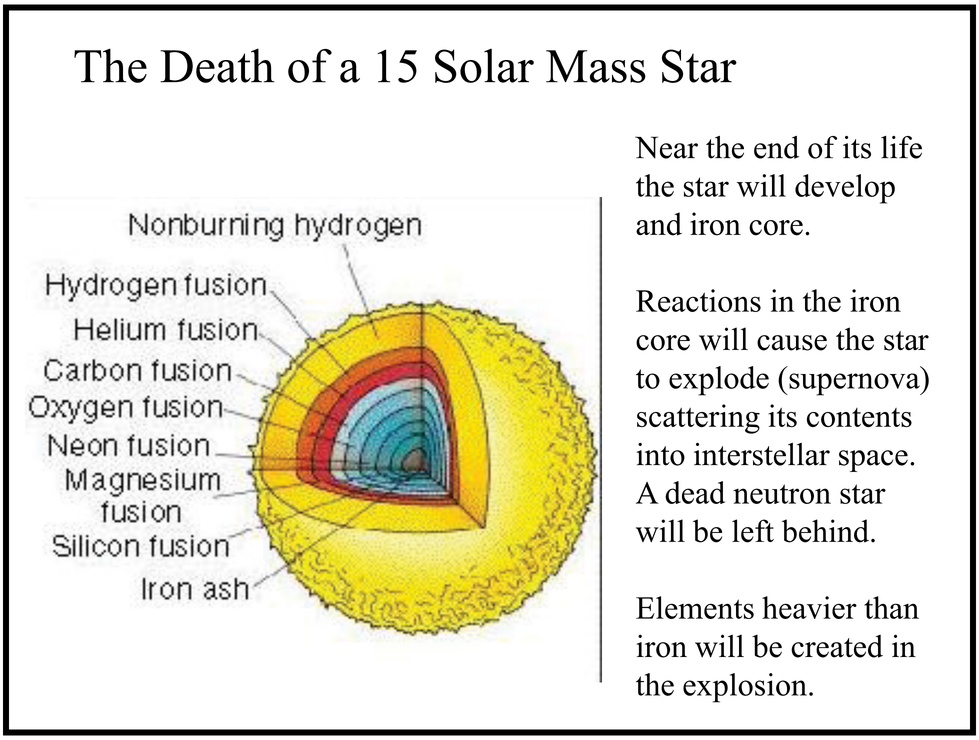


Figure 10.12 Death of a 15 Solar Mass Star

The core temperature becomes so high that the nuclei themselves are broken down into individual protons and neutrons absorbing still more energy. At still higher temperatures, electrons and protons combine to produce neutrons and neutrinos, again, an energy absorbing reaction. Because the number of protons and electrons is exactly the same (the core is electrically neutral), the core now consists entirely of neutrons with the neutrinos being radiated away from the core. In a fraction of a second, the earth-sized iron core is transformed into a neutron core about 10 miles across.

The unsupported outer regions are now free falling toward the neutron core. The implosion is converted into an explosion by a combination of events. The infalling material will hit the surface of the incompressible neutron core with speeds approaching the speed of light. Because the core is incompressible, the infalling material will bounce back at these same speeds. In addition, the pressure of the neutrinos leaving the core, and the energy generated by fusion in the infalling material, add energy to the explosion. The resulting powerful explosion is called a supernova. The energy of the explosion produces elements more massive than iron, all the way up to uranium and beyond. The supernova blasts this debris into interstellar space, enriching the interstellar material with heavier elements.

The supernova may leave behind the neutron core, in which case it becomes a neutron star. If the initial value of the star's mass is high enough (we do not currently know the value of the mass required) the neutron core will be more massive than neutron pressure can support, and the result will be complete gravitational collapse. This state of matter is known as a black hole.

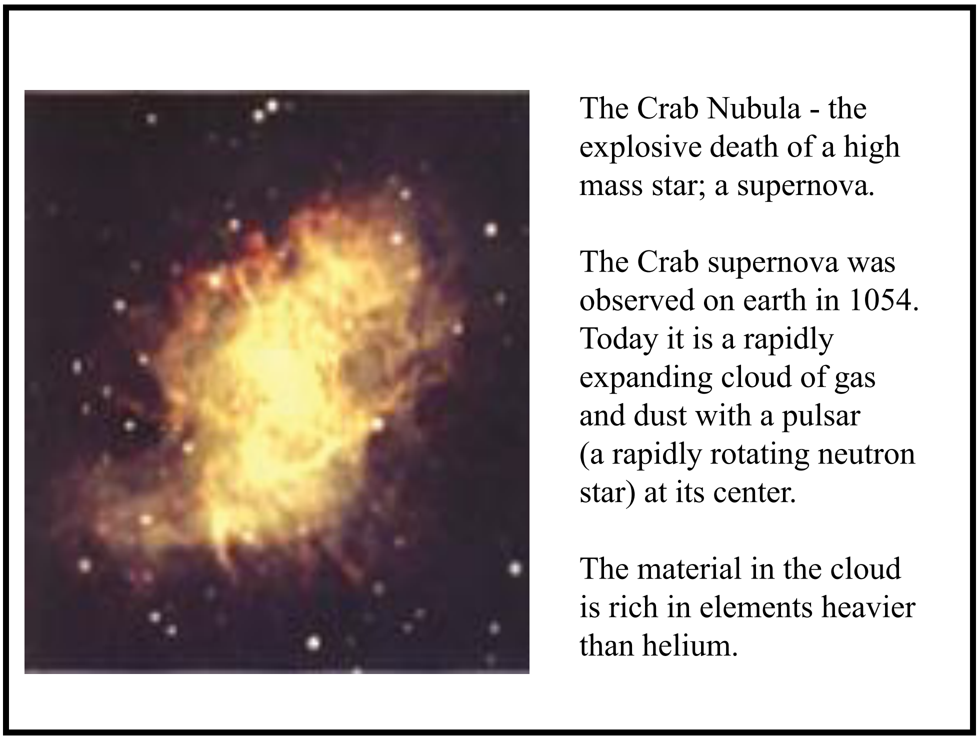


Figure 10.13 Crab Nebula

In 1967, Jocelyn Bell, then a graduate student, observed strong radio pulses coming from space. Initially she and her thesis advisor, Anthony Hewish, were baffled by the seemingly unnatural regularity of the emissions. Although they never took it too seriously, they dubbed the discovery LGM-1, for "little green men" (a common name at the time for intelligent beings of extraterrestrial origin). Soon other LGMs were discovered, and the name was changed to pulsars (a contraction of pulsating star). In 1968, Thomas Gold suggested that pulsars are rapidly rotating neutron stars emitting radiation from their magnetic poles. Because the axis of rotation differs from the axis of the magnetic field, pulsars act as a lighthouse, producing a pulse of radiation when the beam rotates past the earth. The frequency of these pulses is determined by the rotational period of the neutron star. Pulsar periods range from milliseconds to several seconds.

For our story "From Chaos to Consciousness," the most significant result of stellar evolution is that the life and death of high mass stars is the only way in which elements more massive than helium can be produced in the universe, elements such as carbon, oxygen and all the other familiar elements that make up our planet and our bodies. When Carl Sagan said "We are stardust," he is speaking quite literally.

11 The Formation of the Earth

Your goals for this chapter are to know the following:

- Outline the principle features of the nebular hypothesis for the formation of the solar system, including the two additions to this hypothesis made in the twentieth century.
- Compare and contrast the present earth with the earth just after it formed and identify the specific geological processes that produced these differences.
- How is the age of the earth determined?
- Explain the concept of plate tectonics.

11.1 The Solar System

Those bodies caught securely in the gravitational grasp of our star make up the solar system. They constitute our neighborhood, the only immediately accessible portion of the universe. We are in an age of exploration of the solar system, a time when previously inconceivable facts about our sister planets and their moons are found with each new space probe. Moreover it is now clear that the process that led to our solar system has been at work in the formation of other stars and they too have planetary systems of their own.

Just as the name implies, the largest and most massive body in the solar system is the sun. The sun dominates our existence. We have evolved sleep and waking patterns in accord with the sun's appearance and disappearance. Our calendars chronicle the seasonal relationship between the earth and the sun. In fact, our very survival depends on constant nourishment by the sun. Our mythologies, our folklore, and our pseudo-sciences all reflect our innate intuition that our lives are inextricably bound up with the solar system.

The bodies that make up the solar system are the sun, planets, dwarf planets, comets, moons, asteroids, and meteoroids. Approximately 99.9% of the mass of the solar system resides in the sun. Most of the remainder is in the form of planets. Because of the large mass of the sun, it dominates the motion of all the other objects in the solar system.

The planets and most of the remaining matter are distributed in a plane. As seen from a vantage point far north of the solar system, the sun rotates counterclockwise about its axis. The planets and asteroids lie nearly in the sun's equatorial plane and revolve counterclockwise around the sun in elliptical, but nearly circular, orbits. With the exceptions of Venus and Uranus, the planets spin counterclockwise, and moons orbit their planets in the same way. Comets are distinct in the solar system with respect to their orbits around the sun. Rather than all lying in the same flat plane, they are spherically distributed. They also have highly elliptical orbits and are just as likely to have clockwise as counterclockwise orbits. Comets are to the solar system as halo stars are to the Milky Way Galaxy.

In addition to these regularities of motion, there are certain regularities of chemical distribution within the solar system. Solar system material generally falls into one of three broad categories, gases, volatiles, or rocky-metallic materials. The gases are mostly hydrogen and helium. Volatiles include water, methane, ammonia, and carbon dioxide. The solid forms of volatiles are called ices. Rocky-metallic materials are primarily compounds of iron and silicon.

These materials are not randomly distributed within the solar system. The innermost planets and moons are almost entirely rocky-metallic in composition. The asteroids are rocky-metallic, with the outermost ones containing significant fractions of icy material. In addition to rocky-metallic and icy material, the planets Jupiter, Saturn, Uranus, and Neptune contain huge quantities of gas. In fact, Jupiter and Saturn, like the sun, are mostly hydrogen and helium. The dwarf planets and moons of the outer solar system are rocky-metallic and icy in composition.

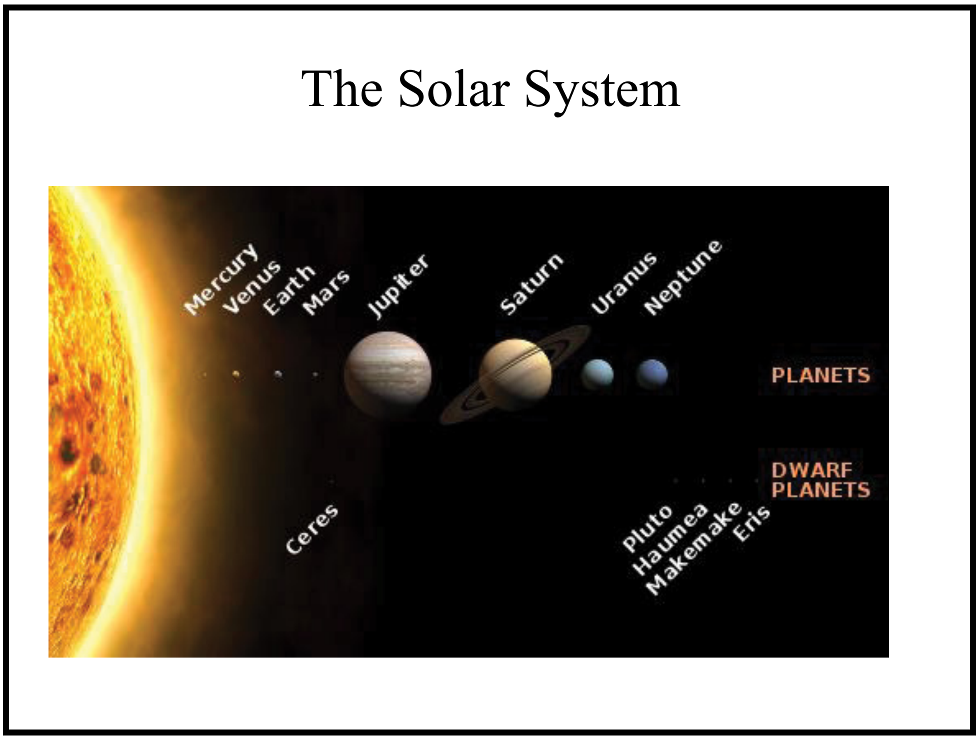


Figure 11.1 The Solar System

11.2 The Origin of the Solar System

The first scientific theory of the origin of the solar system was that of the French philosopher Rene Descartes. In 1644 he proposed that space was initially filled with swirling gas in which large whirlpools evolved into stars while the planets and their satellites formed from much smaller vortices. The German philosopher Immanuel Kant applied Newton's laws to Descartes' model in 1755 and concluded that the swirling gas would assume a disk shape. The French mathematician Pierre-Simon Laplace independently proposed a similar theory in 1796. Known as the nebular hypothesis, the gravitational collapse of an interstellar cloud accounted for the flattened appearance of the solar system, the nearly circular orbits of the planets, and the fact that the planets moved along their orbits and rotated about their axes in the same direction that the sun rotated.

By the end or the 19th century, it had become clear that there were two serious problems with the nebular hypothesis. First, the planets and moons contained insufficient mass to have formed from gravitational collapse. Secondly, the gravitational collapse of a cloud of gas and dust should result in a sun that rotated much more rapidly than our sun does. By the latter half of the 20th century, both these problems had been resolved.

We now understand that the formation of the planets and moons began not with gravitational collapse but with the sticking together of small, solid grains in the solar nebula. This process is called accretion and is exemplified by a snowball rolling down a snow bank and gaining mass as it does. Accretion also explains why dust sometimes concentrates in dust bunnies. Only after the bodies have gained sufficient mass does gravity play a part in the final formation. The problem with the sun's rotation was resolved when it was discovered that the sun constantly emits high speed charged particles. These are known as the solar wind. As the solar wind leaves the sun, it passes through the sun's magnetic field, and the particles are deflected by the magnetic force. For every action there is an equal and opposite reaction (Newton's third law of motion), and therefore, the force the sun exerts on the solar wind is accompanied by a force the solar wind exerts on the sun. This reaction force slows down the sun's rotation. In this way, over the history of the solar system, magnetic braking has slowed the sun's rotational period from its original value of less than ten hours to its present value of about 25 days.

In the process of star formation, the gravitational contraction of the interstellar cloud increases the temperature of the cloud and causes it to spins faster and faster. Both of these results played an important role in the formation of our solar system. As the contracting solar nebula spins faster and faster it begins to flattens out. The rotation caused some of the material to go into orbit around the central object, (which will eventually become our sun). The temperature increase is greatest in the center of the cloud and decreases with distance from the center. The center heats to temperatures sufficient for hydrogen fusion while the outer regions are not much hotter than the original interstellar cloud.

The temperature of the central object and the region immediately surrounding it is sufficient to evaporate all of the solid grains from the original interstellar cloud. Out somewhat further, the temperature is sufficient to evaporate the icy grains, but not the rocky-metallic grains. Still further, past the frost line, both rocky-metallic and icy grains can exist.

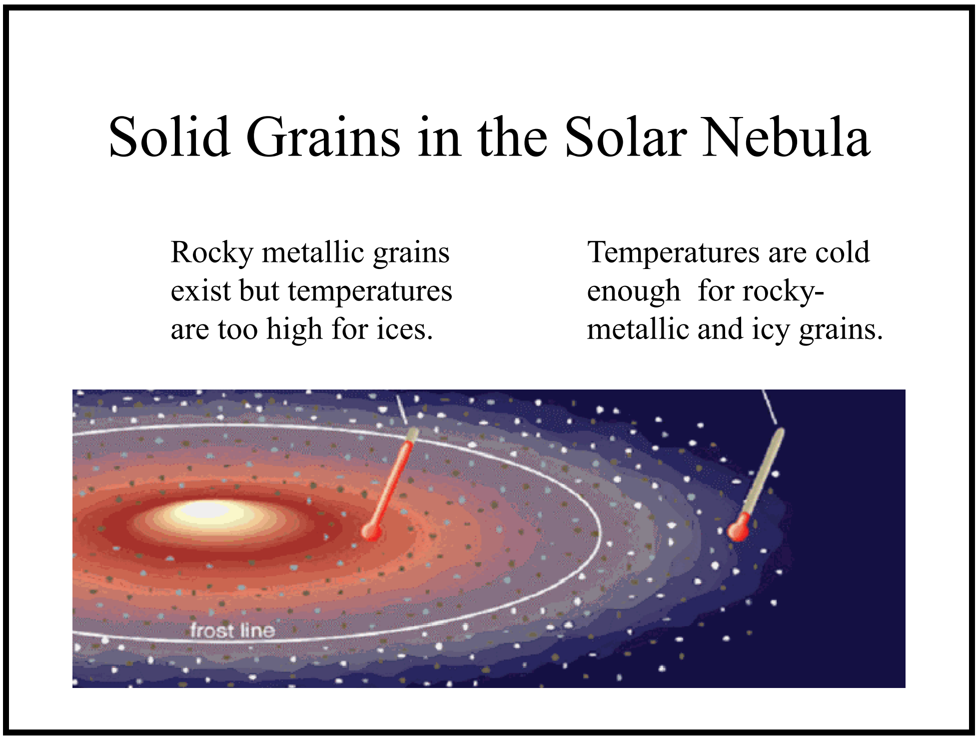


Figure 11.2 The Solar Nebula

Accretion will therefore produce rocky-metallic objects in the inner part of the solar system and rocky-metallic-icy objects in the outer part. Accretion in the inner part of the solar system produced Mercury, Venus, Earth, Mars, and, soon afterwards, our moon. Four of the rocky-metallic-icy objects in the outer part, Jupiter, Saturn, Uranus, and Neptune, had sufficient mass to gravitational capture the gas in their vicinity and are therefore composed mostly of hydrogen and helium. The moons of the four gas planets and the dwarf planets lying beyond them did not have sufficient mass to capture the hydrogen and helium and are rocky-metallic-icy objects.

The nebular hypothesis clearly predicts that stars other than our sun should have planets in orbit about them. This prediction was made well before any planet outside our solar system had been observed. The first extra-solar planet was discovered in 1988. Since then, well over 500 have been discovered with new ones being continuously added. Current data estimates that there are at least 50 billion planets in our galaxy. A small, but non-negligible, fraction of these have masses similar to earth and occupy the habitable zone (temperatures that allow for liquid water) of their planetary system. For this reason, many believe that life is common in our galaxy.

11.3 The Age of the Solar System

Humans have speculated on the age of the earth for thousands of years. The Brahmins of India believed that the Earth was eternal, as did Aristotle. In early Judeo-Christian cultures, estimates of the age of the solar system were based on the Bible. The traditional Jewish calendar starts from 3760 BCE, which is taken to be a date for the creation of the earth. In 1650, Anglican bishop James Ussher used the Jewish Bible to calculate that creation took place on the evening of Sunday, October 23, 4004 BCE of the Julian calendar.

By the 18th century, naturalist had begun to seriously doubt Ussher's value. One of the first estimates based on natural processes was made by George-Louis Leclerc, Comte de Buffon in 1774. He estimated that the time it would take for a molten earth to cool would be about 75,000 years.

By the end of the 18th century, James Hutton, the founder of modern geology, introduced the concept of 'deep time,' pushing the estimate to millions of years. His student and fellow geologist, Charles Lyell, friend and confidant of Charles Darwin, popularized the concept that geological layers were in perpetual change, eroding and reforming continuously at a roughly constant rate. (Darwin read Lyell's landmark text, *Principles of Geology*, while on the *Beagle*, and was much influenced by it.) Lyell's 'uniformitarian' challenged the traditional view that the geological history of the earth was essentially static, with changes brought about by intermittent catastrophes or supernatural interventions such as Noah's flood. Lyell estimated the age of the earth in hundreds of millions of years.

Radioactivity was discovered in 1895 and, because of the constant rate associated with the decay of a particular radioactive substance, was soon recognized as a technique for dating rocks. For example, uranium decays through a sequence of reactions to lead. Because the chemical properties of uranium and lead are different, these two elements would not be found mixed together in a rock that has just formed, say from solidified material from a volcanic eruption. As the rock ages, uranium atoms are gradually transformed to lead atoms, with the lead homogeneously mixed with the uranium. Knowing the rate at which this occurs, the relative proportions of uranium and lead is a direct measure of the time since the rock formed. Because on our geologically active earth, rocks are constantly being recycled, the ages of rocks vary, but by 1907 the age of some earth rocks had been measured in the billions of years.

Ancient rocks exceeding 3.5 billion years in age are found on all of Earth's continents. The oldest rocks on Earth found so far are in northwestern Canada and are dated at just over four billion years. Some rocks in western Greenland are 3.7 to 3.8 billion years old. These ancient rocks have been dated by a number of radiometric dating methods and the consistency of the results give scientists confidence that the ages are correct to within a few percent.

The moon is less active than the earth and some of its rocks are older than any found on earth. A small number of rocks have been returned to earth by lunar missions. The oldest of these rocks formed between 4.4 and 4.5 billion years ago, providing a minimum age for the formation of the solar system. Meteorites are fragments of asteroids that fall to earth. The oldest of these formed between 4.53 and 4.58 billion years ago. Because their structure indicates that they have never been melted, we take their age to be the time at which the first solar system solid objects formed. Thus our solar system is about 4.6 billion years old.

Another way to determine the age of the solar system is to use stellar computer models to determine how long it would take a star with the mass and chemical composition of our sun to reach its current value of luminosity and surface temperature. Although this method is not as accurate as radiometric dating, it too gives and age of somewhat less than 5 billion years for the age of the sun, consistent with the value determined from radiometric dating.

11.4 The Geological Evolution of the Earth

Earth formed by the accretion of rocky-metallic grains in the solar nebula. In its final stages of formation, aided by gravity, most of the remaining debris in its zone was swept up depositing a tremendous amount of energy on the earth's surface. At the same time, radioactive elements were decaying throughout the planet, releasing energy that, like the energy from accretion, was converted into heat. The heat could not be radiated away fast enough and temperatures rose above the melting points of the various materials.

During this molten stage, the denser metallic materials settled toward the center and the lighter, rocky materials floated toward the top. This process is known as chemical differentiation. An iron-nickel core formed, surrounded by the mantle, a layer of rocky material. The mantle extends almost to the earth's surface. The top layer of the earth is the crust, composed of rocks that are generally less dense than the mantle. The crust is several miles thick under the oceans and some twenty or so miles thick under the continents.

When the bombardment diminished, the surface of the earth solidified. Though diminished, bombardment continued leaving the surface covered with impact craters, much as parts of the moon and Mercury are today. At this time, the earth had no oceans and no significant atmosphere.

The early earth would have been unrecognizable as our future home. The processes by which this transformation occurred are known as geological evolution. Geological evolution is primarily driven by the heat produced by radioactive decay. The heat produced not only led to chemical differentiation, it contributed to the formation of our oceans and atmosphere and the elimination of craters.

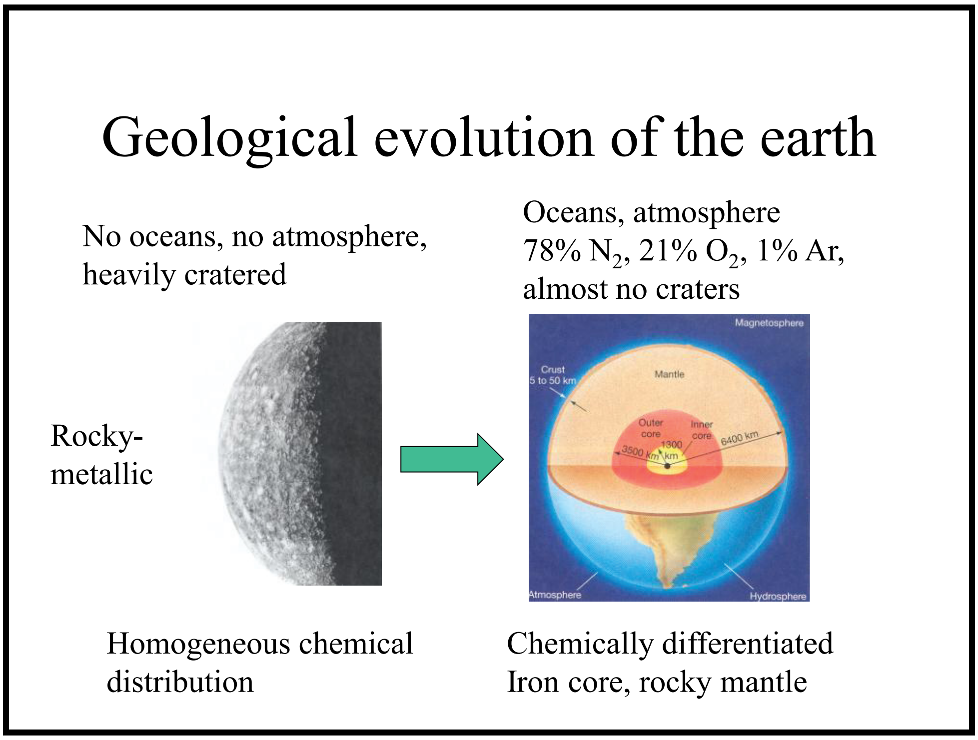


Figure 11.3 Geological evolution

Radioactivity decreases exponentially with time. As a result, the earth has begun to cool. Material that was once liquid has now re-solidified. The iron-nickel core is solid in its center with a surrounding liquid shell. It is this liquid iron and the rotation of the earth that produces the earth's magnetic field. The radius of the iron-nickel core extends to about half the earth's radius. The mantle, for the most part, is not molten. But it is not exactly solid either. It behaves somewhat like a plastic and can flow under pressure.

Liquid rock, lava, occasionally breaks through the crust, releasing into the atmosphere gasses previously trapped in the rock. This is called 'outgassing'. The two most abundantly gasses outgassed are water (steam) and carbon dioxide. At our distance from the sun, the surface temperature is such that the steam condenses to a liquid and falls back to earth as rain. This is in part how the oceans formed.

Carbon dioxide is soluble in water (soda water is an example) and dissolves into the oceans, later to precipitate out as carbonate rock such as limestone. With the two principle outgassed substances removed, the third most abundant, nitrogen, constituted almost all of the early atmosphere. Life evolved on earth and some forms were able to photosynthesize water and carbon dioxide into food (carbohydrates) with oxygen as a byproduct. Gradually the oxygen began to accumulate and today makes up about 21% of our atmosphere.

The third most abundant component of the present atmosphere is argon. The source of the argon is the radioactive decay of potassium mentioned back in Chapter 5. The argon is also released in outgassing.

Our oceans were formed in part by outgassing, but this could not have been the only source. There is too much water on earth to be accounted for by outgassing. A likely alternative source of water is comets. Comets are dirty ice balls, about 80% ice and 20% rocky-metallic materials. Early in the history of the solar system, they were found much more frequently in the inner regions than they are today. It is now well accepted that comet bombardment accounts for some fraction of the earth's oceans, thought the precise value of that fraction is unknown.

Solar system objects like Mercury, our moon and some of the moons in the outer region of the system are heavily cratered, the remnants of the impacts during the final stages of their formation. The earth is not. It might seem logical to assume that this is due to the fact that, unlike the other bodies, the earth has an atmosphere and oceans. Perhaps, weathering has worn down the craters so that they are no longer visible today. This is not the case. The craters have been removed by plate tectonics, another evolutionary mechanism produced by the heat of radioactive decay.

It you look at the globe, the continents of Africa and South America seem to fit together like two pieces of a jigsaw puzzle. The continental shelves of these continents are an even better fit. Early in the 20^{th} century, Alfred Wegner, a German meteorologist, became intrigued by this. He was struck by the similarities in the geological formations and the fossil records at the edges of these widely separated continents. Wegener, noted that the locations of certain fossil plants and animals on present-day, widely separated continents would form definite patterns (shown by the bands of color in the following figure), if the continents are rejoined. He proposed the theory of continental drift, hypothesizing that at one time in the past the continents were connected and have since drifted apart.

Wegner's idea was initially ridiculed by the scientific community. No one could conceive of a mechanism whereby entire continents could be caused to move across the face of the earth. However, as the ocean floors began to be studied in more detail in the late 1950s and early 1960s, scoffers became supporters. Geophysicists developed a theory known as plate tectonics, based on the plastic properties of the upper mantle to explain continental drift. The evidence in favor of the theory soon became overwhelming.

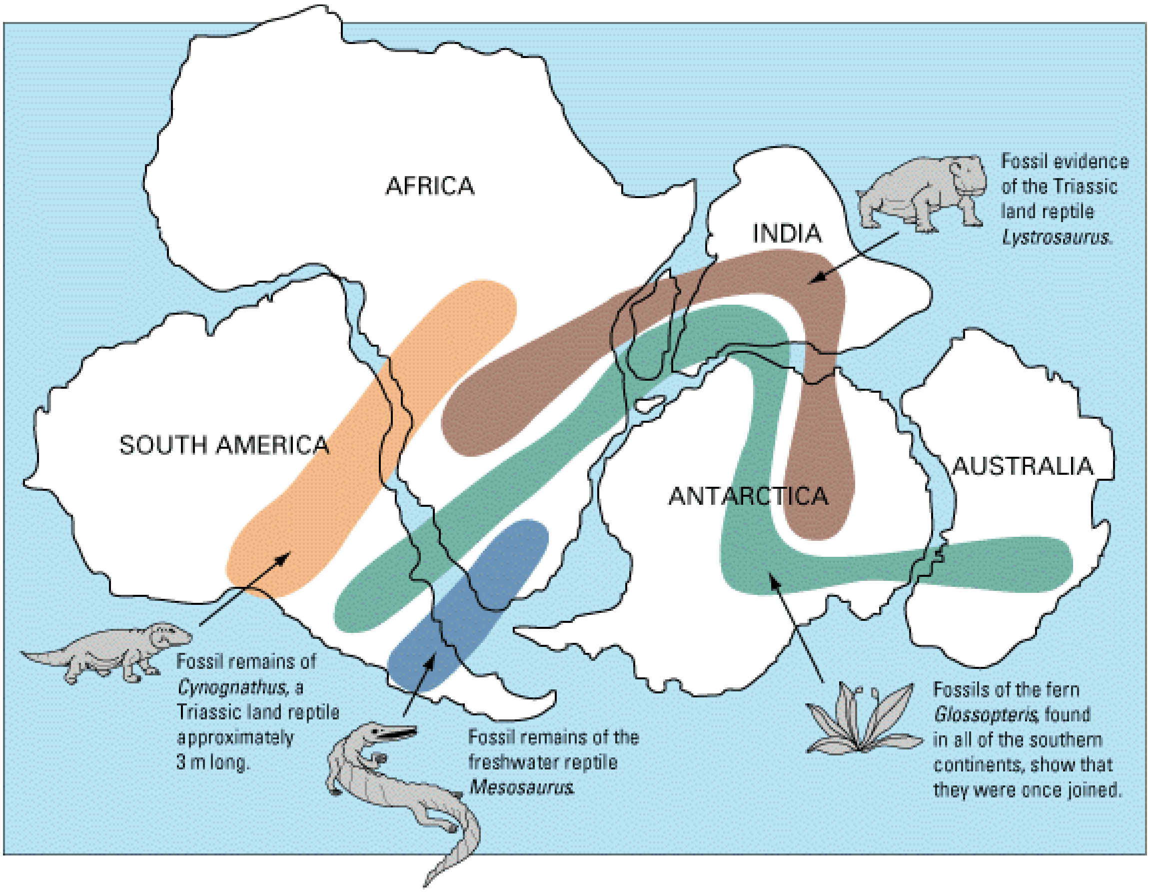


Figure 11.4 Fossil distribution

On the ocean floor, about halfway between Europe and Africa to the east and North and South America to the west, is the Mid-Atlantic Ridge. Dating of rocks at the ridge shows them to be very young. As you move out from the ridge in either direction, the rocks are progressively older. It is clear that the continents of Europe and Africa are moving apart from North and South America. The speed at which they are moving apart has been measured by satellite observations to be about an inch or so a year. In other parts of the globe, the continents are moving together and, in still other parts, moving anti-parallel to one another, all at similar speeds. If we extrapolate the motions of the continents backwards in time, we find that about 200 million years ago all of the continents were together. This single land mass is called 'Pangaea.'

The continents rest on huge plates that are being driven across the earth's surface by convection currents in the upper mantle beneath them. Where the plates collide, one of the plates is driven under the other and forced down into the mantle and melted. The other plate is driven upward forming mountain ranges. At the boundary where two plates move apart, molten material rises up forming new rocks. In this way the crust of the earth is continuously being destroyed and reformed. Although there are some rocks as old as four billion years, most are much younger. The craters formed in the very early history of the earth no longer exist because the ground they formed on no longer exists. It has been recycledby plate tectonics

The boundaries of the plates are regions of extreme geological activity: volcanoes and earthquakes. As can be seen on the map, Japan is on the boundary of two plates. Another consequent of plate tectonics is that India, which was once an island, has collided with the underside of Asia forming the Himalayan Mountains. The Mediterranean Sea is gradually being squeezed closed. Los Angles, parts of southern California, and the Baja peninsula, are located on the Pacific plate and are slowly moving north with respect to the North American plate, on which the rest of California is located. The boundary is known as the 'San Andreas Fault,' an earthquake prone region of the globe.

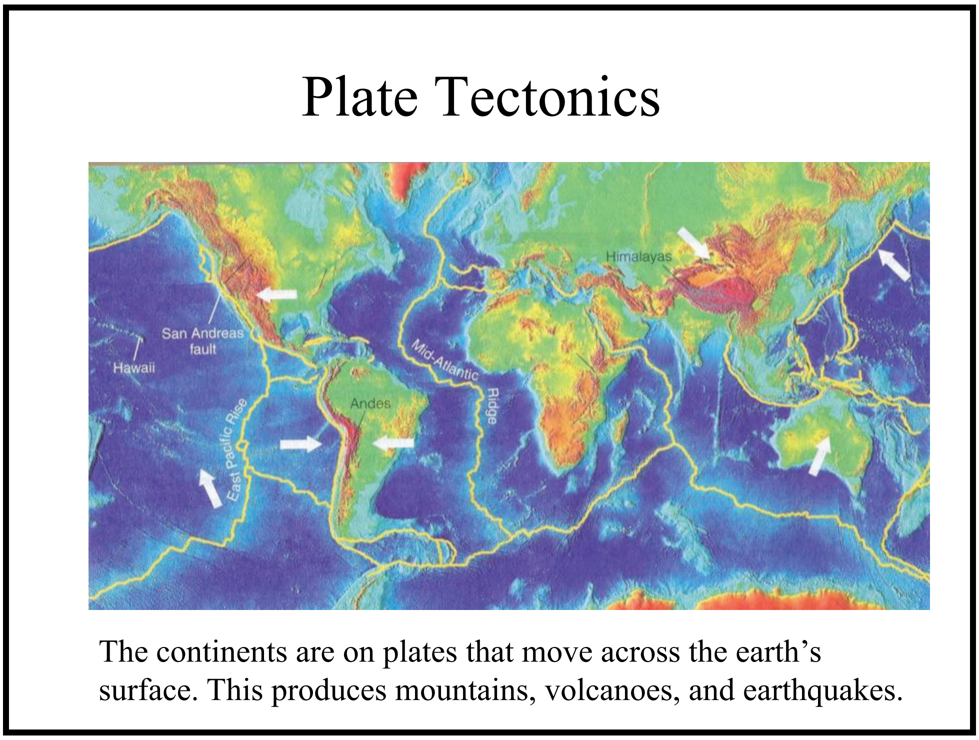


Figure 11.5 Plate Tectonics

12 Biological Evolution

Your goals for this chapter are to know the following:

- What is the significance of the Miller-Urey experiment?
- Identify the defining difference between fish and earlier forms of life.
- Identify the defining difference between amphibians and fish.
- Identify the defining difference between reptiles and amphibians.
- Identify the defining difference between mammals and reptiles.
- Identify the defining difference between primates and mammals.

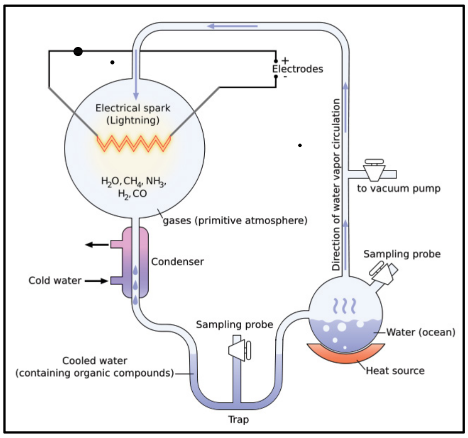
In trying to understand history, it is convenient to divide it into distinct eras, in spite of the fact that it has been one continuous process. We do the same with the evolution of the universe. Each category we've discussed to this point, cosmic, stellar, and geological, flows continuously from its predecessor. Perhaps, as with history, the flow is not inevitable, but it is at least logical. Also, as with history, its continuous nature sometimes makes it hard to know exactly when one type of evolution transitions into one of the other categories.

In the previous chapter, we treated geological evolution. But geological evolution has not been the only evolution transforming the earth throughout its history. Concurrently, chemicals were evolving, producing progressively more complex molecules and microstructures composed of these molecules. Although the details are not known at the present time, it seems clear that at some point, these structures became sufficiently complex to self-replicate; that is life arose from non-life, and chemical evolution became biological evolution.

12.1 Chemical Evolution

In the 1950s 'the origin of life' was a widely discussed scientific topic. It was hypothesized that conditions on the primitive earth favored chemical reactions that synthesized organic compounds from inorganic precursors. The Miller-Urey experiment, conducted in 1952, simulated conditions thought at the time to be present on the early earth. After allowing the experiment to run for a couple of days, several amino acids were identified in the apparatus. Prior to this result, it was generally accepted that only biological processes could synthesize complex organic molecules.

The Miller-Urey Experiment



The flasks contain liquid water representing the sea, and gases representing the early atmosphere. The apparatus is supplied with energy in the form of electrical discharges which simulate lightning. Chemical reactions produce a wide variety of organic molecules, including amino acids, the building blocks of proteins.

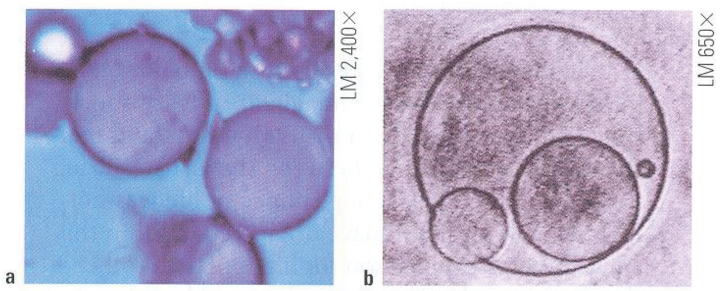
Figure 12.1 The Miller Urey Experiment

Since then, similar experiments have been done, including ones with more realistic conditions representing the early earth. These have produced an even larger variety of amino acids, as well as fatty acids (the building blocks of lipids) and nucleic acids (the building blocks of RNA and DNA).

Other experiments have shown that if these building blocks are dried and heated, they link up, forming long chains similar to the most complex biological molecules. These experiments, together with the discovery of amino acids in meteorites and of complex organic molecules in interstellar space, make it clear that chemical evolution is a fundamental part of the nature of the universe.

If certain complex organic molecules are put in water, they spontaneously form small, cell-like structures. Proteins and lipids placed in water, at appropriate values of acidity, salinity, and temperature, form membranes very much like cell membranes. In addition, these will spontaneously close in on themselves to produce tiny sacks that exhibit properties normally associated with life. They concentrate organic molecules within their membranes, and they can grow in size and divide in two when they get too large.

Formation of proto-cells



Laboratory versions of proto-cells. (a) These microscopic spheres were made by cooling a warn-water solution of amino acids. Although they are not alive, they exhibit many lifelike properties. (b) These microscopic membranes are made from lipids that when mixed with wafer spontaneously form enclosed droplets.

Figure 12.2 Formation of proto-cells

We do not know that these specific mechanisms brought about life on earth. The importance of these experiments is not that they tell us exactly how life arose, but that they demonstrate the tendency of chemical evolution to produce cell-like structures. We may never know exactly how nature accomplished the transition from chemical to biological evolution, but the experimental evidence that this transition represents a natural development is compelling.

12.2 Biological Evolution

In biology, evolution is the process by which populations of organisms acquire and pass on novel traits from generation to generation. Its occurrence over large stretches of time explains the origin of new species and ultimately the vast diversity of the biological world. Contemporary species are related to each other through common descent, products of evolution and speciation over billions of years.

The cell is the fundamental structure of life. It is the least complex bit of matter that can undergo all of the processes that distinguish the living from the nonliving. The first forms of life on earth appeared in the oceans. They were simple one-celled organisms know as bacteria. These original life forms arose very early in the history of the earth. Possible microfossils of one-celled organisms have been found in rocks of Western Australia and South Africa that formed some 3.5 billion years ago. The earliest unambiguous records of life are 2.7 billion year old fossils of blue-green algae colonies found in rocks from Canada and Zimbabwe.

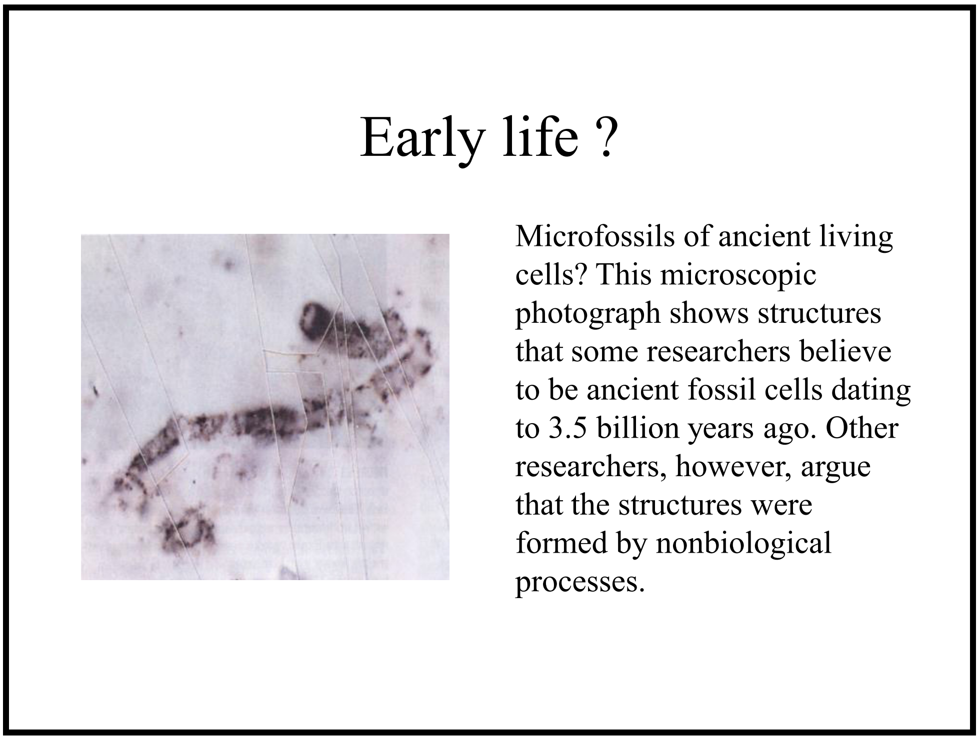


Figure 12.3 Early Life

Photosynthesis is the chemical reaction that allows plants to use the energy of sunlight to make the complex molecules that are the food supply of all living things. The raw materials needed are carbon dioxide and a source of hydrogen. About 3 billion years ago, organisms evolved that were able to utilize water as the hydrogen source. These organisms combined carbon dioxide (CO_2) with water (H_2O) to produce carbohydrates (molecules with a 1:2:1 ratio of carbon, hydrogen, and oxygen). Two oxygen atoms were left over for each water molecule used. The oxygen was simply released into the atmosphere.

Oxygen is a highly reactive element. At first the released oxygen combined with other elements, primarily iron, at the earth's surface. (Rust is oxidized iron.) Then, about 2 billion years ago, once most of the materials at the earth's surface had been oxidized, oxygen began to accumulate in the atmosphere. When this process began, the atmosphere was almost 100% nitrogen. Some 600 million years ago, by the start of the Cambrian period, oxygen made up about 5% of the atmosphere. Today it is accounts for 21%.

The increase of oxygen in the atmosphere had two important consequences. First, organisms developed that could use the oxygen to break down food molecules. This reaction, called respiration, is the reverse of photosynthesis, and allows much more energy to be extracted from food molecules than had been possible previously. The greater efficiency provided by respiration played a major role in the later diversification of life forms. Secondly, atmospheric oxygen also began to produce a layer of ozone (O_3) in the earth's upper atmosphere. Prior to the development of the ozone layer, the deadly ultraviolet radiation of the sun confined life to the seas in which it had originated. However, ozone is a strong absorber of ultraviolet radiation, particularly in the shorter wavelength region that is so dangerous to living organisms. A growing ozone layer made it progressively safer for life to rise to the surface of the water and eventually to occupy the land.

Evolution of life on earth

- **The basic timeline is a 4600 million year old Earth, with**
  - **4000 my (million years) of simple cells,**
  - **3000 my of photosynthesis,**
  - **2000 my of complex cells,**
  - **1000 my of multi-cellular life,**
  - **600 my of simple animals,**
  - **500 my of fish and proto-amphibians,**
  - **475 my of land plants,**
  - **400 my of insects, seeds and amphibians,**
  - **300 my of reptiles,**
  - **200 my of mammals,**
  - **150 my of birds,**
  - **100 my of flowers,**
  - **65 my since the non-avian dinosaurs died out**
  - **7 my of hominids, and**
  - **0.1 my of Homo sapiens.**

Figure 12.4 Evolution of life on earth

The Cambrian period, between about 500 and 600 million years ago, was characterized by a great proliferation of new species. There were probably many factors that contributed to this explosion of life, but certainly the increase in the oxygen content of the atmosphere was crucial. About 450 million years ago an important milestone in the evolution of humans occurred – the first backbones took shape when fish evolved from an invertebrate ancestor. Life made its first appearance on land about 440 million years ago and consisted of simple plants. Spiders, scorpions, and millipedes soon followed. About 370 million years ago, the first amphibians evolved from fish, and emerged from the water to establish the evolutionary branch that today included all terrestrial vertebrates, including us.

Reptiles evolved from amphibians about 300 million years ago. They evolved to fill a variety of terrestrial niches and soon became the most diversified vertebrate form on land. About 200 million years ago, a subclass of reptiles developed special secreting glands to nourish their young and became the first mammals. The mammals lived in the shadows of the dinosaurs until about 65 million years ago when the giant reptiles became extinct.

The extinction of the dinosaurs coincided with the extinction of about 75% of all species alive on earth at the time, something known as a mass extinction. It seems that the mass extinction of 65 million years ago was caused by the impact of a meteorite near the Yucatan peninsula. The meteorite had sufficient mass to cause extended, global weather changes. The extinctions may have taken several million years to complete, but this is practically overnight by geological timescales. There is strong fossil evidence for many additional mass extinctions, one of which eliminated about 90% of the then-existing species. Some of these may also have been caused by meteorite impact.

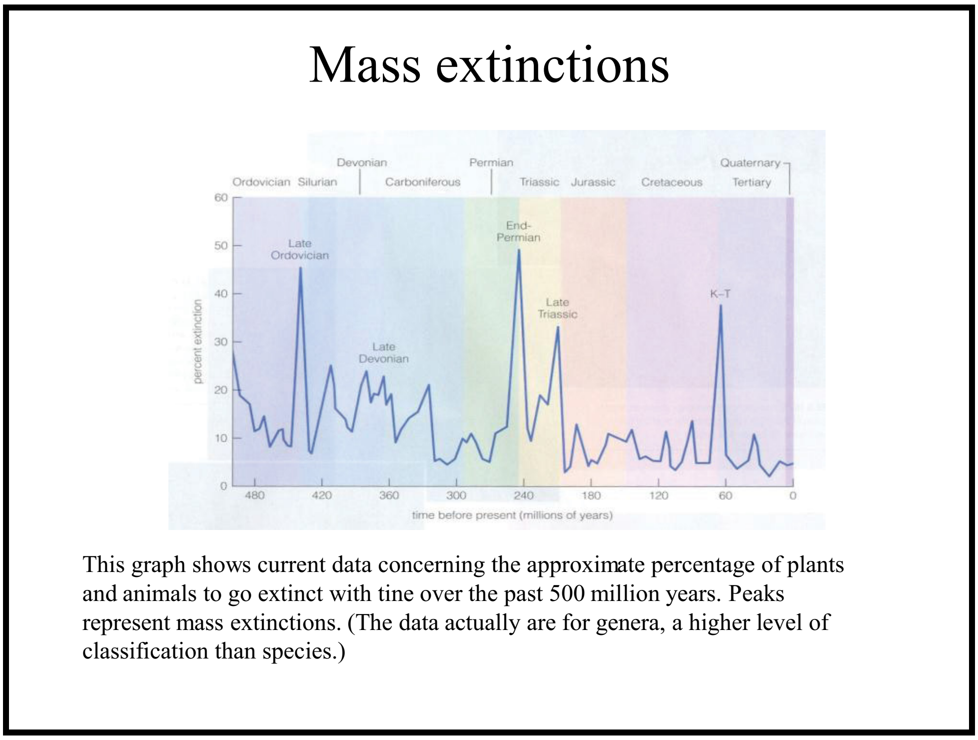


Figure 12.5 Mass extinctions

12.3 Primates

After the extinction of the dinosaurs, mammals flourished. At about this same time, early primates – small, forest-dwelling, squirrel-like creatures. – evolved. Primates are distinguished from other mammals by several features, including shoulder joints that allow a high degree of movement in all directions, opposable thumbs, and a trend toward a reduced snout and flattened face, attributed to a greater reliance on vision at the expense of olfaction.

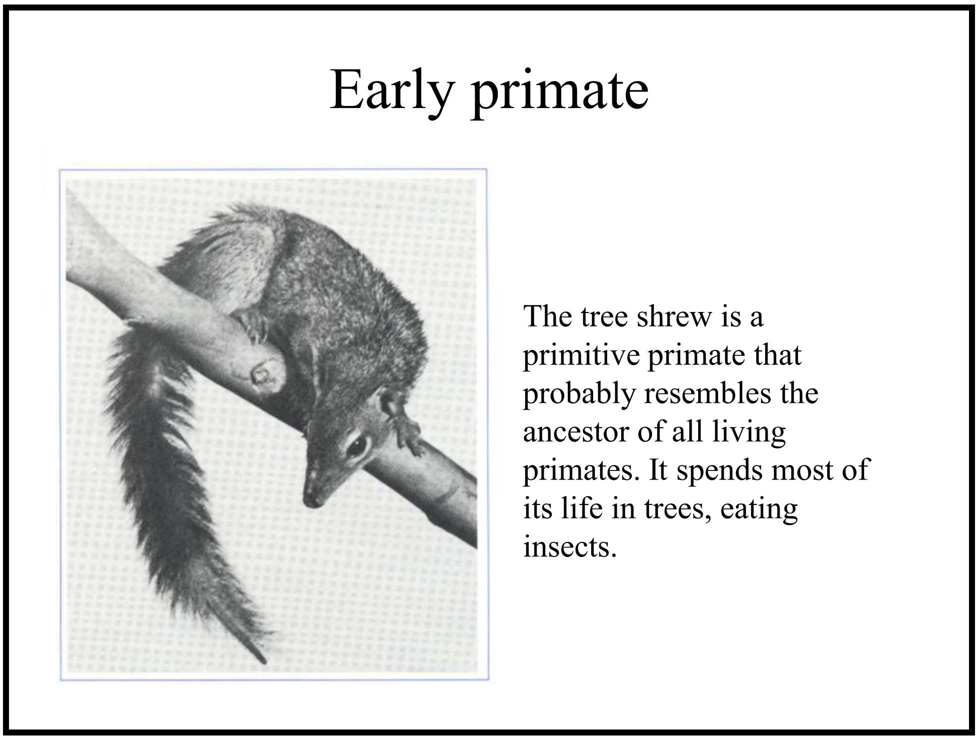


Figure 12.6 Early primate

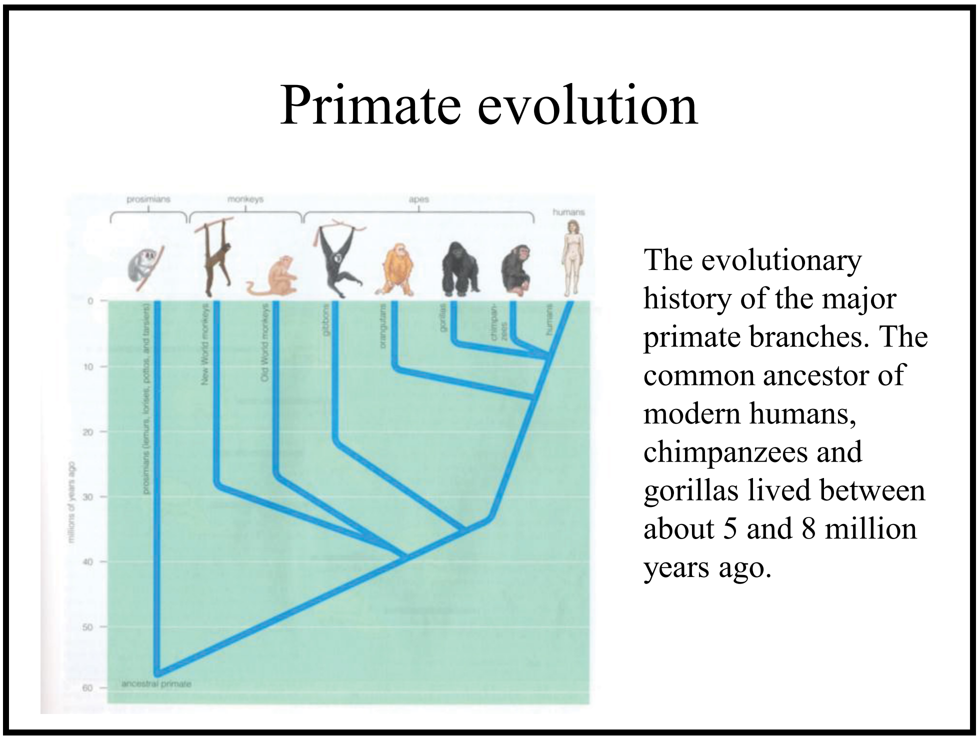


Figure 12.7 Primate evolution

About 40 million years ago, the evolutionary branch that would eventually lead to humans branched and the lines leading to the old world and new world monkeys split off. This branching continued as lines leading to today's gibbons, orangutans, and gorillas split off. Finally, some 5 or so million years ago, the line leading to chimpanzees diverged.

Chimpanzees are our closest living relatives. About 96% of our respective genomes are identical, although recent studies have shown that the vast differences between humans and chimpanzees are due more to changes in gene regulation than differences in individual genes themselves. Certain genes control the expression of other genes, so that even identical genes may be expressed differently in different species. Thus small changes in these regulatory genes can have a substantial impact.

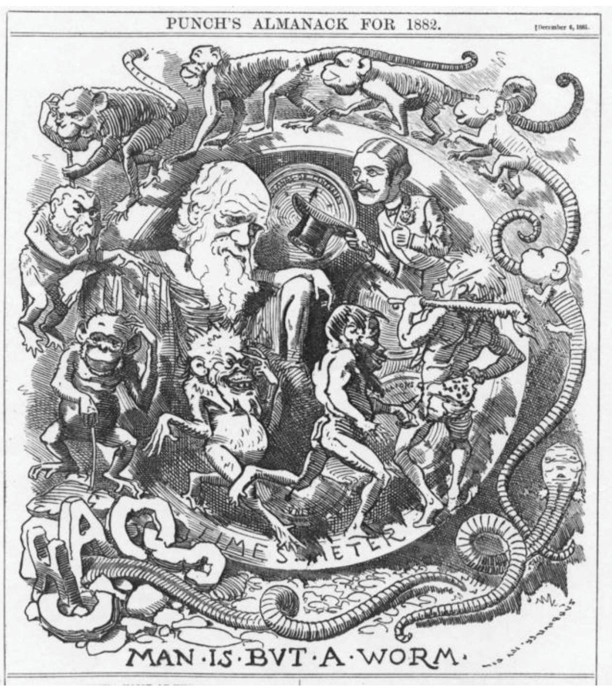
After the lines leading to chimps and humans split, species on the line leading to humans are known as hominids. Dozens of hominid fossil species have been identified beginning in 1856 with the discovery fossils now known as *Homo neanderthalensis*. The next chapter will discuss many of these species.

13 Human Evolution

Your goals for this chapter are to know the following:

- List at least five species on the hominid line of descent.
- Compare and contrast Australopithecus africanus and Homo erectus.
- Compare and contrast Homo erectus and Homo neanderthalensis.
- Compare and contrast Homo neanderthalensis and Homo sapiens.
- Explain the two theories as to how Homo sapiens came to be the single hominid species on earth.

In *The Origin of Species*, Darwin dealt only with plants and nonhuman animals. He did not directly address the question of human origins, fearing that it would prejudice readers against his general theory. The book's only mention of human evolution was the prediction that "light will be thrown on the origin of man and his history." The implication that humans descended from ape-like ancestors, however, was clear and immediately became the focus of public ridicule and outrage.



Man is But a Worm, from the Punch Amanack of 1882, published soon after Darwin's last book *The Formation of Vegetable Mould through the Action of Worms*.

Figure 13.1 Darwin cartoon

In 1860, at the annual meeting of the British Association for the Advancement of Science, Oxford's scholarly Anglican bishop, Samuel Wilberforce, delivered a long speech raising scientific objections to Darwin's theory. The meeting was attended by Thomas Huxley, an outspoken defender of Darwin. Near the end of his talk, Wilberforce jokingly asked Huxley if he was related to apes on his grandfather's or grandmother's side. Huxley is said to have whispered as he rose to respond, "The Lord hath delivered him into mine hands."

No transcript of the meeting exists, but Huxley claimed to have said, "If then the question is put to me would I rather have a miserable ape for a grandfather or a man highly endowed by nature and possessed of great means and influence and yet who employs those faculties for the mere purpose of introducing ridicule into a grave scientific discussion, I unhesitatingly affirm my preference for the ape." The exchange quickly became legendary.

Darwin, by nature, was not combative. In the years following the publication of *The Origin of Species*, he steered clear of the intense public debate on the topic of human evolution, relying on more combative defenders to press the case. Huxley, known as 'Darwin's bulldog,' was perhaps the best known and most effective of these. Finally, in 1871, he addressed the question in a massive two-volume book, *The Descent of Man*.

This book did not have the impact of *The Origin of Species*. The arguments were not as compelling, and by then everyone had already made up their mind on the subject. Most rejected the idea that humans were simply the product of natural selection, even many who readily accepted evolution as the explanation for the origins of 'the lower forms.' Among these was Alfred Wallace, the co-founder of evolution through natural selection. (Wallace believed that natural selection was insufficient to explain the development of consciousness and the human mind.) It would take the abundant fossil discoveries of the 20th century to provide the evidence that put human evolution on solid scientific ground.

13.1 Lucy

Perhaps the most famous fossil relating to human evolution is Lucy. The fossil was discovered in 1974 in Ethiopia by a team lead by Donald Johanson and is estimated to be about 3.2 million years old. She got her name from the Beatles song *Lucy in the Sky with Diamonds*, that played repeatedly at the party celebrating her discovery.

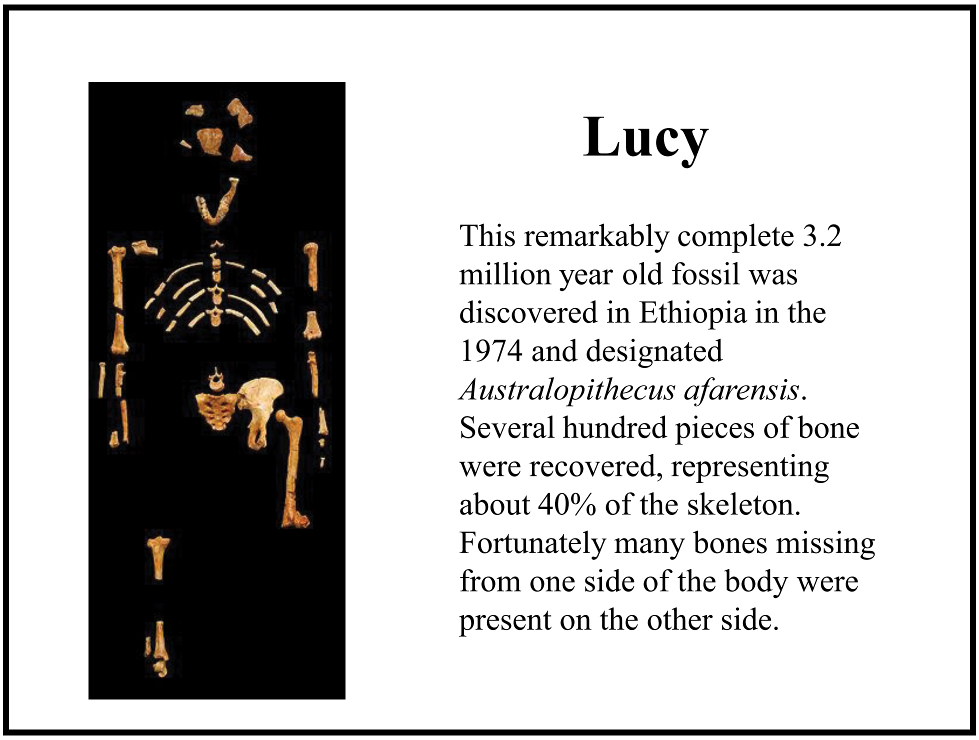


Figure 13.2 Lucy

Lucy was 3 feet, 8 inches tall and weighed about 65 pounds. She had a relatively small brain, a little bigger that a chimpanzee, but only about 30% that of modern humans. The structure of her hips showed that Lucy walked upright. This result was important as it showed that bipedalism evolved before brain size increases. Johanson suggested the species name of *Australopithecus afarensis* for Lucy, *afarensis* for Afar, the region of Ethiopia where the fossil was discovered. Since then there have been many additional discoveries of *A. afarensis* fossils making it one of the best documented pre-human hominid species.

Initially some suggested *A. afarensis* was at least partially arboreal and that, when not in the trees, walked with an ape-like gait, much as chimps sometimes do. Recent discoveries of fossil foot bones however, clearly show that Lucy had a modern human-like foot and walked fully erect, much as we do today.

Australopithecus afarensis

Compared to the modern and extinct great apes, *A. afarensis* has reduced canines and molars, although they are still relatively larger than in modern humans. *A. afarensis* also has a relatively small brain size (~380–430 cm^3) and a face with forward projecting jaws.

Figure 13.3 Australopithecus

13.2 The Earliest Hominids

Based on fossil and genetic evidence, it is estimated that the lines leading to chimpanzees and humans diverged about 5 million years ago. Fossils dating from that era are rare and much of the interpretation is highly speculative.

Sahelanthropus tchadensis is an ape-like fossil approximately 7 million years old. The fossils found include only fragments of the skull. These show a mixture of human and chimpanzee features. The braincase suggests a chimpanzee, but the teeth are closer to those of humans. The point at the back of the skull where the neck muscles attach suggests that this species walked upright.

The discovery was made in Chad in the central part of Africa resulting in both its scientific name and its nickname, 'Toumai' meaning 'hope of life' in the local language of Chad. *S. tchadensis* may be a common ancestor of both humans and chimpanzees, though it is unlikely to be the most recent one. The evidence from DNA mutation clocks suggest humans and chimps diverged about 5 million years ago, some 1 or 2 million years later.

Orrorin tugenesis is another fossil from approximately the time that humans and chimpanzees diverged. It is represented only by some teeth, parts of two lower jaws, and fragments of arm and leg bones. These are accurately dated between 6.1 and 5.8 million years.

The leg bones strongly suggest that *O. tugenensis* was bipedal. The teeth are small and remarkably human-like, in fact, more human-like that the later australopithecines such as Lucy. The discovers claim on the basis of dental anatomy and advanced bipedality, that *O. tugensis* was in the direct line of descent to modern humans and that all of the *australopithecine* species, with their slightly more ape-like teeth, eventually became extinct This view is extremely controversial.

13.3 Australopithecines

Chronologically, the next important group of species is the *australopithecines*. Unlike S. Tchadensis and O. tegenensis, australopithecines are unambiguously on the hominid side of decent from the common ancestor of chimpanzees and humans, although whether they are direct ancestors or a branch of the line that became extinct, is still being debated. The oldest of these species is *Australopithecus amenensis* dated at about 4 million years ago. They were originally classified as *A. afarensis*, but were reclassified in 1995 based primarily on differences in teeth. The next species in the australopithecus genus is *A. afarensis* (Lucy) discussed earlier.

The first australopithecine fossil was discovered and named by Raymond Dart, a South African anthropologist, in 1924. He named it *Australopithecus africanus* (southern ape of Africa). Although primitive in many of its features, this juvenile specimen had teeth that were distinctly more human-like than ape-like.

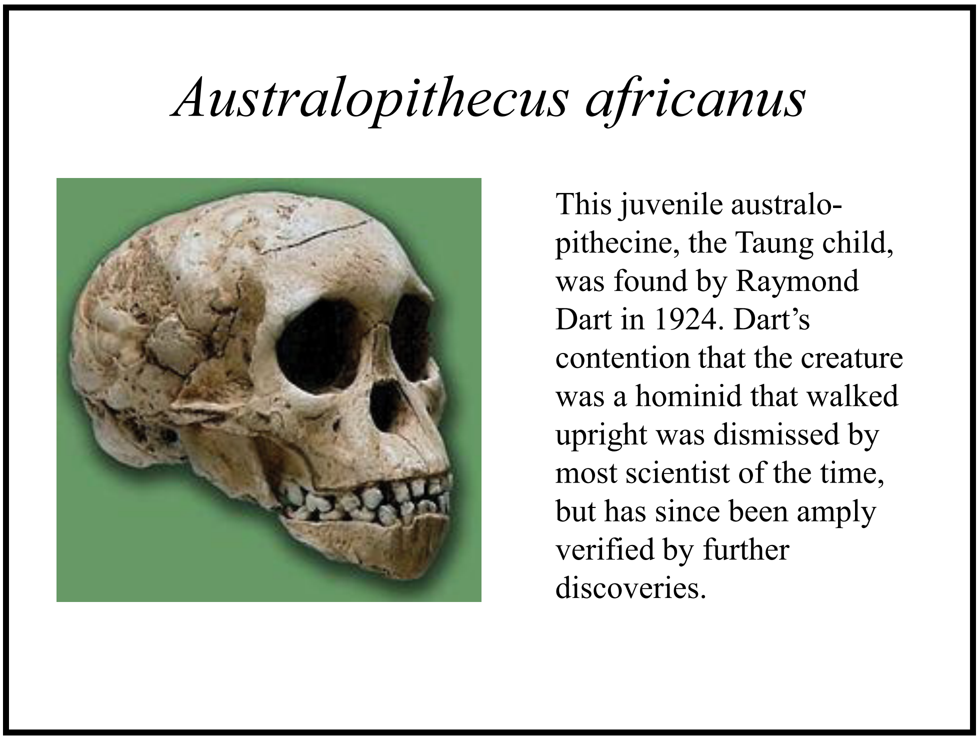


Figure 13.4 Australopithecus africanus

However, due to other more primitive features, some researchers believe A. africanus, instead of being a direct ancestor of humans, evolved into the genus Paranthropus, a more robust group of species definitely not in the direct line of human descent. One species seen as a descendent of *A. africanus* is Paranthropus robustus. Both *P. robustus* and *A. africanus* crania seem very similar despite the more heavily built features of *P. robustus* that are adaptations for heavy chewing, suggesting a diet similar to that of gorillas.

A later australopithecine species is *A. garhi*, was found in Ethiopia in 1996 and dated at about 2.5 million years ago. The species name '*garhi*' means surprise in the local Afar language. *A, garhi* is suspected by some to be the ancestors of the *Homo* line leading to us. In fact, primitive stone tools have been found with the *A. garhi* fossils, the criterion for being considered in the genus *Homo*.

13.4 Homo habilis

The genus *Homo* is reserved for species capable of manufacturing and using tools. In modern taxonomy, *Homo sapiens* is the only extant species of this genus. The fossil record shows that there were earlier *Homo* species, all of which are now extinct. While some of these other species might have been ancestors of *H. sapiens*, many were likely our "cousins", having speciated away from our ancestral line. There is not yet consensus as to which of the *Homo* species are our direct ancestors, but one frequently mentioned in this context is *Homo habilis*.

The first species found in close association with primitive stone tools was given the name *Homo habilis* (literally, handy man). Mary and Louis Leakey found the first fossils in Tanzania, East Africa, in 1962. Many additional *H. habilis* fossils have been found, all exclusively in Africa. They date to approximately 2.3 to 1.4 million years ago. In 2010, *Homo gautengensis* was discovered and is believed to be even older than *H. habilis*.

The tools found with *H. habilis* are crude stone tools – core stones flaked to produce sharp edges. These are similar to, but slightly more sophisticated than those found in association with *A. garhi* fossils. These types of tools are referred to as 'Oldowan' after the gorge in Tanzania where they were first found. Oldowan tools were used by hominids from about 2.5 to 1.7 million years ago.

H. habilis was short and had disproportionately long arms compared to modern humans. However, it had a less protruding face than the australopithecines from which it may have descended. *H. habilis* had a cranial capacity slightly less than half the size of modern humans.

In the mid-20th century, when the fossil record was much more incomplete than now, many anthropologists believed that humans descend from our earliest hominid ancestor in a single, un-branching line. We now know that that is not true. There have been many branches, all but one leading to eventual extinction. For example, *Homo habilis* co-existed with other *Homo*-like bipedal primates, such as *Paranthropus robustus and P. boisei*. However, *H. habilis*, possibly because of its early tool innovation and a less specialized diet, very likely became the precursor of an entire line of new species, whereas the *Paranthropus* line disappeared from the fossil record.

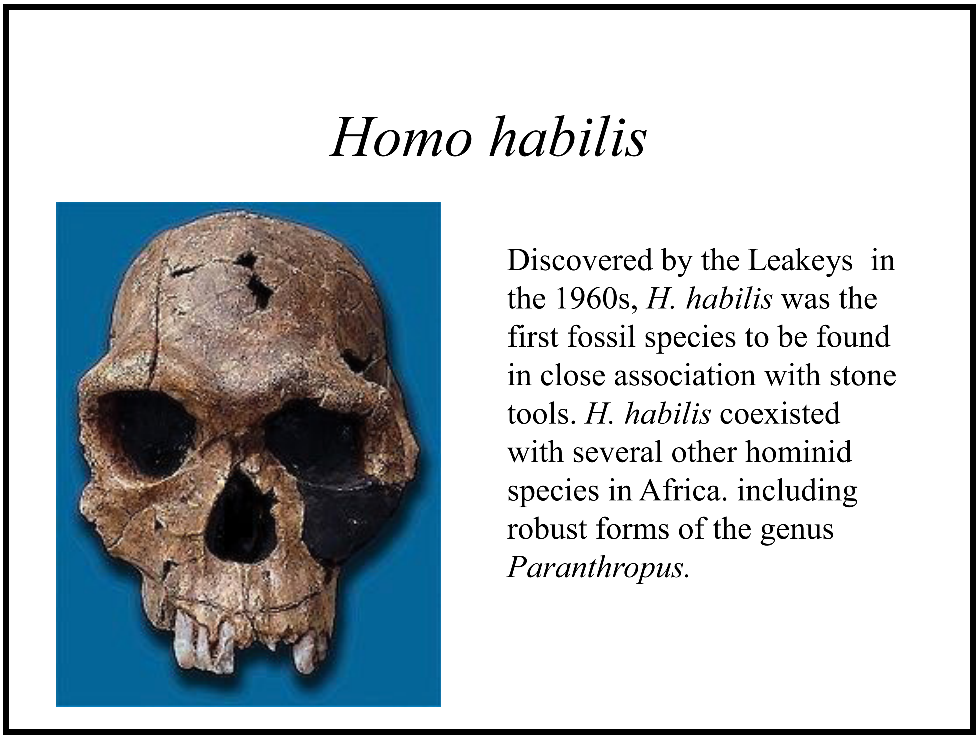


Figure 13.5 Homo habilis

13.5 Homo erectus

The fossils that eventually became known as *Homo erectus*, were among the earliest hominid fossils discovered. Eugene Dubois, a Dutch physician/anatomist, became obsessed with Darwin's evolutionary theory, particularly as it applied to humans. In 1886, he gave up a successful medical practice to travel to Asia in search of the 'missing link.' In 1891 on the Indonesian island of Java, he found a fossilized skullcap and femur that show several features of modern humans but are clearly more primitive. He was sure he had found the missing link and named his find *Pithecantropus erectus* (literally, erect walking ape man).

Dubois made his find public a few years later, and was met by derision from the dominant British paleontological hierarchy. Dubois was disillusioned and eventually refused to allow other researchers to examine the fossils, hiding them away in his house.

In the 1920s and 30s, similar fossils were discovered near Beijing, that became known as Peking man. Similarly, Dubois' find was by then known as Java man. The two were recognized as the same species and renamed *Homo erectus*. Other than our species, *H. erectus* was the most far-ranging hominid species with remains being found throughout Africa, Europe, and Asia. *H. erectus* first appeared in Africa some 1.9 million years ago and became extinct in Africa and Europe some 300 thousand years ago. However they may have survived until as recently as 50,000 years ago in Asia. It is believed by many that African *H. erectus* gave rise to modern humans while the lines in Europe and Asia were dead-ends.

H. erectus used more diverse and sophisticated stone tools than the Oldowan technology, its predecessor. The newer tool industry is known as Acheulean. *H. erectus* was probably the first hominid to live in small, familial societies similar to modern hunter-gatherers. *H. erectus* is thought to be the first hominid to hunt in coordinated groups, use complex tools, and care for infirm or injured companions. Sites in Europe and Asia seem to indicate controlled use of fire by *H. erectus*, some dating back 1.5 million years.

Early African *Homo erectus* fossils are sometimes referred to as *H. ergaster*, although the distinction is not accepted by everyone. In addition to *H. ergaster*, several additional species names appear in the literature for fossils others classify as *H. erectus*.

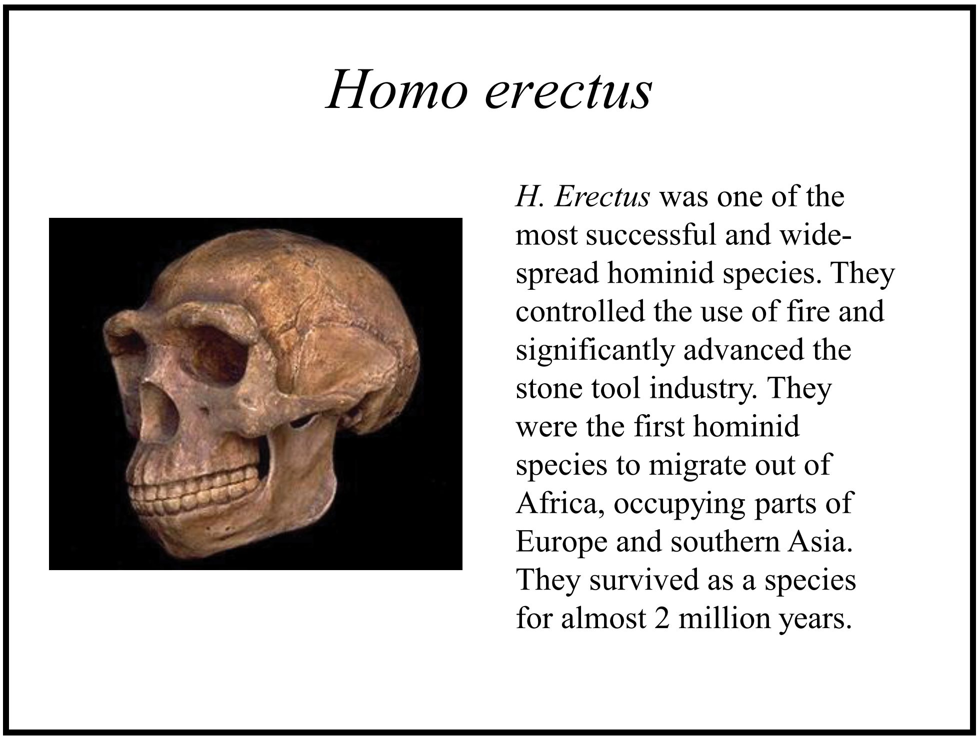


Figure 13.6 Homo erectus

13.6 Homo neanderthalensis

Between 250,000 and 30,000 years ago, another species of *Homo* inhabited territory from England across southern Europe into Asia. Since the first fossils were found in the Neander valley of Germany, these people are often referred to as Neanderthals. The Neanderthals had an average brain volume as large as, or perhaps even slightly larger than, modern humans. They built hut-like structures and produced a more advanced stone tool industry with which they hunted big game. They also developed ceremonial rituals. The remains of this relatively short-lived group reflect a state of consciousness akin to our own. They clearly had a sense of physical mortality. Ceremonial burials show a concern with life after death. In the Neanderthals there are signs of a greater consciousness of the subtleties of the world and the roots of our own complicated beliefs, societies, and religious sensibilities.

Neanderthals earliest ancestors evolved, like all hominids did, in Africa and migrated outwards into Europe and Asia. Recent mitochondrial DNA studies suggest that Neanderthals and modern humans had a common ancestor about 550,000 years ago. In Europe and Asia, Neanderthals lived following a combined scavenger and hunter-gatherer lifestyle until about 30,000 years ago, when they disappeared. For the last 10,000 years of their existence, Neanderthals shared Europe with anatomically modern humans and apparently the two types of humans led fairly similar lifestyles.

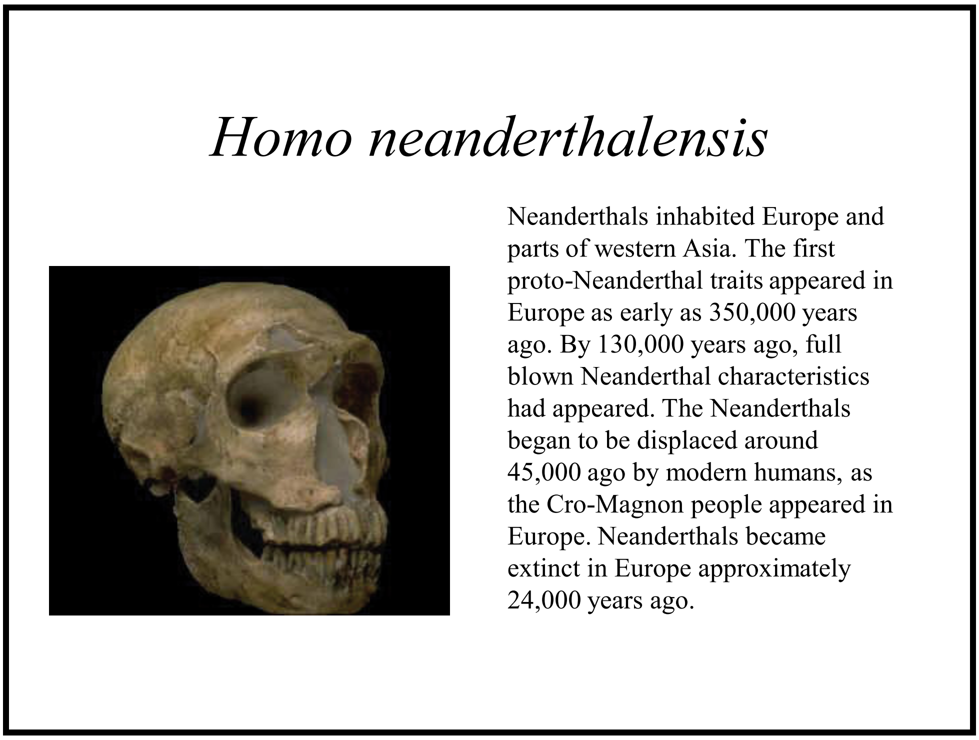


Figure 13.7 Homo neanderthalensis

One of the most hotly debated topics in human evolution is the relationship between modern humans and Neanderthals. Some researchers believe that they are closely enough related that they should be classified as subspecies rather than separate species. Thus, *Homo sapiens neanderthalensis* and *Homo sapiens sapiens*. This is currently a minority opinion. A related question is, Did they interbred? The answer seems to be yes, at least in a limited way. The Neanderthal Genome Project appears to have settled the issue, by uncovering evidence that some modern humans from Eurasia, but not from Africa, hold a tiny percentage (1–4%) of Neanderthal genes. Now that the issue of interbreeding seems to have been settled, why modern humans survived while Neanderthals did not is the issue being hotly debated. Reasons discussed include climate change, better language and thus cooperative hunting skills, and greater intellectual capabilities. Some have even suggested out and out genocide.

13.7 Homo sapiens

> "The creature called man has a strange history. He is not of one piece, nor was he born of a single moment in time. His elementary substance is stardust almost as old as the universe. His living organs, his eyes, backbone, his hands and feet – even his remarkable brain – have originated in far places and in different eras of time." The Lost Notebooks of Loren Eiseley, pp 105–106.

Anatomically modern humans arose in Africa between 200,000 and 150,000 years ago. Somewhere between 100,000 and 50,000 years ago, a small group, or several small groups migrated out of Africa and eventually populated the entire world. There are two distinct theories as to how this happened. One, called the Multi-regional hypothesis, asserts that as *Homo sapiens* spread, they interbreed with existing *H. neanderthalensis* and *H. erectus* populations. Thus today's humans are a blend of species. It seems well established that there was limited interbreeding between *H. sapiens* and Neanderthals, but, in general, the current evidence tends to favor the alternative hypothesis.

The Out of Africa hypothesis (more scientifically referred to as the 'recent single origin' hypothesis) asserts that all humans today are descended from the relatively small population that left Africa less than 100,000 years ago, and replaced the existing hominid species present at that time. The evidence for this is primarily genetic in nature. Regardless of how it occurred, the fact is that today there is but a single species of hominids alive on the planet. Not only a single species, but a single subspecies. Racial differences are obvious, but superficial, and not nearly sufficient to divide humans into subspecies.

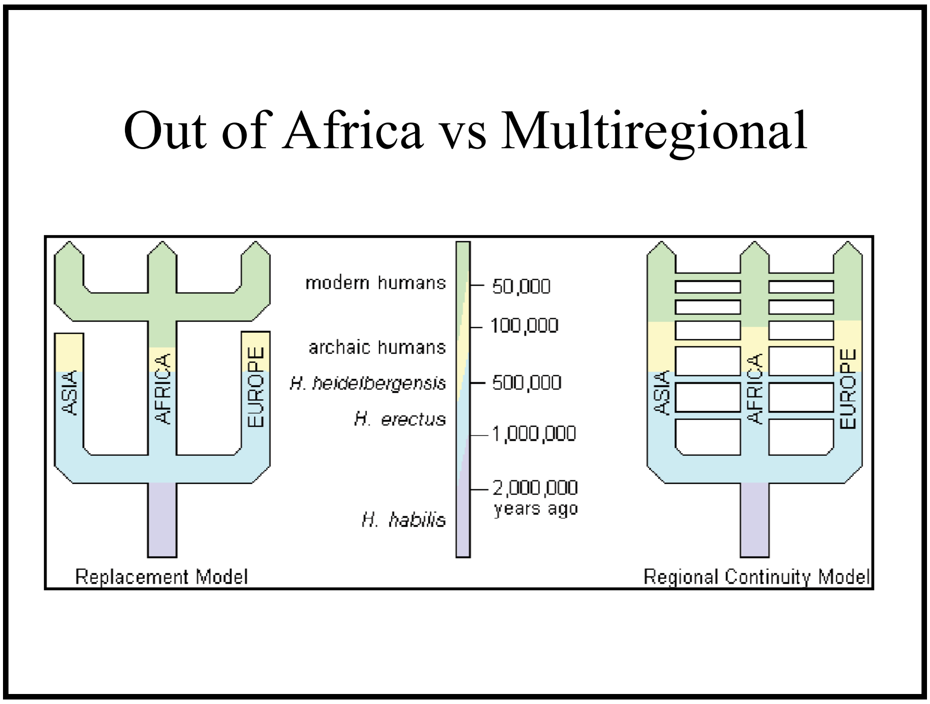


Figure 13.8 Out of Africa vs Multiregional

The Latin meaning of *Homo sapien* is 'wise man.' It reflects the greater endowment of brain power compared to predecessors. The species is defined primarily in terms of anatomy. Compared to other hominids, the species is characterized by a higher and more vertical forehead, a round and gracile cranium, small face and teeth, a prominent chin, and a more slender and elongated post-cranial skeleton.

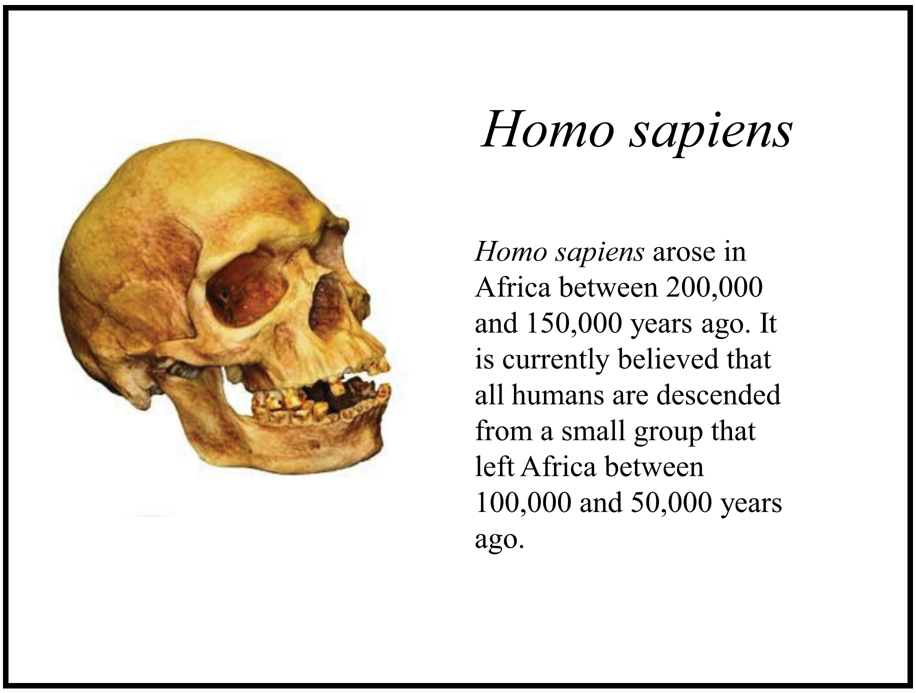


Figure 13.9 Homo sapiens

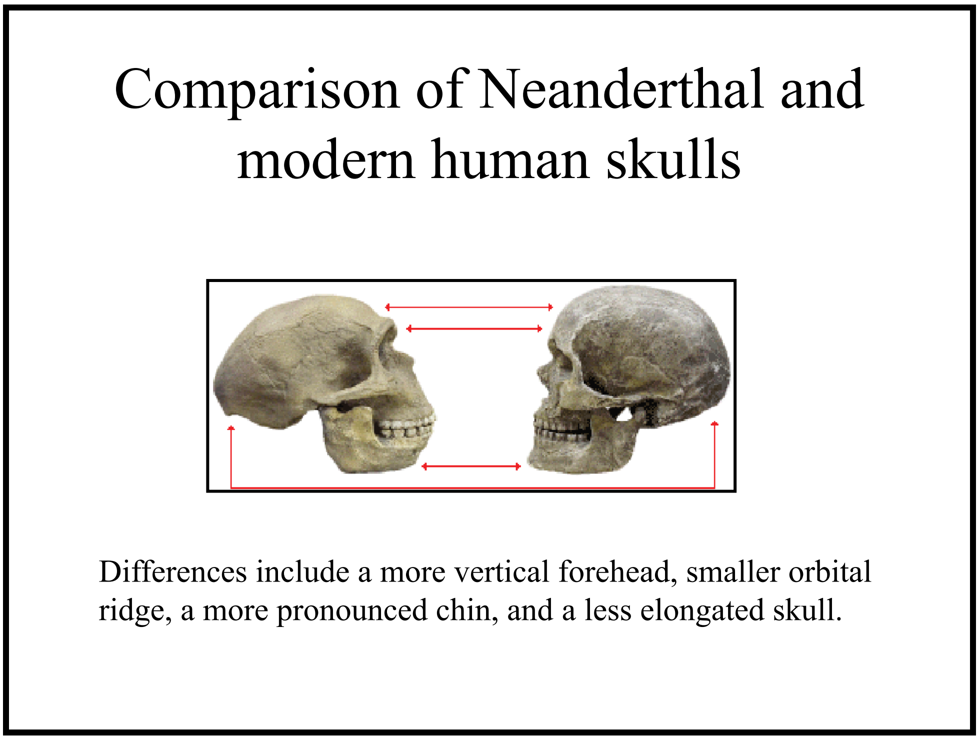


Figure 13.10 Comparison

If the record of the rocks had never been, if the stones had remained closed, if the dead bones had never spoken, still man would have wondered. He would have wondered every time a black ape chattered from the trees as they do in the Celebes where, of old, simple forest people had called them ancestors of the tribe. He would have wondered when he saw the huge orangs pass in the forest, their bodies festooned with reddish hair like moss and on their faces the sad expression of a lost humanity. He would have seen, even in Europe, the mischievous fingers and half-human ways of performing monkeys. He would have felt, aloof in religious pride and the surety of revelation though he was, a vague feeling of unease. It is a troubling thing to be a man, with a very special and assured position in the cosmos, and still to feel those amused little eyes in the bush – eyes so maddeningly like our own.

Loren Eiseley, *Darwin and the Mysterious Mr. X, page 187.*

14 Consciousness

Your goals for this chapter are to know the following:

- Identify the principle regions of the human brain
- Compare and contrast the brain and the mind.
- Define the following terms: neuron, synapse, functional neuroimaging.
- What evidence exists suggesting the evolution of consciousness?

Discuss evidence for the evolution of human consciousness.

The discussion of human evolution to this point has focused on the morphological changes that resulted in the transition from an ape-like ancestor to modern humans: bipedalism, reduction of tooth size, changes in the shape of the face, and increases in brain size. But there has also been a concurrent change that has taken place, one that is recorded only indirectly in the fossil record. As best we can tell, we are the only species driven to make sense of the universe, the only species able to stand in wonder and awe before the mystery and beauty of the Universe. How did this come about? Clearly it had something to do with evolutionary changes in brain size and structure. But what exactly is it that has produced these abilities and propensities? What is the relationship between mind and body? This is an old question, one that has only recently begun to be addressed by empirical methods.

14.1 The Human Brain

The brain controls the other organ systems of the body by causing secretion of chemicals such as hormones and neurotransmitters. This centralized control allows rapid and coordinated responses to changes in the environment. In vertebrates, the spinal cord by itself contains neural circuitry capable of generating reflex responses as well as simple motor patterns such as swimming or walking. However, sophisticated control of behavior based on complex sensory input, requires the information-integrating capabilities of a centralized brain.

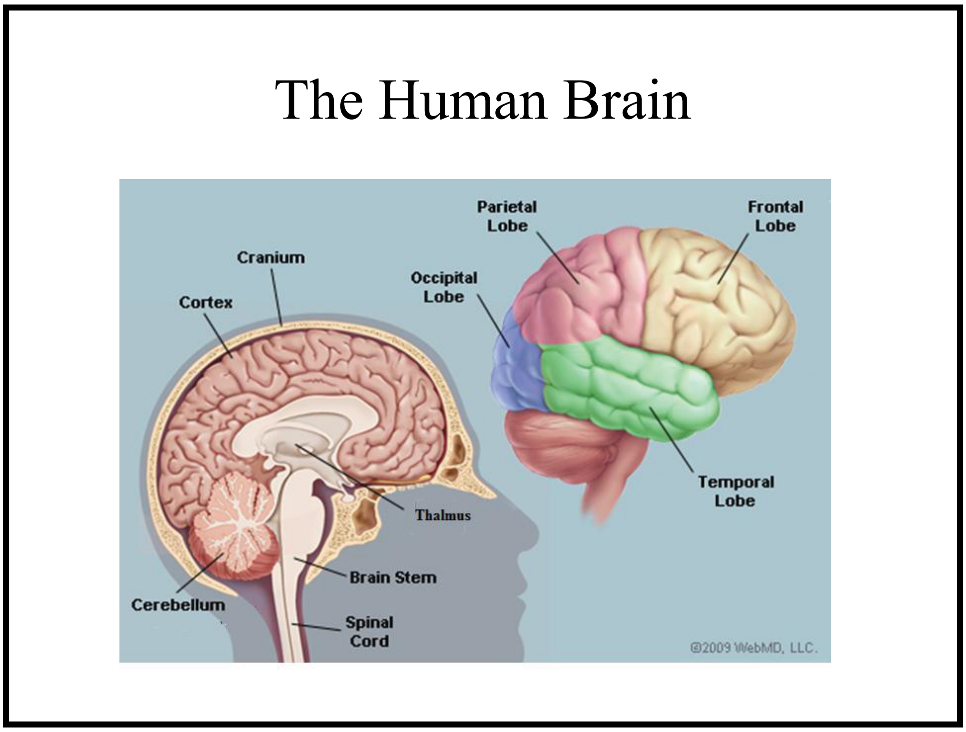


Figure 14.1 The Human Brain

Despite rapid scientific progress, much about how brains work remains a mystery. The operations of individual neurons and synapses are now understood in considerable detail, but the way they cooperate in ensembles of thousands or millions has been very difficult to decipher. Methods of observation such as EEG recording and functional neuroimaging techniques tell us that brain operations are highly organized.

The **cerebrum** is the largest part of the human brain, accounting for 85% of the organ's weight. The distinctive, deeply wrinkled outer surface is the **cerebral cortex**, which consists of gray matter. Beneath this lies the white matter which connects cortical neurons to other neurons in the brain and to other parts of the body. It's the cerebrum that makes the human brain, and the mental capacities associated with it, so distinct from that of other species.

The brain is made up of many specialized areas that work together. The cortex is the outermost layer of the cerebrum. Thinking and voluntary movements begin in the cortex. The brain stem is between the spinal cord and the rest of the brain. Basic functions like breathing and sleep are controlled here. The basal ganglia are a cluster of structures in the center of the brain. The basal ganglia coordinate messages between multiple other brain areas which often coordinate control of movement. The cerebellum is at the base and the back of the brain. The cerebellum is responsible for coordination and balance.

The cerebrum has two halves, or hemispheres. Research suggests that the two different sides of the brain control two different 'modes' of thinking. 'Left brain' thinking is frequently characterized by logic, reason, and objective analysis, while 'right brain' thinking is more random, intuitive, holistic and subjective. The research also suggests that most of us prefer one mode over the other. Broad generalizations should be avoided, however, as no one uses one side exclusively, and many people appear to utilize both sides equally.

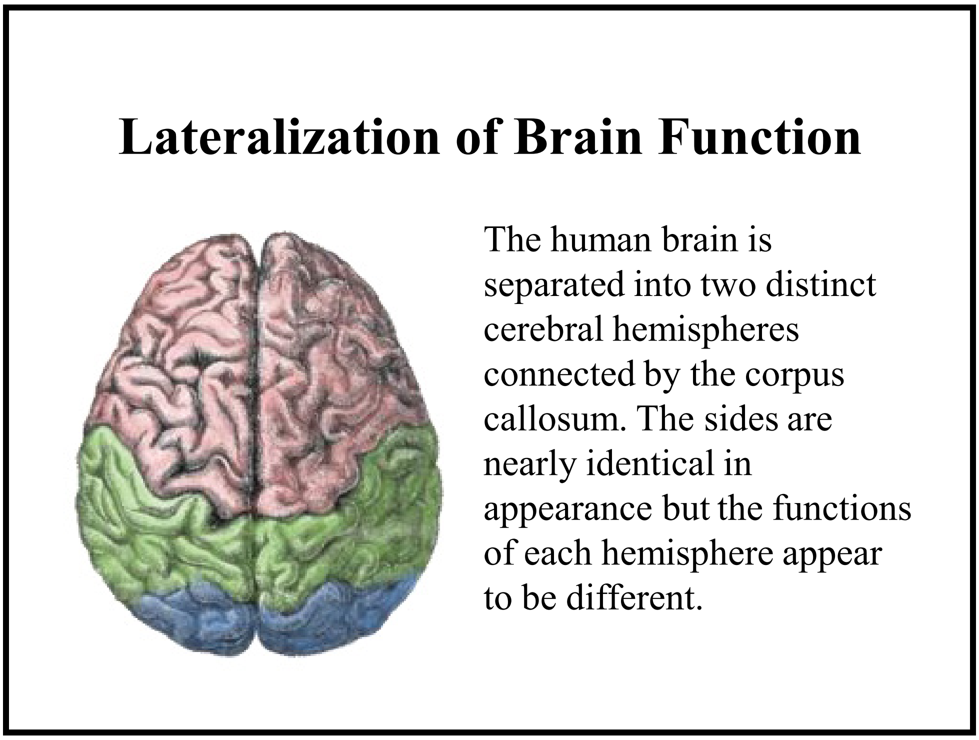


Figure 14.2 Localization of Brain Function

The two hemispheres are further divided into four regions, or lobes, in each hemisphere. The **frontal lobes**, located behind the forehead, are involved with speech, thought, learning, emotion, and movement. Behind them are the **parietal lobes**, which process sensory information such as touch, temperature, and pain. At the rear of the brain are the **occipital lobes**, dealing with vision. Lastly, there are the **temporal lobes**, near the temples, which are involved with hearing and memory.

14.2 The Mind-Body Problem

The tern 'mind' refers to the mixture of thought, perception, memory, emotion, will, and imagination, including all unconscious cognitive processes that as human beings we experience. The human mind creates an inner space that represents the outer world. We can picture our bodies, our cities, the earth, or our galaxy. We can imagine things that don't exist. In fact, we seem capable of incorporating virtually anything into this mind-space. For example, we can even visualize the non-spatial dimension of time as a spatial entity spread out before us from the Big Bang to the present. The mind's contents constitute our sense of self, a self that no one else has direct access to.

Descartes said, "I think, therefore, I am."

The idea that our mind and our body are distinct from one another arises because mental experiences seem qualitatively different from more concrete bodily processes. The body is a physical entity composed of matter and can be understood mechanistically, while consciousness seems more metaphysical. This has led in some cases to a dualistic approach; a belief that the mind and body are in some categorical way separate from one another.

Although a dualistic solution can be traced back to Plato and Aristotle, it was most precisely formulated by Rene' Descartes in the 17th century. Descartes argued that the mind is an independent existing substance that cannot be reduced to the brain. That is, mind could exist without brain and brain could exist without mind. Part of Descartes' motivation for this point of view was religious in that it provides a rational basis for believing in the soul's immortality. For Descartes, mind and soul are the same thing, and thus the soul could exist after the body has ceased to exit. Also, the complete separation of mind and body allowed Descartes to explain the material world, including the human body, in a mechanistic way without reducing humans to machines.

Part of Descartes' argument for the distinction between mind and body was that body is extended, bounded, and divisible, while mind is non-extended, unbounded, and indivisible. The mind cannot be measured in terms of size and weight as the body can.

Most modern philosophers have rejected the view of Descartes that mind and matter are different substances. Other formulations of dualism have been proposed. For example, property dualism holds that, although the world is constituted of just one kind of substance – the physical kind – there exist two distinct kinds of properties: physical properties and mental properties. In other words, it is the view that non-physical, mental properties (such as beliefs, desires and emotions) are intrinsic to some physical substances (namely brains).

In contrast to dualism, monism asserts that the mind is not something separate from the body. This view was advocated by Parmenides in the 5th century BCE and was later espoused by the 17th century philosopher Baruch Sponoza. Today, the most common form of monism is physicalism which asserts that the only existing substance is physical and predicts that as physical theory advances, the mind will eventually be explained in terms of the neural processes of the brain. Thus, the relationship between mental activity and brain activity is one that can be addressed through science.

14.3 How Does the Brain Think?

At one time consciousness was viewed with skepticism by many scientists and considered to be solely in the domain of philosophers and theologians. However, since the mid-20th century, it has been an increasingly significant topic of scientific research. Cognitive neuroscience is the academic field concerned with the study of the biological basis underlying cognition. The goal of cognitive neuroscience is to explain mental processes and behavior in terms of underlying brain mechanisms.

Brain activity arises from neurons – electrically excitable cells that process and transmit information to other cells by electrical and chemical means. The signals are transmitted across specialized connections between the cells called synapses. Through synapses, neurons connect to each other to form intricate networks of electrical and chemical activity. The human brain is a system of networks consisting of 10^{11} neurons with 10^{15} connections, making it the most complex system in the universe.

There are several empirical approaches used in cognitive neuroscience. Perhaps the earliest studies involved brain lesions. By correlating specific losses of function with damage to specific areas of the brain, the lesion method has provided a great deal of information about functional localization in the brain.

One of the earliest and most famous studies of localization of function was done by the French physician Paul Broca. In 1861, reporting on the autopsy of a patient suffering from the speech disorder known as aphasia, he identified the damaged area of the patient's brain. Broca showed that damage to the third convolution of the left frontal lobe was associated with the loss of the ability to speak. This segment of the brain came to be called Broca's area. Then, in 1873, Carl Wernicke, a German, identified an area controlling the understanding of speech, now called Wernicke's area.

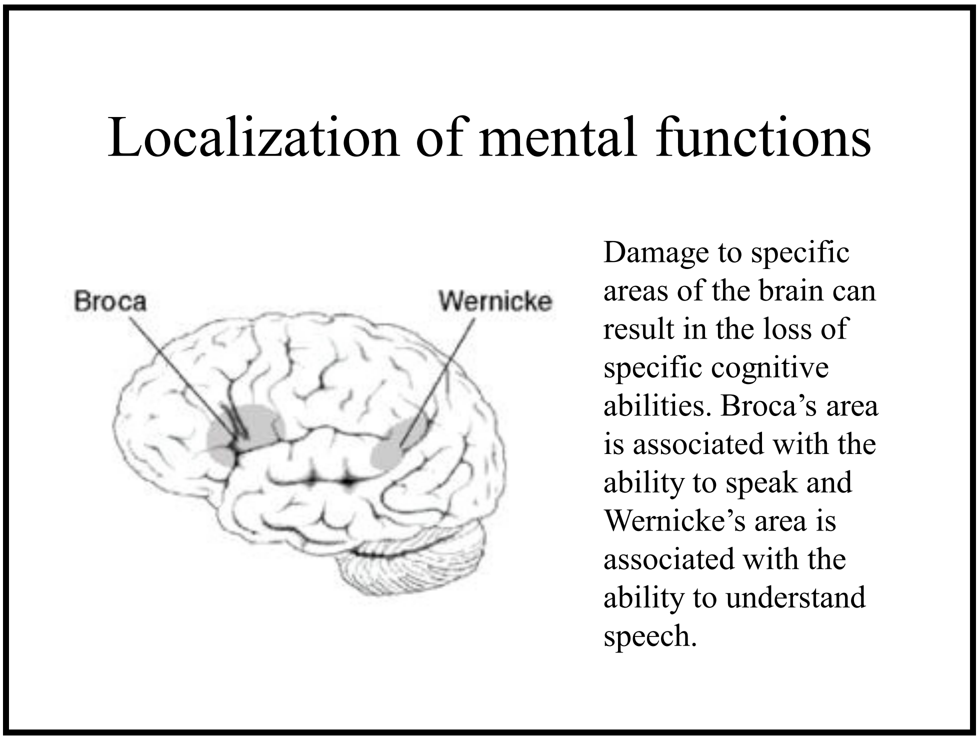


Figure 14.3 Localization of mental functions

Another technique used is electrical stimulation of the brain. Since the brain works by means of electrical impulses, brain areas may be tricked into revealing their function by applying brief pulses of electricity to the brain and seeing what happens. When brain areas are stimulated electrically, this mimics what happens when they are activated naturally. In this way, it is possible to induce thoughts and perceptions which to the experimental subject are indistinguishable from real experience. By the proper stimulation of the brain it is possible to induce the recall of forgotten events. Electrical stimulation of the brain is capable of inducing emotions such as fear, rage, or love in the complete absence of any reality that might produce these feelings.

Perhaps one of the most obvious indications of the physical nature of mind is psychoactive drugs. Psychoactive drugs are chemical substances that can cross the blood-brain barrier and affect brain function, resulting in changes in perception, mood, consciousness, cognition, and behavior.

An empirical technique more recently utilized in mind-brain studies is functional neuroimaging. This refers to the use of non-invasive brain imaging methods, such as functional magnetic resonance imaging (fMRI) or positron emission tomography (PET), to localize neural activity in the human brain related to specific mental functions. Such methods have been in use since the late 1980s to investigate the brain regions activated by sensory, motor, and cognitive processing.

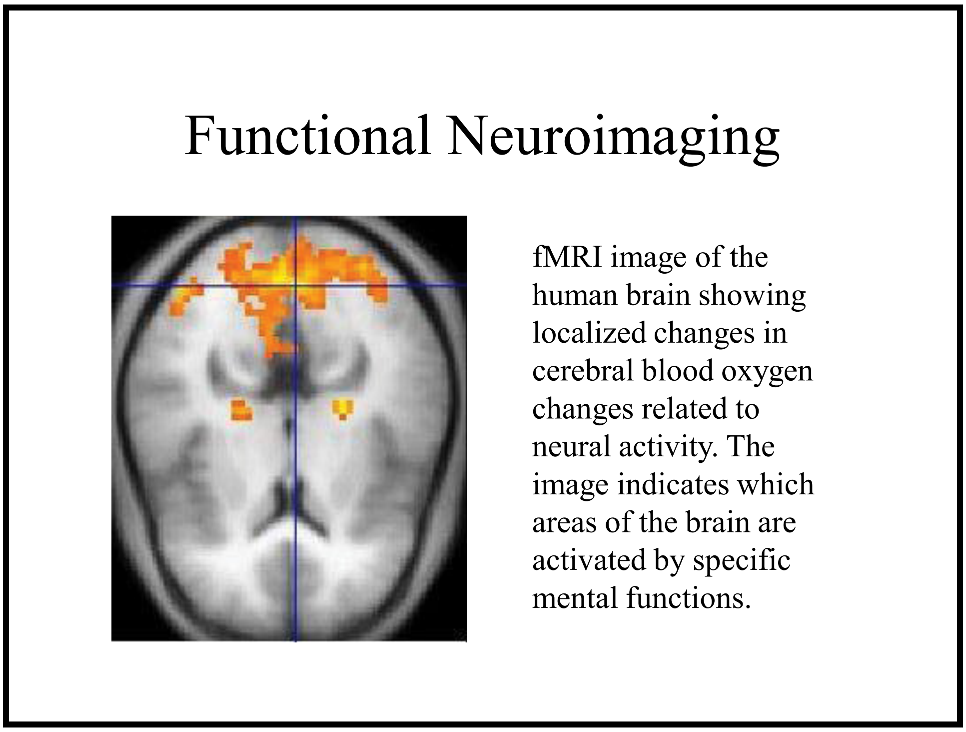


Figure 14.4 Functional Neuroimaging

Although significant progress has been made in the understanding of the physical nature of mind, it is far from possible at present to draw a strict, reducible identity between brain states and the mind.

14.4 The Evolution of Consciousness

The intimate connection between mind and body is well established, even if all the details are not yet known. It is also clear that some of our mental capabilities are possessed by other species. The intelligence of dolphins, chimps, and even crows is well established. Who could doubt that their dog possesses some level of consciousness akin to our own? If consciousness is not unique to humans, it is clear that changes in the levels of consciousness have evolved concurrently with speciation.

The evolution of consciousness undoubtedly results from the evolution of the brain. In the long evolution of our brain, several important changed in the nervous system have occurred. First, it became increasingly centralized in architecture, evolving from a loose network of nerve cells (as in the jellyfish) to a spinal column and complex brain with impressive swellings at the hindbrain and forebrain (vertebrates). This increasingly centralized structure also became increasingly hierarchical. The newer additions to the brain took over control from the previous additions and in effect became their new masters.

The evolution of the mammalian brain occurred in several stages, initially involving significant increases in the ability to smell. The evidence comes from the tiny skulls of two species of reptile that lived in China about 200 million years ago. Relative to other reptiles these fossils showed a huge increase in the size of the olfactory region of the brain. These reptiles are believed to have evolved into the first mammals. Dinosaurs evolved about the same time and soon came to dominate the landscape. They probably hunted during the day, so the early mammals that survived tended to be nocturnal insectivores.

In mammals, visual processing did not assume increased importance until the demise of the dinosaurs. At night, olfaction rather than vision was the more beneficial sense. However, by 55 million years ago, some primates with more evolved occipital (vision related) lobes had emerged from the underbrush and the darkness of night.

Mammals have the largest brains relative to body size of all the organisms on Earth. In addition, only mammals have a neocortex, the outer layer of the cerebral hemispheres that participates in higher functions like sensory perception, refined motor coordination, and (eventually) language.

Throughout mammalian and primate evolution, there has been a gradual increase in brain size, superimposed with spikes of fast growth. In addition, over time, different parts of our brain have increased in size at different rates. Brain evolution, from the earliest shrew-like mammals through primates to hominids, is marked by a steady increase in the ratio of brain to body size. The cerebral cortex has expanded more than other areas, and within the cortex, some areas have expanded to a much greater extent than others.

The mammalian brain is built on the components of the reptilian brain, but with two new structures. The neocerebellum ("new cerebellum") is added to the cerebellum at the base of the brain, and the neocortex ("new cortex") grows out of the front of the forebrain. The cerebral cortex, specifically the frontal lobes is the part of the brain that most strongly distinguishes mammals from other vertebrates, primates from other mammals, and humans from other primates. In most mammals, these new additions are not particularly large relative to the brain stem. In primates they are much larger, and in the human they are so large that the original brain stem is almost completely hidden by this large convoluted mass of grey neural matter. The initiation of voluntary behavior as well as the ability to plan, engage in conscious thought, and use language, all depend on the new neocortical structures.

In the last 3–4 million years brain volume within the hominid lineage has increased from less than 400 cc. to roughly 1400 cc. This is an evolutionarily significant change that cannot be simply accounted for in terms of increased body size alone. From the australopithecines through *H. habilis* brain size doubled from about 400 cc to 800 cc. From the appearance of *H. erectus* to the present, brain size doubled again, this time without any significant change in body size.

Evolutionary changes in the brain clearly correlate with evolutionary changes in consciousness. In addition to the fossil record documenting brain evolution, another line of evidence related to the evolution of consciousness is the gradual development of culture. Culture in this context refers to knowledge, belief, and behavior that arise from symbolic thought and social learning. Clues to the cultures of the various species of the genus *Homo,* and hence to their levels of consciousness, can be found in the artifacts they left behind.

The genus *Homo* is defined by the ability to make and utilize stone tools. The stone-tool industries are divided into three main categories; Paleolithic (old stone age), Mesolithic (middle stone age), and Neolithic (new stone age), the categories based on the increasing sophistication of the tools. Although it is probable that the initiation of stone tool use resulted from the development of new mental functions, further advances in stone tool technology may not have required new mental capabilities. However, the gradual development of language and art probably did result from an increased level of consciousness.

Anatomical fossil evidence for the capability of speech is controversial. Less controversial is the archaeological discovery in South Africa, dating from about 75,000 years ago, of abstract geometric designs incised into lumps of red ochre. These designs are generally accepted by archaeologists as evidence for language and symbolic reasoning. John Hoffecker, the discoverer, concludes that the symbolism inherent in the South African artifacts "indicates a thoroughly modern capacity for novelty and invention." Also found at the South African site were perforated shells that appear to have been strung as beads and worn as jewelry. From this point onward there is a growing variety of new types of artifacts.

Another artifact showing evidence of workmanship with an artistic purpose is a carved piece of mammoth ivory dated to the Upper Palaeolithic, more than 40,000 years ago. It appears to be the head of a small figurine. From the Upper Palaeolithic through the Mesolithic, cave paintings and portable art, such as figurines and beads, predominated. Decorative workings are also seen on some utilitarian objects. In the Neolithic, evidence of early pottery appeared, as did sculpture and the construction of megaliths. Ancient musical instruments and figurative art discovered in caves in France and Germany date to before 30,000 years ago.

Exactly how these cultural advances relate to the evolution of consciousness will probably never be known. However, the fact that they are related to the evolution of consciousness is difficult to deny.