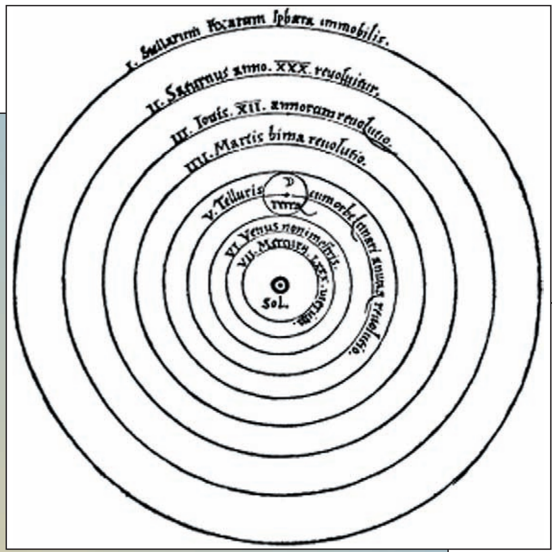
CLBE001-ESS2E November 11, 2006 4:27

A History of Solar System Studies

David Leverington
BAE Systems, England (retired)



CHAPTER 3

This chapter gives a brief overview of the history of solar system research from the earliest times up to the start of the space age.

1. Babylonians and Greeks

Many early civilizations studied the heavens, but it was the Babylonians of the first millennium B.C. who first used mathematics to try to predict the positions of the Sun, Moon, and visible planets (Mercury, Venus, Mars, Jupiter, and Saturn) in the sky. In this they differed from the Greeks, as the Babylonians were priests trying to predict the movement of the heavenly bodies for religious purposes, whereas the Greeks were philosophers trying to understand why they moved in the way they did. The Babylonians were fascinated by numbers, whereas the Greeks were more interested in geometrical figures.

The accuracy of the Babylonian predictions in the 2nd century B.C. is remarkable. For example, their estimate of the length of the **sidereal** year was within 6 minutes of its true value, and that of the average **anomalistic month** was within 3 seconds. In addition, Jupiter's sidereal and **synodic periods** were within 0.01% of their correct values.

Pythagoras (c. 580–500 B.C.) was a highly influential early Greek philosopher who set up a school of philosophers, now known as the Pythagoreans. None of Pythagoras' original writings survive, but later evidence suggests that the Pythagoreans were probably the first to believe that the Earth is spherical, and that the planets all move in separate orbits inclined to the celestial equator. But the Pythagorean spherical Earth did not spin and was surrounded by a series of concentric, crystalline spheres supporting the Sun, Moon, and individual planets. Each had its own sphere, which revolved around the Earth at different speeds, producing a musical sound, the "music of the spheres," as they went past each other.

Hicetus of Syracuse (fl. 5th century B.C.) was the first person to specifically suggest that the Earth spun on its axis, at the center of the universe. This model was further developed by Heracleides who proposed that Mercury and Venus orbited the Sun as it orbited the Earth. Then Aristarchus (c. 310–230 B.C.), who was one of the last of the Pythagoreans, went one step further and proposed a heliocentric (i.e., Sun-centered) universe in which the planets orbit the Sun in the (correct) order of Mercury, Venus, Earth, Mars, Jupiter and Saturn, with the Moon orbiting a spinning Earth. This was 1700 years before Copernicus came up with the same idea. Aristarchus was also the first to produce a realistic estimate for the Earth–Moon distance, although his estimate of the Earth–Sun distance was an order of magnitude too low.

While the Pythagoreans were developing their ideas, Plato (c. 427–347 B.C.) was developing a completely different school of thought. Plato, who was a highly respected philosopher, was not too successful with his geocentric (i.e., Earth-centered) model of the universe. His main legacy to

Encyclopedia of the Solar System 2e ©2007 by Academic Press. All rights of reproduction in any form reserved.

astronomy was his teaching that all heavenly bodies must be spherical, as that is the perfect shape, and that they must move in uniform circular orbits, for the same reason. Aristotle (384–322 B.C.), a follower of Plato, was one of the greatest of Greek philosophers. His ideas were to hold sway in Europe until well into the Middle Ages. However, his geocentric model of the universe was highly complex, requiring a total of 56 spheres to explain the motions of the Sun, Moon, and planets. Unfortunately, many of its predictions were wrong, and it soon fell into disuse.

Hipparchus (c. 185–120 B.C.), who was the first person to quantify the **precession of the equinoxes,** was aware that the Sun's velocity along the ecliptic was not linear. This was known to the Babylonians and to Callippus of Cyzicus, but they did not seek an explanation. Hipparchus, on the other hand, in adopting Plato's philosophy of uniform circular motion in a geocentric universe, realized that this phenomenon could only be explained if the Sun was orbiting an off-center Earth. However, his estimate of the off-center amount was far too large, although his **apogee** position was in error by only 35′.

The mathematician Apollonius of Perga (c. 265–190 B.C.) appears to have been the first to examine the properties of epicycles. These were later adopted by Ptolemy (c. A.D. 100–170) in his geocentric model of the universe. In Ptolemy's scheme (Fig. 1), the Moon, Sun, and planets each describe a circular orbit called an epicycle, the center of which goes in a circle, called a deferent, around a non-spinning Earth. Because the inferior planets, Mercury and Venus, each appear almost symmetrically on both sides of the Sun at maximum **elongation**, he assumed that the centers of their epicycles were always on a line joining the Earth and Sun. For the superior planets he assumed that the lines linking these with the center of their epicycles were always parallel to the Earth–Sun line. Unfortunately, this simple system did not provide accurate enough position estimates, and so Ptolemy introduced a number of modifications. In the case of the Moon, he made the center of the Moon's deferent describe a circle whose center was the Earth. For the planets he introduced the concept of an equant, which was a point in space equidistant with the Earth from the center of the deferent (Fig. 2). The equant was the point about which the planet's angular velocity appeared to be uniform. Other modifications were also required, but by the time he had finished, he was able to make accurate position estimates for all but the Moon and Mercury. In addition, assuming that there were no gaps between the furthest part of one epicycle and the nearest part of the next, he was able to produce an estimate for the size of the solar system of about 20,000 times the radius of the Earth (or about 120 million km). Although this was a gross underestimate, it gave, for the first time, an idea of how large the solar system really was.

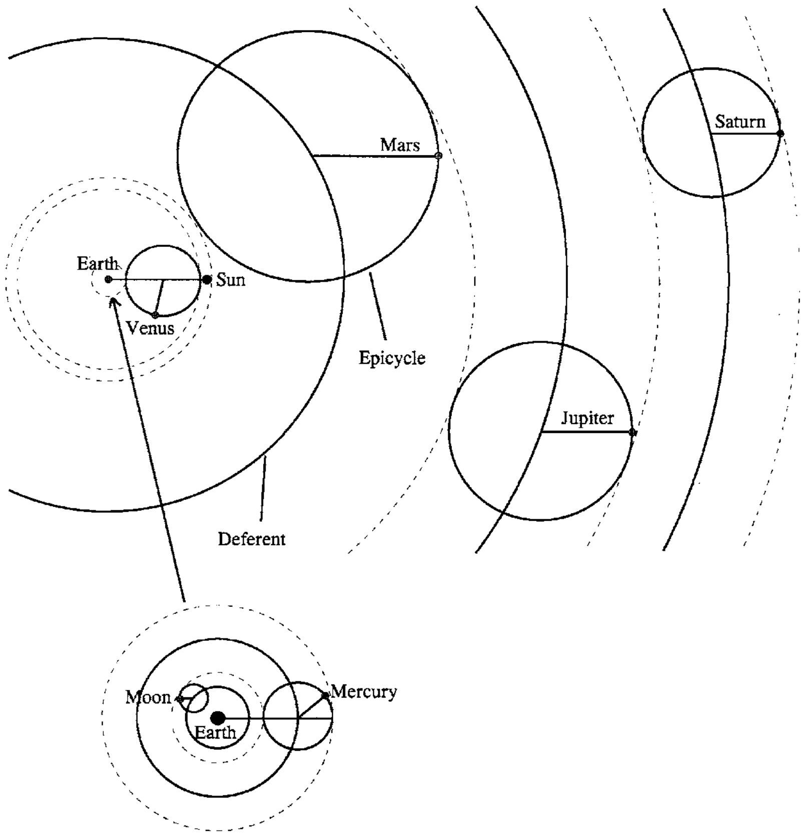


FIGURE 1 Ptolemy's model of the universe in which all bodies, except the Sun (and stars), describe epicycles, the centers of which orbit the Earth in deferents. He assumed that there were no gaps between the circle enclosing the furthest distance of one planet, and that just touching the epicycle of the next planet out from the Earth.

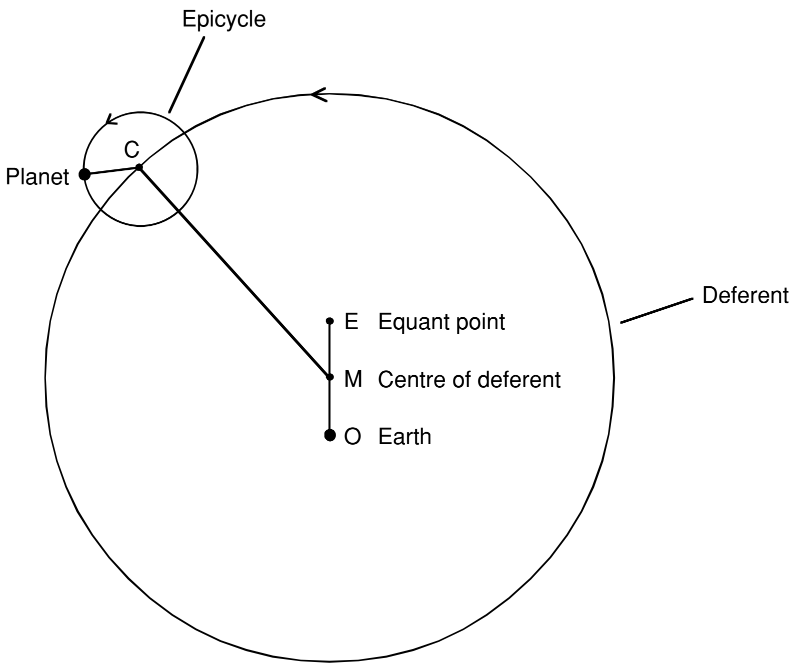


FIGURE 2 Ptolemy modified his epicycle theory for the superior planets by moving the Earth *O* from the center *M* of the deferent, and by defining an equant point *E* such that the distance $EM = MO$. He then assumed that the angular velocity of *C*, the center of the epicycle, is uniform about the equant point *E*, rather than about the center *M* of the deferent.

2. Copernicus and Tycho

There was virtually no progress in astronomy over the next one thousand years, and during this time many of the Greek texts had been lost in Europe. But in the 12th century Arab translations found their way to Europe, mainly via Islamic Spain. Then in the 14th century Ibn al-Shātir (1304–1375), working in Damascus, improved Ptolemy's model by modifying his epicycles and deleting his equant. Interestingly, al-Shātir's system was very much like Copernicus' later system, but with the Earth, not the Sun, at the center.

Copernicus' heliocentric theory of the universe (Fig. 3) was published in his *De Revolutionibus Orbium Caelestium* in 1543, the year of his death. Interestingly, in the light of Galileo's later problems with the Church, the book was well received. This is probably because of the Foreword, which had been written by the theologian Andreas Osiander and explained that the book described a mathematical model of the universe, rather than the universe itself.

Copernicus (1473–1543) acknowledged that his idea of a spinning Earth in a heliocentric universe was not new, having been proposed by Aristarchus. In addition, Copernicus' theory was based on circular motion and still depended on epicycles, although he deleted the equant. But he had resurrected the heliocentric theory, which had not been seriously considered for almost two thousand years, at the height of the Renaissance, which was eager for new ideas.

In the Middle Ages, Aristotle's ideas were taught at all the European universities. But now Copernicus had broken with the Aristotelian concept of a nonspinning Earth at the center of the universe. Then in 1577 Tycho Brahe (1546–1601) disproved another of Aristotle's ideas. Aristotle had believed that comets are in the Earth's atmosphere, but Tycho was unable to measure any clear parallax for the comet of that year. Finally, Tycho, in his book of 1588, rejected another of Aristotle's ideas, that the heavenly bodies are carried in their orbits on crystalline spheres. This is because,

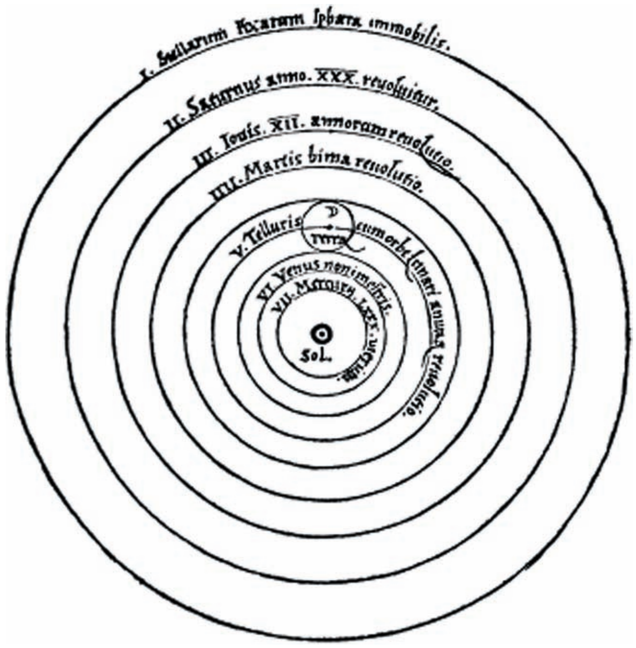


FIGURE 3 Copernicus' heliocentric universe, as described in his *De Revolutionibus,* in which the planets orbit the Sun (Sol) and the Moon orbits the Earth (Terra).

in Tycho's new model of the universe, all the planets, except the Earth, orbit the Sun as the Sun orbits the Earth. This meant that the sphere that carried Mars around the Sun would intercept that which carried the Sun around the Earth, which was clearly impossible if they were crystalline.

3. Kepler and Galileo

Johannes Kepler (1571–1630) looked at the universe in an entirely different way than his predecessors. The Babylonians had examined it arithmetically, and the Greeks and later astronomers had considered it in geometrical terms. Kepler, on the other hand, tried to understand the structure of the solar system by considering physical forces.

Kepler conceived of a force emanating from the Sun that pushed the planets around their orbit of the Sun such that planetary movement would stop if the force stopped. The magnitude of his force, and hence the linear velocity of the planets, decreased linearly with distance. This should have resulted in the period of the planets varying as their distance squared, but Kepler made a mathematical error and came up with another relationship. Fortuitously, however, his analysis produced remarkably accurate results.

Although Kepler was having some success with this and other theories, he thought he could improve them if he had access to Tycho Brahe's accurate observational data. So Kepler went to see Tycho; a visit that ended with him joining Tycho and eventually succeeding him after his death.

Tycho had initially asked Kepler to analyze Mars' orbit, a task that he continued well after Tycho's death. Kepler published his results in 1609 in his book *Astronomia Nova*, in which he reintroduced the equant, previously deleted by Copernicus. In Kepler's model, all the planets orbited the Sun in a circle, with the Sun off-center, but he could not find a suitable circle to match Mars' observations, even with an equant. So he decided to reexamine the Earth's orbit, as the Earth was the platform from which the observations had been made.

Copernicus had proposed that the Earth moved around the Sun in a circle at a uniform speed, with the Sun off-center. So there had been no need for an equant. But Kepler found that an equant was required to explain the Earth's orbit. However, even adding this, he could not fit a circle, or even a flattened circle to Mars' orbit. And so in desperation he tried an ellipse, with the Sun at one focus, and, much to his surprise, it worked.

Kepler now considered what type of force was driving the planets in their orbits, and concluded that the basic circular motion was produced by vortices generated by a rotating Sun. Magnetic forces then made the orbits elliptical. So Kepler thought that the Sun rotated on its axis, and that the planets and Sun were magnetic.

Initially, Kepler had only shown that Mars moved in an ellipse, but in his *Epitome* of 1618–1621 he showed that this was the case for all the planets, as well as the Moon and the satellites of Jupiter. He also stated what we now know as his third law, that the square of the periods of the planets are proportional to the cubes of their mean distances from the Sun. Finally, in his *Rudolphine Tables,* he listed detailed predictions for planetary positions and predicted the transits of Mercury and Venus across the Sun's disc.

Galileo Galilei (1564–1642) made his first telescopes in 1609 and started his first telescopic observations of the Moon in November of that year. He noticed that the **terminator** had a very irregular shape and concluded that this was because the Moon had mountains and valleys. It was quite unlike the pure spherical body of Aristotle's cosmology.

Galileo undertook a series of observations of Jupiter in January 1610 and found that it had four moons that changed their positions from night to night (Fig. 4). Galileo presented his early Moon and Jupiter observations in his *Sidereus Nuncius* published in March 1610. By 1612, he had determined the periods of Jupiter's moons to within a few minutes.

Galileo's *Sidereus Nuncius* created quite a stir, with many people suggesting that Galileo's images of Jupiter's moons were an illusion. Kepler, who was in communication with Galileo, first saw the moons himself in August 1610 and supported Galileo against his doubters. The month before, Galileo had also seen what he took to be two moons on either side of Saturn, but for some reason they did not move. Finally in late 1610 he observed the phases of Venus, finally proving that Ptolemy's structure of the solar system was incorrect. As a result, Galileo settled on the Copernican heliocentric system.

Sunspots had been seen from time to time in antiquity, but most people took them to be something between the

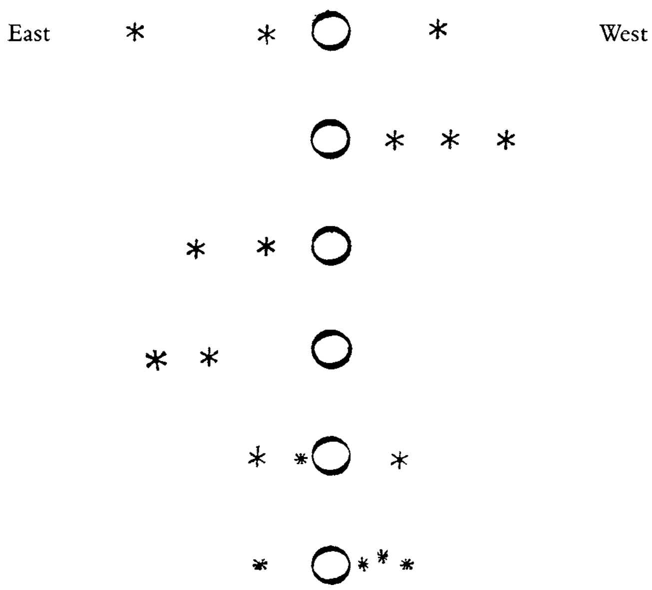


FIGURE 4 Galileo's observations of the moons of Jupiter on consecutive nights from 7 to 13 January (excluding 9 January) 1610, as shown in his book *Sidereus Nuncius.*

Earth and Sun. Although Thomas Harriot and Galileo had both seen sunspots telescopically in 1610, it was Johann Fabricius who first published his results in June 1611. He concluded that they were on the surface of the Sun, and that their movement indicated that the Sun was rotating. This was completely against Aristotle's teachings that the Sun was a perfect body.

In the meantime, Galileo had visited the Jesuits of the Roman College to get their support for his work and, in particular, their support for Copernicus' heliocentric cosmology. His reception was very warm, and he was even received in audience by the pope. But, although the Roman Catholic Church did not argue with his observations, outlined above, there was considerable unease at his interpretation. Initially, the Church was prepared to tolerate Galileo's support of the Copernican cosmology, provided he presented this cosmology as a working hypothesis, rather than as a universal truth. But Galileo was stubborn and tried to take on the Church in its interpretation of theology. In this he could not win, of course, and the Church put him on trial, where he was treated very well. Nevertheless, he was forced in 1633 to recant his views and was then placed under house arrest for the remaining nine years of his life.

4. Second Half of the 17th Century

4.1 The Moon

Thomas Harriot (1560–1621) was the first astronomer to record what we now know as the libration in latitude of the Moon, which has a period of one month. This occurs because the Moon's spin axis is not perpendicular to its orbit. A little later Galileo detected a libration in longitude, which he thought had a period of one day. In fact, it has a period of one month and is caused by the eccentricity of the Moon's orbit.

Although Galileo thought that the Moon has an atmosphere, he concluded that there was very little water on the surface as there were no clouds. His early telescopes were not sufficiently powerful, however, to show much surface detail. But over the next few decades, maps of the Moon were produced by a number of astronomers. The most definitive of which were published in 1647 by Johannes Hevelius (1611–1687). They were the first to show the effect of libration.

By midcentury, it was clear that there were numerous craters on the Moon, and in 1665 Robert Hooke (1635–1703) speculated on their cause in his *Micrographia*. He undertook laboratory-like experiments and noted that if round objects were dropped into a mixture of clay and water, features that resemble lunar craters were produced. But he could not think of the source of large objects hitting the Moon. However, he also found that he could produce crater-like features if he boiled dry alabaster powder in a container. As a result, he concluded that lunar craters are produced by the collapsed blisters of warm viscous lava.

4.2 Saturn

Christiaan Huygens (1629–1695) and his brother Constantyn finished building a state-of-the-art telescope in early 1655. Shortly afterwards Christiaan discovered Saturn's first Moon, Titan, which he announced in his *De Saturni* of 1656. The next four moons of Saturn were discovered by Gian Domenico Cassini (1625–1712); Iapetus in 1671, Rhea in 1672, and both Tethys and Dione in 1684.

Huygens had also mentioned in *De Saturni* that he had solved the problem of Saturn's two "moons" observed by Galileo. In fact, the behavior of these moons had been very odd, as they had both completely disappeared in November 1612, reappearing again in mid 1613. Since then, their shape had gradually changed. In 1650, Francesco Grimaldi discovered Saturn's polar flattening, but still the behavior of the moons, then called ansae, was unexplained. Finally, Huygens announced, in his *Systema Saturnium* of 1659, that the ansae were actually a thin, flat, solid ring, which was inclined to the ecliptic, and so changed its appearance with time. Then in 1675 Cassini noticed that Saturn's ring was divided in two by a dark line, now called the Cassini Division, going all the way around the planet. Cassini speculated that the two rings were not solid but composed of swarms of small satellites.

Other major observational discoveries of this period are listed in Table 1.

4.3 Newton

Kepler had thought that the planets were being pushed around their orbits by a vortex emanating from the Sun but attributed the tides on Earth to the combined attraction of the Sun and Moon by a gravitational force. It seems strange to us that he did not think of this attractive force as having some effect on the orbits of the planets.

René Descartes (1596–1650) also developed a vortex theory to explain the motion of the planets. In his theory, the vortices are in the ether, which is a frictionless fluid filling the universe. In his *Principia* of 1644, Descartes stated that each planet had two "tendencies": one tangential to its orbit and one away from the orbit's center. It is the pressure in the vortex that counterbalances the latter and keeps the planet in its orbit.

In 1664, Isaac Newton (1642–1727) started to consider the motion of a body in a circle. In the following year, he proved that the force acting radially on such a body is proportional to its mass multiplied by its velocity squared, and divided by the radius of the circle (i.e., mv^2/r). From this, he was able to prove that the force on a planet moving in a circular orbit is inversely proportional to the square of its distance from the center. Newton realized that this outward centrifugal force on a planet must be counterbalanced by an equal and opposite centripetal force, but it was not obvious at that time that this force was gravity.

CLBE001-ESS2E November 11, 2006 4:27

TABLE 1 Key Solar System Discoveries and Observations, 1630–1700

Sun-Earth distance	
1672	Richer, Cassini, and Picard deduce a solar parallax of 9.5 minutes of arc from observations of the parallax of Mars. John Flamsteed independently deduces a similar value. This implied a Sun-Earth distance of about 22,000 earth radii, or 140 million km.
Moon	
See main text	
Mercury	
1631	First observation of a transit of Mercury by Gassendi, Remus, and Cysat—all independently. It occurred on the date predicted by Kepler.
1639	Phases of Mercury first observed by Zupus.
Venus	
1639	First observation of a transit of Venus by Horrocks and Crabtree.
1646	Fontana observes that Venus' terminator is uneven, attributing the cause to high mountains. (This is now known to be incorrect; Venus is covered in dense clouds.)
1667	Cassini deduces a rotation period of about 24 hours. (This is now known to be incorrect).
Mars	
1659	Huygens observes Syrtis Major and deduces a planetary rotation period of about 24 hours.
1672	Huygens first unambiguously records the south polar cap.
Jupiter	
c. 1630	Fontana, Torricelli, and Zucchi independently observe the main belts.
1643	Riccioli observes the shadows of the Galilean satellites on Jupiter's disc.
1663	Cassini deduces a Jupiter rotation period of 9 h 56 min.
1665	Cassini observes a prominent spot that may be an early appearance of the Great Red Spot.
1690	Cassini observes the differential rotation of Jupiter.
1691	Cassini observes Jupiter's polar flattening, which he estimates to be about 7%.
Saturn	
See main text	

At this time, it was known that gravity acted on objects on the Earth's surface, but it was not known how far from Earth gravity extended. To get a better understanding of this, Newton devised his so-called Moon test. In this test, he compared the force acting on the Moon, because of its motion in a circle, with the force of the Earth's gravity at the Moon's orbit and found that they were not the same. The difference was not large, but it was sufficient to cause Newton to stop work on gravity. In fact, at that time, Newton appears to have thought that the centripetal force was a mixture of the gravitational force and the force created by vortices in the ether, so he may not have been too surprised by his result.

Newton was finally prompted to return to the subject of gravity by an exchange of letters with Robert Hooke in 1679. In the following year, Newton proved that, assuming an inverse square law of attraction, planets and moons will orbit a central body in an ellipse, with the central body at one focus. Then in 1684 he finally rejected the idea of etherial vortices and started to develop his theory of universal gravitation.

It was during this period that the comet of 1680 appeared. At that time, most astronomers, including Newton, believed that comets described rectilinear orbits. John Flamsteed (1646–1719), on the other hand, believed that comets described closed orbits, and he suggested, in a letter to Edmond Halley (1656–1742), that the 1680 comet had passed in front of the Sun. Newton, who had been sent a copy of this letter, thought, like a number of astronomers, that there had been two comets, one approaching the Sun and one retreating. Further communications between Flamsteed and Newton in 1681 did not resolve their disagreements, causing Newton to drop the subject of cometary orbits. Eventually, Newton returned to the subject, and by 1686 he had changed his position entirely, as he proved that cometary orbits are highly elliptical or parabolic, to a first approximation. So the 1680 comet had been one comet after all. Newton now felt, having solved the problem of cometary orbits, that he could complete his *Principia*, which was published in 1687.

Newton developed his universal theory of gravitation in his *Principia,* which ran to three editions. For example, he used Venus to "weigh" the Sun, and planetary moons to weight their parent planets, and by the third edition he had deduced the masses and densities for the Earth, Jupiter, and Saturn relative to the Sun (Table 2).

TABLE 2 **A Comparison of Newton's Results (Relative to the Sun) with Modern Values**

	Mass		Density	
	Principia	Modern Value	Principia	Modern Value
Sun	1	1	100	100
Earth	1/169,282	1/332,980	400	392
Jupiter	1/1,067	1/1,047	94.5	94.2
Saturn	1/3,021	1/3,498	67	49

Newton realized that if gravity was really universal, then not only would the Sun's gravity affect the orbit of a planet, and the planet's gravity affect the orbit of its moons, but the Sun would also affect the orbits of the moons, and one planet would affect the orbits of other planets. In particular, Newton calculated that Jupiter, at its closest approach to Saturn, would have about 1/217 times the gravitational attraction of the Sun. So he was delighted when Flamsteed told him that Saturn's orbit did not seem to fit exactly the orbit that it should if it was only influenced by the Sun. Gravity really did appear to be universal.

Richer, Cassini, and Picard had found evidence in 1672 that the Earth had an equatorial bulge. Newton was able to use his new gravitational theory to calculate a theoretical value for this **oblateness** of 1/230 (modern value 1/298). He then considered the gravitational attraction of the Moon and Sun on the oblate Earth and calculated that the Earth's spin axis should precess at a rate of about 50″.0 per annum (modern value 50″.3). This explained the precession of the equinoxes.

5. The 18th Century

5.1 Halley's Comet

Halley used Newton's methodology to determine the orbits of 24 comets that had been observed between 1337 and 1698. None of them appeared to be hyperbolic, and so the comets were all clearly permanent members of the solar system. Halley also concluded that the comets of 1531, 1607, and 1682 were successive appearances of the same comet as their orbital elements were very similar. But the time intervals between successive perihelia were not the same; a fact he attributed to the perturbing effect of Jupiter. Taking this into account, he predicted in 1717 that the comet would return in late 1758 or early 1759.

Shortly before the expected return of this comet, which we now called Halley's comet, Alexis Clairaut (1713–1765) attempted to produce a more accurate prediction of its **perihelion** date. He used a new approximate solution to the three-body problem that allowed him to take account of planetary perturbations. This showed that the return would be delayed by 518 days due to Jupiter and 100 days due to Saturn. As a result, he predicted that Halley's comet would reach perihelion on about 15 April 1759 ± 1 month. It did so on 13 March 1759, so Clairaut was just 33 days out with his estimate.

5.2 The 1761 and 1769 Transits of Venus

James Gregory (1638–1675) had suggested in 1663 that observations of a transit of Mercury could be used to determine the **solar parallax**, and hence the distance of the Sun from Earth. Such a determination required observations from at least two different places on Earth, separated by as large a distance as possible. In 1677, Edmond Halley observed such a transit when he was on St. Helena observing the southern sky. But, when he returned, he found that Jean Gallet in Avignon seemed to have been the only other person who had recorded the transit. Unfortunately, there were too many problems in comparing their results, which resulted in a highly inaccurate **solar parallax.**

In 1678, Halley reviewed possible methods of measuring the solar parallax and suggested that transits of Venus would produce the most accurate results. The problem was, however, that these occur in pairs, 8 years apart, only every 120 years. The next pair were due almost one hundred years later, in 1761 and 1769.

Joseph Delisle (1688–1768) took up Halley's suggestion and tried to motivate the astronomical community to undertake coordinated observations of the 1761 transit. After much discussion, the French Academy of Sciences sent observers to Vienna, Siberia, India, and an island in the Indian Ocean, while other countries sent observers to St. Helena, Indonesia, Newfoundland, and Norway. Unfortunately, precise timing of the planetary contacts proved much more difficult than expected, resulting in solar parallaxes ranging from 8″.3 to 10″.6. Interestingly, several observers noticed that Venus appeared to be surrounded by a luminous ring when the planet was partially on the Sun. Mikhail Lomonsov (1711–1765) correctly concluded that this showed that Venus was surrounded by an extensive atmosphere.

The lessons learned from the 1761 transit were invaluable in observing the next transit in 1769. This was undertaken from over 70 different sites, and analysis of all the results eventually yielded a best estimate of 8″.6 (modern value 8″.79) for the solar parallax.

5.3 The Discovery of Uranus

On 13 March 1781, William Herschel (1738–1822), whilst looking for double stars, noticed what he thought was a comet. Four days later, when he next saw the object, it had clearly moved, confirming Herschel's suspicion that it was a comet. He then wrote to Nevil Maskelyne (1732–1811), the Astronomer Royal, notifying him of his discovery. As a result, Maskelyne observed the object on a number of occasions, but he was unsure as to whether it was a comet or a new planet.

Over the next few weeks a number of astronomers observed the object and calculated its orbit, which was found to be essentially circular. So it was a planet, now called Uranus. It was the first planet to be discovered since ancient times, and its discovery had a profound effect on the astronomical community, indicating that there may yet be more undiscovered planets in the solar system.

A few years later Herschel discovered the first two of Uranus' satellites, now called Titania and Oberon, with orbits at a considerable angle to orbit.

5.4 Origin of the Solar System

Immanuel Kant (1724–1804) outlined his theory of the origin of the solar system in his *Universal Natural History* of 1755. In this he suggested that the solar system had condensed out of a nebulous mass of gas, which had developed into a flat rotating disc as it contracted. As it continued to contract, it spun faster and faster, throwing off masses of gas that cooled to form the planets. However, Kant had difficulty in explaining how a nebula with random internal motions could start rotating when it started to contract.

Forty years later, Laplace (1749–1827) independently produced a similar but more detailed theory. In his theory, the mass of gas was rotating before it started contracting. As it contracted, it spun faster, progressively throwing from its outer edge rings of material that condensed to form the planets. Laplace suggested that the planetary satellites formed in a similar way from condensing rings of material around each of the protoplanets. Saturn's rings did not condense to form a satellite because they were too close to the planet. At face value, the theory seemed plausible, but it became clear in the 19th century that the original solar nebula did not have enough angular momentum to spin off the required material.

5.5 The First Asteroids

A number of astronomers had wondered why there was such a large gap in the solar system between the orbits of Mars and Jupiter. Then in 1766 Johann Titius (1729–1796) produced a numerical series that indicated that there should be an object orbiting the Sun with an orbital radius of 2.8 **astronomical units** (AUs). Johann Elert Bode (1747–1826) was convinced that this was correct and mentioned it in his book of 1772. However, what is now known as the Titius–Bode series was not considered of any particular significance, until Uranus was found with an orbital radius of 18.9 AU. This was very close to the 19.6 AU required by the series.

In 1800, a group of astronomers, who came to be known as the Celestial Police, agreed to undertake a search for the missing planet. But before they could start Giuseppe Piazzi (1746–1826) found a likely candidate by accident in January 1801. Unfortunately, although he observed the object for about 6 weeks, he was unable to fit an orbit, and wondered if it was a comet. But Karl Gauss (1777–1855) had derived a new method of determining orbits from a limited amount of information, and in November of that year he was able to fit an orbit. It was clearly a planet, now called Ceres, at almost exactly the expected distance from the Sun. But it was much smaller than any other planet. Then in March 1802 Heinrich Olbers (1758–1840) found another, similar object, now called Pallas, at a similar distance from the Sun. At first Olbers thought that these two objects may be the remnants of an exploded planet. But he dropped the idea after the discovery of the fourth such asteroid, as they are now called, in 1807, because its orbit was inconsistent with his theory.

6. The 19th Century

6.1 The Sun

Sunspots were still an enigma in the 19th century. Many astronomers thought that they were holes in the photosphere, but because the Sun was presumably hotter beneath the photosphere, the Sunspots should appear bright rather than dark. Then in 1872 Angelo Secchi suggested that matter was ejected from the surface of the Sun at the edges of a sunspot. This matter then cooled and fell back into the center of the spot, so producing its dark central region.

In 1843, Heinrich Schwabe found that the number of sunspots varied with a period of about 10 years. A little later Rudolf Wolf analyzed historical records that showed periods ranging from 7 to 17 years, with an average of 11.1 years. Then in 1852, Sabine, Wolf, and Gautier independently concluded that there was a correlation between sunspots and disturbances in the Earth's magnetic field. There were also various unsuccessful attempts to link the sunspot cycle to the Earth's weather. But toward the end of the century, Walter Maunder pointed out that there had been a lack of sunspots between about 1645 and 1715. He suggested that this period, now called the Maunder Minimum, could have had a more profound effect on the Earth's weather than the 11-year solar cycle.

In 1858, Richard Carrington discovered that the latitude of sunspots changed over the solar cycle. In the following year, he found that sunspots near the solar equator moved faster than those at higher latitudes, showing that the Sun did not rotate as a rigid body. This so-called differential rotation of the Sun was interpreted by Secchi as indicating that the Sun was gaseous. In the same year, Carrington and Hodgson independently observed two white light solar flares moving over the surface of a large sunspot. About 36 hours later, this was followed by a major geomagnetic storm.

Astronomy was revolutionized in the 19th century by Kirchoff's and Bunsen's development of spectroscopy in the early 1860s, which, for the first time, enabled astronomers to determine the chemical composition of celestial objects. Kirchoff measured thousands of dark Fraunhofer lines in the solar spectrum and recognized the lines of sodium and iron. By the end of the century, about 40 different elements had been discovered on the Sun.

Solar prominences had been observed during a total solar eclipse in 1733, but it was not until 1860 that they were proved to be connected with the Sun rather than the Moon. Spectroscopic observations during and after the 1868 total eclipse showed that prominences were composed of hydrogen and an element that produced a bright yellow line. This was initially attributed to sodium, but Norman Lockyer suggested that it was caused by a new element that he called helium. This was confirmed when helium was found on Earth in 1895.

6.2 Vulcan

Newton's gravitational theory had been remarkably accurate in explaining the movement of the planets, but by the 19th century there appeared to be something wrong with the orbit of Mercury. In 1858, Le Verrier analyzed data from a number of transits and concluded that the perihelion of Mercury's orbit was precessing at about 565″/century, which was 38″/century more than could be accounted for using Newton's theory. As a result, Le Verrier suggested that there was an unknown planet called Vulcan, inside the orbit of Mercury, causing the extra precession. A number of astronomers reported seeing such a planet, but none of the observations stood up to detailed scrutiny, and the idea was eventually dropped.

Einstein finally solved the problem of Mercury's perihelion precession in 1915 with his general theory of relativity. No extra planets were required.

6.3 Mercury

There was considerable disagreement among astronomers in the 19th century on what could be seen on Mercury. Some thought that they could see an atmosphere around the planet, but others could not. Hermann Vogel detected water vapor lines in its spectrum, and Angelo Secchi saw clouds in its atmosphere. However, Friedrich Zöllner concluded, from his photometer measurements, that Mercury was more like the Moon with, at most, a very thin atmosphere.

A number of astronomers detected markings on Mercury's disc in the middle of the 19th century and concluded that the planet's period is about 24 hours. On the other hand, Daniel Kirkwood maintained that it should have a **synchronous rotation** period because of tidal effects of the Sun on its crust. In the 1880s, Giovanni Schiaparelli confirmed this synchronous rotation observationally, and in 1897 Percival Lowell came to the same conclusion. So at the end of the century, synchronous rotation was thought to be the most likely.

6.4 Venus

In the 18th century, Venus was thought to have an axial rotation rate of about 24 hours. In fact, a 24-hour period was generally accepted until in 1890 Schiaparelli and others concluded that it, like Mercury, has a synchronous rotation period.

Spectroscopic observations of Venus yielded conflicting results in the 19th century. A number of astronomers detected oxygen and water vapor lines in its atmosphere; however, W. W. Campbell, who used the powerful Lick telescopes, could find no such lines.

6.5 The Moon

The impact theory for the formation of lunar craters was resurrected at the start of the 19th century, after the discovery of the first asteroids and a number of meteorites. There now seemed to be a ready source of impacting bodies, which Hooke had been unaware of when he had abandoned his impact hypothesis. But both the impact and volcanic theories still had problems. Most meteorites would not hit the lunar surface vertically, and so the craters should be elliptical, but they were mostly circular. Also, as Grove K. Gilbert pointed out, the floors of lunar craters are generally below the height of their surrounding area, whereas on Earth the floors of volcanic craters are generally higher than their surroundings.

Edmond Halley had discovered in 1693 that the Moon's position in the sky was in advance of where it should be based on ancient eclipse records. This so-called secular acceleration of the Moon could be because the Moon was accelerating in its orbit, and/or because the Earth's spin rate was slowing down. In 1787, Laplace had shown that the observed effect, which was about 10″/century^2, could be completely explained by planetary perturbations. But in 1853, John Couch Adams included some of Laplace's second-order terms, which Laplace had omitted, so reducing the calculated figure from 10″/century^2 to just 6″/century^2. Charles Delaunay suggested that the missing amount was probably due to tidal friction, but it was impossible at that time to produce a reasonably accurate estimate of the effect. In the early 20th century, Taylor and Jeffreys produced the necessary calculations, showing that Delaunay was correct.

In 1879, George Darwin developed a theory of the origin of the Moon. In this the proto-Earth had gradually contracted and increased its spin rate as it cooled. Then, when the spin rate had reached about 3 hours per revolution, it had broken into two unequal parts: the Earth and the Moon.

After breakup, tidal forces had caused the Earth's spin rate to slow down and the Moon's orbit to gradually increase in size.

A major problem with this theory was that the Earth would have had a tendency to break up the Moon shortly after separation. It was not clear whether the Moon could have passed through the danger zone before this could have happened.

6.6 The Earth

Karl Friedrich Küstner undertook precise position measurements of a number of stars in 1884 and 1885 from the Berlin Observatory. When he analyzed his results, however, he found that the latitude of the observatory had apparently decreased by about 0.20″ in a year. Intrigued, the International Commission for Geodesy (ICG) decided to organize a series of observations around the world to define the effect more precisely. These results indicated that the Earth's spin axis was moving, relative to its surface, with a period of about 12 or 13 months.

Seth Chandler had also noticed slight variations in the latitude of the Harvard College Observatory, at about the same time as Küstner was making his measurements, but Chandler had not taken the matter further. Galvanized by Küstner's and the ICG's results, however, he undertook a thorough review of all available data. As a result, he concluded that the observed effect had two components. One had a period of 14 months, and was due to the nonrigid Earth not spinning around its shortest diameter. The other, which had a period of a year, was due to the seasonal movement of water and air from one hemisphere to the other and back.

6.7 Mars

The first systematic investigation of Mars' polar caps had been undertaken in the 18th century by Giacomo Maraldi, who found that the south polar cap had completely disappeared in late 1719, only to reappear later. William Herschel suggested that this was because it consisted of ice and snow that melted in the southern summer.

At the end of the 18th century, most astronomers thought that the reddish color of Mars was due to its atmosphere. But in 1830, John Herschel suggested that it was the true color of its surface. Camille Flammarion, on the other hand, hypothesized that it was the color of its vegetation.

It was generally believed by astronomers in the mid-19th century that there must be some form of life on Mars, even if it was only plant life, because the planet clearly had an atmosphere and a surface that exhibited seasonal effects. The polar caps were apparently made of ice or snow, and there were dark areas on the surface that may be seas.

Schiaparelli produced a map of Mars, following its 1877 **opposition**, that showed a network of linear features that he called *canali*. This was translated incorrectly into English as canals, which implied that they had been built by intelligent beings. Schiaparelli and others saw more *canali* in subsequent years (Fig. 5), but other, equally competent observers could not see them at all. Percival Lowell then went further than Schiaparelli in not only observing many canali, but interpreting them to be a network of artificial irrigation channels. At the end of the century, the debate as to whether these *canali* really existed was still in full swing.

Spectroscopic observations of Mars in the late 19th century yielded conflicting results. Some astronomers detected oxygen and water vapor lines, whereas Campbell at the Lick Observatory could find none. There was also a problem with the polar caps: Calculations showed that the average temperature of Mars should be about −34°C, yet both polar caps clearly melted substantially in summer, which they should not have done if they had been made of water ice or snow. In 1898, Ranyard and Stoney suggested that the caps could be made of frozen carbon dioxide. But there appeared to be a melt band at the edge of the caps in spring, yet carbon dioxide should sublimate directly into gas on Mars.

Two satellites of Mars, now called Phobos and Deimos, were discovered by Asaph Hall in 1877. Their orbits were extremely close to the planet, and the satellites were both very small. As a result, they were thought to be captured asteroids.

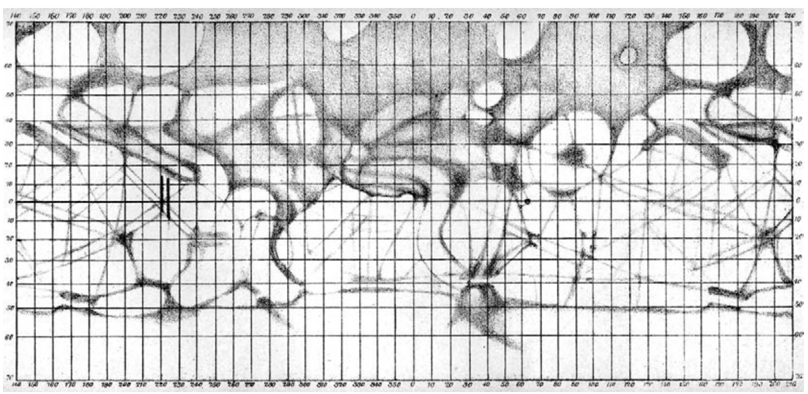


FIGURE 5 Schiaparelli's map of Mars produced following the 1881 opposition. A large number of *canali* are seen, many of them double. (From Robert Ball, 1897, "The Story of the Heavens," Plate XVIII.)

6.8 Jupiter

The Great Red Spot (GRS) was first clearly observed in the 1870s. Then in 1880 an unusually bright, white equatorial spot appeared; it rotated around Jupiter over 5 minutes faster than the GRS. This gave a differential velocity of about 400 km/h. But the rotation rates of both the white spot and the GRS were not constant, indicating that neither could be surface features as some astronomers had supposed.

White and dark spots were continuously appearing and disappearing on Jupiter, suggesting that they were probably clouds. But the GRS was completely different because, although it changed its appearance and size over time, it was still there at the end of the century. This longevity led astronomers to wonder if it could really be a cloud system.

In 1778, Leclerc, Compte de Buffon, had suggested that rapid changes in Jupiter's appearance showed that it had not completely cooled down since its formation. In the 19th century, Jupiter's differential rotation and low density, which were both similar in nature to those of the Sun, caused some astronomers to go even further and wonder if Jupiter was self-luminous. Although this was considered unlikely, the idea had not been completely ruled out by the end of the century.

William Herschel had concluded in 1797 that the axial rotation rates of the four Galilean satellites were synchronous. However, it was not until the 1870s that Engelmann and Burton independently confirmed this for Callisto and the 1890s that Pickering and Douglass confirmed it for Ganymede. The rotation rates of Io and Europa were still unclear.

In 1892, Edward Barnard discovered Jupiter's fifth satellite, now called Amalthea, very close to the planet, when he was observing Jupiter visually through the 36-in. Lick refractor. Amalthea was very small compared to the four Galilean satellites. It was the last satellite of any planet to be discovered visually.

6.9 Saturn

In 1837, Johann Encke found that the A ring was divided into two by a clear gap, now called the Encke Division. Then in 1850 W. C. and G. P. Bond discovered a third ring, now called the C ring, inside the B ring. The new ring was very dark (Fig. 6) and partly transparent. In 1867, Kirkwood pointed out that any particles in the Cassini Division would have periods of about one-half that of Mimas, one-third that of Enceladus, one-quarter that of Tethys, and one-sixth that of Dione. He concluded that these resonances had created the Cassini Division, which would be clear of particles.

The true nature of Saturn's rings had been a complete mystery in the 18th century. Cassini had thought that they may be composed of many small satellites, and Laplace had suggested that they were made of a number of thin solid rings. Others thought that they may be liquid. But in 1857, James Clerk Maxwell proved mathematically that they could not be solid or liquid. Instead, he concluded that they were composed of an indefinite number of small particles.

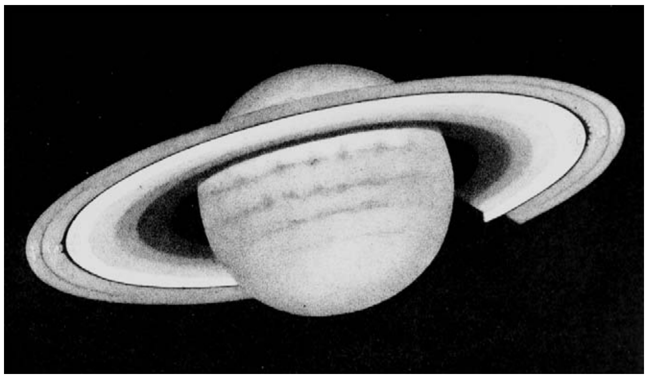


FIGURE 6 Trouvelot's 1874 drawing of Saturn. It clearly shows the dark C ring extending from the inner edge of the B ring to about half-way to the planet. (From Edmund Ledger, 1882, "The Sun: Its Planets and Their Satellites," Plate IX.)

Two new satellites were found in the 19th century: Hyperion by G. P. Bond in 1848 and Phoebe by William Pickering 50 years later. Phoebe was the first satellite in the solar system to be discovered photographically. It was some 13 million kilometers from Saturn, in a highly eccentric, **retrograde** orbit. So it appeared to be a captured object.

6.10 Uranus

Little was know about Uranus in the 19th century. William Herschel had noticed that Uranus had a polar flattening, its orientation indicating that its axis of rotation was perpendicular to the plane of its satellites. But observations of apparent surface features produced very different orientations. Uranus' spectrum appeared to be clearly different from those of Jupiter and Saturn, but it was very difficult to interpret. There was even confusion about the discovery of new satellites. It was not until 1851 that William Lassell could be sure that he had discovered two new satellites, now called Ariel and Umbriel within the orbit of Titania. He had, in fact, seen them both some years before, but his earlier observations had been too infrequent to produce clear orbits.

6.11 The Discovery of Neptune

In 1821, Alexis Bouvard tried to produce an orbit for Uranus using both prediscovery and postdiscovery observations. But he could not find a single orbit to fit them. The best he could manage was an orbit based on only the postdiscovery observations; he published the result but admitted that it was less than ideal. However, it did not take long for Uranus to deviate more and more from even this orbit. One possible explanation was that Uranus was being disturbed by yet another planet, and if the Titius–Bode series was correct it would be about 38.8 AU from the Sun.

In 1843, the Englishman John Couch Adams set out to try to calculate the orbit of the planet that seemed to be disturbing the orbit of Uranus. By September 1845, he had calculated its orbital elements and its expected position in the sky, and over the next year, he progressively updated this prediction. Unfortunately, these predictions varied wildly, making it impossible to use them for a telescopic search of the real planet. In parallel, and unknown to both men, Urbain Le Verrier, a French astronomer, undertook the same task. He published his final results in August 1846 and asked Johann Galle of the Berlin Observatory if he would undertake a telescope search for it. Galle and his assistant d'Arrest found the planet within an hour of starting the search on 23 September 1846. There then followed a monumental argument between the English and French astronomical establishments on the priority for the orbital predictions. But much of the evidence on the English side was never published, and an "official line" was agreed. That evidence has recently come to light, however, and it is currently being analyzed to establish the exact sequence of events. What is clear, however, is that when Neptune's real orbit was calculated, it turned out to be quite different from either of the orbits predicted by Le Verrier or Adams. So its discovery had been somewhat fortuitous.

Less than a month after Neptune's discovery, William Lassell observed an object close to Neptune, which he thought may be a satellite. It was not until the following July that he was able to confirm his discovery of Neptune's first satellite, now called Triton. Triton was later found to have a retrograde orbit inclined at approximately 30° to the ecliptic.

6.12 Asteroids

The fourth asteroid, Vesta, had been discovered in 1807, but it was not until 1845 that the fifth asteroid was found. Then the discovery rate increased rapidly so that nearly 500 asteroids were known by the end of 1900. As the number of asteroids increased, Kirkwood noticed that there were none with certain fractional periods of Jupiter's orbital period. This he attributed to resonance interactions with Jupiter.

All the early asteroids had orbits between those of Mars and Jupiter, and even as late as 1898 astronomers had discovered only one that had part of its orbit inside that of Mars. But in 1898, Eros was found with an orbit that came very close to that of the Earth, with the next closest approach expected in 1931. This could be used to provide an accurate estimate of solar parallax.

In 1906, two asteroids were found at the Lagrangian points, 60° in front of and behind Jupiter in its orbit. They were the first of the so-called Trojan asteroids to be discovered.

6.13 Comets

Charles Messier discovered a comet that passed very close to the Earth in 1770. Anders Lexell was the first to fit an orbit to it, showing that it had a period of just 5.6 years. With such a short period it should have been seen a number of times before, but it had not. As Lexell explained, this comet had not been seen because it had passed very close to Jupiter in 1767, which had radically changed its orbit. In the late 19th century, Hubert Newton examined the effect of such planetary perturbations on the orbits of comets and found that, for a random selection of comets, they were remarkably inefficient. Lexell's comet appeared to be an exception.

Jean Louis Pons in 1818 discovered a comet that, on further investigation, proved to have been seen near previous perihelia. In the following year, Johann Encke showed that the comet, which now bears his name, has an orbit that takes it inside the orbit of Mercury. When the comet returned in 1822, Encke noticed that it was a few hours early and suggested that it was being affected by some sort of resistive medium close to the Sun. In 1882, however, a comet passed even closer to the Sun and showed no effect of Encke's medium. Then in 1933, Wolf's comet was late, rather than early. The problem of these cometary orbits was finally solved in 1950 when Fred Whipple showed that the change in period was caused by jetlike, vaporization emissions from the rotating cometary nucleus.

The first successful observation of a cometary spectrum was made by Giovanni Donati in 1864. When the comet was near the Sun, it had three faint luminous bands, indicating that it was self-luminous. Then four years later, William Huggins found that the bands were similar to those emitted by hydrocarbon compounds in the laboratory.

Quite a number of cometary spectra were recorded over the next 20 years. When they were first found, they generally exhibited a broad continuous spectrum like that of the Sun indicating that they were scattering sunlight. As they got closer to the Sun, however, the hydrocarbon bands appeared. Then in 1882 Wells' comet approached very close to the Sun. Near perihelion its bandlike structure disappeared to be replaced by a bright, double sodium line. In the second comet of 1882, this double sodium line was also accompanied by several iron lines when the comet was very near the Sun. As the comet receded, these lines faded and the hydrocarbon bands returned.

6.14 Meteor Showers

A spectacular display of shooting stars was seen in November 1799, and again in November 1833. They seemed to originate in the constellation Leo. In the following year, Denison Olmsted pointed out the similarities between these two meteor showers and a less intense one in 1832. These so-called Leonid meteors seemed to be an annual event occurring on or about 12 November. Olmsted explained that the radiant in Leo was due to a perspective

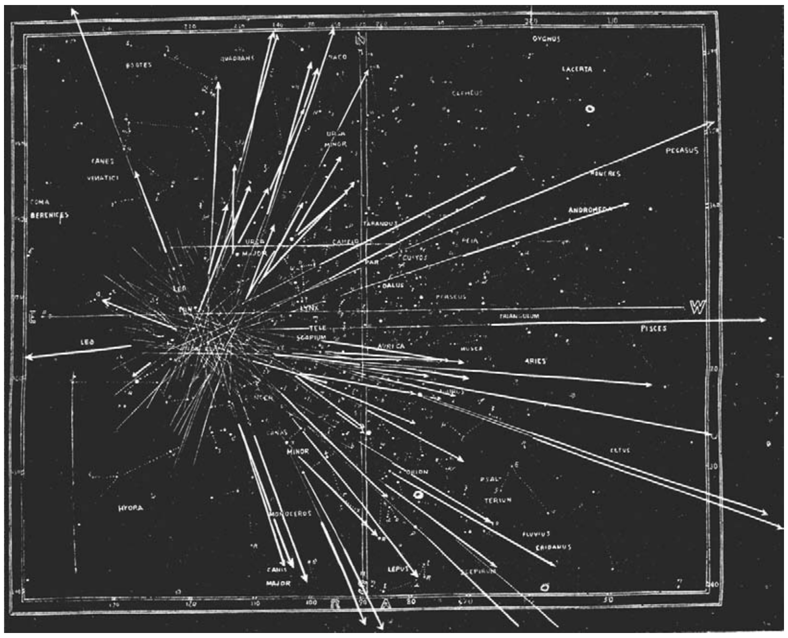


FIGURE 7 Paths of the Leonid meteors showing their apparent origin from a common radiant due to parallax. (From Simon Newcomb, 1898, "Popular Astronomy," p. 403.)

effect (Fig. 7). A similar effect was then observed for a meteor shower on 8 August 1834, which appeared to have a radiant in Perseus. Shortly afterward, Lambert Quetelet showed that these were also an annual event.

In 1839, Adolf Erman suggested that both the Leonid and Perseid meteor showers were produced by the Earth passing through swarms of small particles that were orbiting the Sun and spread out along Earth's orbit. But it was still unclear as to the size of the orbit. In 1864, Hubert Newton found that the node of the Leonids' orbit was precessing at about 52″/year. John Couch Adams then showed that only a particle in a 33.25-year orbit would have this nodal precession. So the Leonids were orbiting the Sun in a diffuse cloud every 33.25 years, which explained why the most intense showers occurred with this frequency. The stragglers all around the orbit explained why we saw the Leonids on an annual basis. In 1867, Carl Peters recognized that the source of the Leonid meteor stream was a periodic comet called Tempel–Tuttle. This was just after Schiaparelli had linked the Perseids to another periodic comet, Swift–Tuttle.

7. The 20th Century Prior to the Space Age

7.1 The Sun

In the 19th century, most physicists had thought that heat was transported from the interior to the exterior of the Sun by convection. But in 1894, R. A. Sampson suggested that the primary mechanism was radiation. Then, 30 years later, Arthur Eddington used the concept of radiative equilibrium to calculate the temperature at the center of the Sun and found it to be about 39 million K. At about the same time, Cecilia Payne showed that hydrogen and helium were the most abundant elements in the stars. Although this idea was initially rejected, it was soon accepted for both the Sun and stars. As a result, in 1935 Eddington reduced his temperature estimate for the center of the Sun to 19 million K.

However, Eddington's calculations made no assumption on how the Sun's heat was produced, which was still unknown at the time. Earlier, in 1920, Eddington himself had proposed two alternative mechanisms. The heat could be produced either by the mutual annihilation of protons and electrons or by the fusion of hydrogen atoms into helium atoms in some unknown manner. There were other mechanisms suggested by other physicists, but the issue could not be resolved at the time because nuclear physics was still in its infancy. The breakthrough came in 1938 when Charles Critchfield explained how energy could be produced at high temperatures by a chain reaction starting with proton–proton collisions and ending with the synthesis of helium nuclei. Hans Bethe then collaborated with Critchfield to develop this idea. But Bethe also examined an alternative mechanism that relied on carbon as a catalyst to produce helium from hydrogen, in the so-called carbon cycle. Carl von Weizsäcker independently developed this same scheme. Which mechanism was predominant in the Sun depended crucially on temperature, and it was not until the 1950s that it became clear that the proton–proton chain is dominant in the Sun.

In the 19th century, the **corona** had been found to have a faint continuous spectrum crossed by Fraunhofer absorption lines, but the conditions in the corona were unclear. Of particular interest was a bright green emission line in the coronal spectrum; Young and Harkness found it in 1869 and originally attributed it to iron. In 1898, however, it was found to have a slightly different wavelength than the iron line. Because no known element generated the required line, it was attributed to a new element called coronium.

At that time, it was assumed that the temperature of the Sun and its corona gradually reduced from the center moving outwards. But in the early part of the 20th century, competing theories were put forward, one for a low-temperature corona and another for a high-temperature one. In 1934, Walter Grotrian analyzed the coronal spectrum and concluded that the temperature was an astonishing 350,000 K. A few years later Bengt Edlén, in a seminal paper, showed that coronal lines are produced by highly ionized iron, calcium, and nickel at a temperature of at least 2 million K. The "coronium" line, in particular, was the product of highly ionized iron. How the temperature of the corona could be so high, when the photosphere temperature is only of the order of 6,000 K, was a mystery, which has not been completely resolved even today.

Charles Young discovered in 1894 that, at very high dispersions, many absorption lines in sunspot spectra appeared to have a sharp bright line in their centers. In 1908, George Ellery Hale and Walter Adams found that photographs of the Sun taken in the light of the 656.3-nm hydrogen line showed patterns that looked like iron filings in a magnetic field. This caused Hale to examine sunspot spectra in detail. He found that the Young effect was actually caused by Zeeman splitting of spectral lines in a magnetic field, which was of the order of 3,000 gauss. So sunspots were the home of very high magnetic fields.

Hale then started to examine the polarities of sunspots, and found that spots generally occur in pairs, with the polarity of the lead spot, as they crossed the disc, being different in the two hemispheres. This pattern was well established by 1912 when the polarities were found to be reversed at the solar minimum. They reversed yet again at the next solar minimum in 1923. So the solar cycle was really 22 years, not 11.

Walter Maunder found in 1913 that large magnetic storms on Earth start about 30 hours after a large sunspot crosses the center of the solar disc. Later work showed that the most intense storms were often associated with solar flares. In 1927, Chree and Stagg found that smaller storms, which did not seem to be associated with sunspots, tended to recur at the Sun's synodic period of 27 days. Julius Bartels called the invisible source on the Sun of these smaller storms, M regions. Both the so-called flare storms and the M storms were assumed to be caused by particles ejected from the Sun. In 1951, Ludwig Biermann suggested that, to explain the behavior of cometary ion tails, there must be a continuous stream of charged particles emitted by the Sun. Then in 1957, Eugene Parker proposed his theory of the solar wind, which was later confirmed by early spacecraft.

Marconi noticed in 1927 that interference with radio signals in September and October of that year coincided with the appearance of large sunspots and intense aurorae. In the late 1930s, Howard Dellinger carried out a detailed examination of the timing of shortwave radio fadeouts, at numerous receiving stations, and solar flares. He found a reasonable but by no means perfect correlation. The fadeouts seemed to start almost instantaneously after the flare was seen, and they only occurred when the receiving station was in daylight. So Dellinger concluded that they were caused by some form of electromagnetic radiation from the Sun, rather than particles.

7.2 Mercury

The synchronous rotation period of Mercury was gradually accepted as a fact in the 20th century. But in 1962, W. E. Howard found that Mercury's dark side seemed to be warmer than it should be if it were permanently in shadow. Then 3 years later, Dyce and Pettengill found, using radar, that Mercury's rotation period was not synchronous, but represented two-thirds of its orbital rotation period.

7.3 Venus

There was considerable confusion in the first half of the 20th century about Venus' rotation period. All sorts of periods were proposed between about 24 hours and synchronous (225 days). Then in 1957 Charles Boyer found a distinctive V-shaped pattern of Venus' clouds that had a 4-day period. In 1962, however, Carpenter and Goldstein deduced a period of about 250 days retrograde using radar, which was modified to 243 days in 1965 for the rotation period of Venus' surface. So Venus has a 243-day period, whilst its clouds have a period of about 4 days, both periods being retrograde.

In 1932, Adams and Dunham concluded that there was no oxygen or water vapor on Venus, but carbon dioxide was clearly present. A few years later, Rupert Wildt calculated that the greenhouse heating of the latter could produce a surface temperature as high as 400 K. Then in 1956, Mayer, McCullough, and Sloanaker deduced a surface temperature of about 600 K by analyzing Venus' thermal radio emissions. The suggestion that Venus' surface temperature could be so high was naturally treated with caution. Shortly afterward, Carl Sagan estimated that the surface atmospheric pressure was an equally incredible 100 bar.

7.4 The Moon

The idea that there may be life on the Moon had fascinated people for centuries. Even respected astronomers like William Herschel had thought that there would be "lunarians" as he called them. But by the start of the 20th century, it was thought that the most complex lifeforms would be some sort of plant life. However, by the 1960s, when the Americans were planning their lunar landings, even this concept had been rejected. Nevertheless, it was thought that there may be some sort of very elemental life, like bacteria, on the Moon.

Bernard Lyot had concluded in 1929, from polarization measurements, that the Moon was probably covered by volcanic ash. Then in the 1950s, Thomas Gold suggested that the Moon may be covered with dust up to a few meters deep. If this was so, it would have provided a major problem for the manned *Apollo* missions.

At the end of the 19th century, the key objection to the impact theory for the formation of lunar craters had been that the craters were generally circular, when they should have been elliptical, because most of the impacts would not be vertical. However, after the First World War it was realized that the shape of the lunar craters resembled shell craters. The shell craters were formed by the shock wave of the impact or explosion, so a nonvertical impact could still produce a circular crater. Nevertheless, not all lunar craters have the same general appearance. So, by the start of the space age it was still unclear if they had been produced by volcanic action, meteorite impact, or both.

7.5 The Earth

It was known in the 19th century that temperatures in deep mines on Earth increased with depth. That, together with the existence of volcanoes, clearly indicated that the Earth has a molten interior. Calculations indicated that the rocks would be molten at a depth of only about 40 km.

In 1897, Emil Wiechert suggested that the Earth has a dense metallic core, mostly of iron, surrounded by a lighter rocky layer, now called the mantle. A little later, Richard Oldham found clear evidence for the existence of the core from earthquake data. Then in 1914, Beno Gutenberg showed that the interface between the mantle and the core, now called the Wiechert–Gutenberg discontinuity, is at about $0.545r$ from the center of the Earth (where r is its radius).

A little earlier, Andrija Mohorovičić had discovered the boundary between the crust and mantle, now called the Mohorovičić discontinuity, by analyzing records of the Croatian earthquake of 1909. The depth of this discontinuity was later found to vary from about 70 km under some mountains to only about 5 km under the deep oceans.

A number of theories were proposed to try to define and explain the internal structure of the Earth. In particular, Harold Jeffreys produced a theory that assumed that all the **terrestrial planets** and the Moon have a core of liquid metals, mostly iron, and a silicate mantle. But it could not explain how those planets with the smallest cores could have retained a higher percentage of lighter material in their mantles. In 1948, William Ramsey solved this problem when he proposed that the whole of the interior of the terrestrial planets consists of silicates, with the internal pressure in the largest planets causing the silicates near the center to become metallic. Unfortunately this idea became unviable when Eugene Rabe found in 1950 that Mercury's density was much higher than originally thought. It was even higher than that of Venus and Mars, which were much larger planets.

In the mid-20th century, most astronomers believed that the planets had been hot when first formed from the solar nebula, but in 1949 Harold Urey suggested that the nebula had been cold. According to Urey, the Earth had been heating up since it was formed because of radioactive decay. Internal convection had then started as iron had gradually settled into the core. Urey believed that the Moon was homogenous because it was relatively small.

At the turn of the 19th century, it was thought that radio waves generally traveled in a straight line. So it was a great surprise when Marconi showed in 1901 that radio waves could be successfully transmitted across the Atlantic. Refraction could have caused them to bend to a limited degree, but not enough to cross the ocean. In the following year, Heaviside and Kennelly independently suggested that the waves were being reflected off an electrically conducting layer in the upper atmosphere.

The structure of what we now call the E or Heaviside layer, and of other layers in the ionosphere, was gradually clarified over the next 20 years or so. The 80 km high D layer was found to largely disappear at night, and the higher E layer was found to maintain its reflectivity for only 4 or 5 hours after sunset. In addition, it was found that solar flares can cause a major disruption to the ionosphere (see Section 7.1). However, it was not until after the Second World War that the cause of these effects could be examined in detail by first sounding rockets and then by spacecraft. The first major discovery was made by Herbert Friedman in 1949 when he showed that the Sun emits X-rays, which have a major effect on the Earth's ionosphere.

7.6 Mars

There was a great deal of uncertainty about the surface of Mars in the first half of the 20th century. It was thought unlikely that the linear markings called *canali* really existed, but they were still recorded from time to time by respected observers. In addition, some astronomers thought that the bluish green areas on Mars were vegetation, while others thought that they were volcanic lava.

There was also considerable uncertainty about the spectroscopic observations of Mars. Some observers recognized water vapor and oxygen lines, whereas others found none. But in 1947 Gerard Kuiper clearly found evidence for a small amount of carbon dioxide, and in 1963 Andouin Dollfus found a trace amount of water vapor. Estimates of the surface atmospheric pressure varied from about 25 to 120 millibars. Then in 1963, shortly before the first spacecraft reached Mars, a figure of 25 ± 15 millibars was estimated by Kaplan, Münch, and Spinrad.

It seemed clear that the yellow clouds seen on Mars were dust. In 1909, Fournier and Antoniadi found that they appeared to cover the whole planet for a while. Later Antoniadi found that they tended to occur around perihelion when the solar heating is greatest, and so appeared to be produced by thermally generated winds. Thirty years later, De Vaucouleurs measured the wind velocities as being typically in the range of 60 to 90 km/h when the clouds first formed.

7.7 Internal Structures of the Giant Planets

It was known in the 19th century that the densities of Jupiter, Saturn, Uranus, and Neptune were similar to that of the Sun, and were much less than that of the terrestrial planets. At that time, it was thought that Jupiter, and probably Saturn, had not yet fully cooled down since their formation. As a result, they were probably emitting more energy than they received from the Sun.

In 1923, Donald Menzel found that the cloud top temperatures of Jupiter and Saturn were about 160 K. This compares with temperatures of 120 and 90 K for Jupiter and Saturn, respectively, that would be maintained solely by incident solar radiation. Three years later, Menzel produced modified observed temperatures of 140, 120, and 100 K, for Jupiter, Saturn, and Uranus. So any internally generated heat would be rather low.

In 1923, Harold Jeffreys pointed out that the ratio of the densities of Io and Europa, the innermost of Jupiter's large satellites, to that of Jupiter, was about the same as the ratio of the density of Titan, Saturn's largest satellite, to that of Saturn. He then assumed that the density of the cores of Jupiter and Saturn were the same as these their large satellites. In that case, the thickness of the planetary atmospheres would be about 20% of their radii.

In the following year, Jeffreys included consideration of the moments of inertia of Jupiter and Saturn in his analysis and concluded that their atmospheres would have depths of $0.09R_J$ and $0.23R_S$, respectively (where R_J and R_S are the radii of Jupiter and Saturn, respectively). He assumed that beneath their atmospheres there was a layer of ice and solid carbon dioxide, which in turn was surrounded a rocky core.

Various schemes were then produced by a number of physicists, of which those of Rupert Wildt in 1938 and William Ramsey in 1951 were probably the most significant. Wildt, who was particularly interested in internal pressures, wanted to find out if matter at the core of the large planets was **degenerate**. His calculations indicated that it was not. Ramsey, on the other hand, developed his theory assuming that the giant planets were made of hydrogen. He then added helium and other ingredients until their densities and moments of inertia were correct. On this basis, he concluded that Jupiter and Saturn were composed of 76% and 62% hydrogen, by mass, respectively, with central pressures of 32 and 6×10^6 bar. At these pressures, most of the hydrogen would be metallic.

The structures of Uranus and Neptune were a problem in Ramsey's analysis because the heavier planet, Neptune, was the smaller. So their constituents could not be the same. Then in 1961 William Porter produced a model that seemed to fit; in this model, Neptune had 74% ammonia and 26% heavier elements, whereas Uranus had less heavy elements and a small amount of hydrogen.

7.8 Atmospheres of the Giant Planets

Vesto Slipher undertook a detailed investigation of the spectra of Jupiter, Saturn, Uranus, and Neptune in the early decades of the 20th century. He recorded numerous bands for all the planets but had trouble interpreting them. In 1932, Rupert Wildt deduced that a number of the bands in all four planets were due to ammonia and methane. However, subsequent work by Mecke, Dunham, Adel, and Slipher showed that some of the lines had been misattributed, so there was no ammonia in the atmospheres of Uranus and Neptune. This was, presumably, because it had been frozen out at their lower temperatures. Adel and Slipher also concluded that the methane concentration reduced in going from Neptune to Uranus to Saturn to Jupiter.

7.9 Jupiter

In 1955, Burke and Franklin made the unexpected discovery that Jupiter was emitting radio waves at 22.2 MHz. Subsequently, it was found that Jupiter emitted energy at many radio frequencies. Some of it was thermal energy, with an effective temperature of 145 K, but some was clearly non-thermal. The latter was taken to indicate that Jupiter had an intense magnetic field, with radiation belts similar to those that had, by then, been found around the Earth.

Our knowledge of Jupiter's Galilean satellites changed little in the 20th century before the space age. In 1900, Bernard had observed that the poles of Io appeared to be reddish in color. Then in 1914 Paul Guthnik showed that all four Galilean satellites exhibited synchronous rotation. In the 19th century, it was thought that all four satellites probably had atmospheres, but this was considered more and more unlikely as the 20th century progressed.

7.10 Saturn

A prominent white equatorial spot had been observed on Saturn in 1876. Then in 1903 Edward Barnard discovered another temporary prominent white spot at about 36°N, but its rotation period around Saturn was some 25 minutes slower. Another equatorial spot that had a similar period to the 1876 equatorial spot appeared in 1933, and another spot that had a similar period to the 1903 spot was observed at about 60°N in 1960. The velocities of these spots showed that there was an equatorial current on Saturn, similar to that on Jupiter. But the one on Saturn had a velocity of about 1400 km/h, compared with just 400 km/h for Jupiter. It was unclear why Saturn, which is farther from the Sun, and so receives less heat than Jupiter, should have a much faster equatorial current.

Markings on Saturn's rings were seen by a number of observers in the late 19th and early 20th centuries, including the respected observers Etienne Trouvelot and Eugène Antoniadi. In 1955, Guido Ruggieri noticed clear radial streaks at both ansae of the A ring, but after further investigation he concluded that they were an optical illusion. It is unclear whether any of these observations were early observations of spokes, of the sort discovered by the *Voyager* spacecraft on the B ring, or not.

In the winter of 1943–1944, Gerard Kuiper photographed the spectrum of the ten largest satellites of the solar system and found evidence for an atmosphere on Titan and possibly Triton. He could find no such evidence for the Galilean satellites of Jupiter, however.

7.11 Uranus and Neptune

In the 19th century, Triton had been found to orbit Neptune in a retrograde sense, and it was unclear at the time whether Neptune's spin was also retrograde. But in 1928 Moore and Menzel found, by observing the **Doppler shift** of its spectral lines, that Neptune's spin was **direct** or **prograde**. So Neptune's largest satellite was orbiting the planet in the opposite sense to the planet's spin. This phenomenon had not been observed before in the solar system for a major satellite.

Kuiper discovered Uranus' fifth satellite, now called Miranda, in 1948. It was orbiting the planet in an approximately circular orbit inside that of the other four satellites. Then in the following year he discovered Neptune's second satellite, now called Nereid, orbiting Neptune in the opposite sense to Triton. Nereid was in a highly elliptical orbit well outside the orbit of Triton. So Nereid was the "normal" satellite in orbiting Neptune direct or prograde, whereas the larger Triton, which was nearer to Neptune in an almost circular orbit, appeared to be the abnormal one.

7.12 The Discovery of Pluto

The discoveries of Uranus and Neptune made astronomers realize that there may well be planets even farther out from the Sun. As Neptune had only been discovered in 1846, and as it was moving very slowly, its orbit was not very well known in the second half of the 19th century. However astronomers had much better information on Uranus' orbit, and so they reexamined it to see if there were any unexplained deviations that might indicate the whereabouts of a new planet. Such deviations were soon found, and a number of possible locations for the new planet proposed by various astronomers, including Percival Lowell. A photographic search for the new planet was started at Lowell's observatory, but this was abandoned when Lowell died in 1916.

In 1929, Vesto Slipher, the new director of Lowell's observatory, recruited Clyde Tombaugh to undertake a search for the new planet using a photographic refractor that had been specifically purchased for the task. Tombaugh photographed the whole of the zodiac, and used a blink comparator to find objects that had moved over time. The task was very tedious, but he discovered Pluto in February 1930 after working for 10 months. However, although the planet's orbit was very similar to that predicted by Lowell (Fig. 8), it was far too small to have perturbed Uranus in the way that Lowell had estimated.

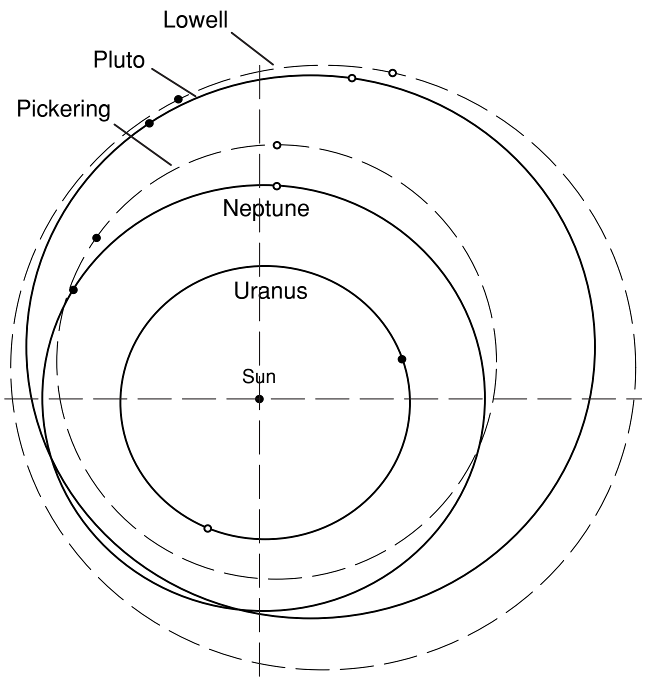


FIGURE 8 A comparison between the true orbit of Pluto and that predicted by Lowell and Pickering. Although Lowell's orbit was reasonably close to that of Pluto, the agreement was fortuitous. (The open circles show the positions of the planets in 1900, and the closed circles represent those in 1930.)

Over the years, the estimated mass of Pluto has gradually reduced from 6.6 M_E (M_E is the mass of the Earth) predicted by Lowell, to 0.7 M_E (maximum) at the time of its discovery, to 0.002 M_E now. Its orbit is highly eccentric, and it has the largest inclination of the traditional planets.

In 1955, Walker and Hardie deduced a rotation period of 6d 9h 17min from regular fluctuations in Pluto's intensity. Little more was known about the planet when the space age started.

7.13 Asteroids

In 1918, Kiyotsugu Hirayama identified families of asteroids based on their orbital radius, eccentricity, and inclination. Initially, he identified three families, Themis (22 members), Eos (21 members), and Koronis (13 members). Hirayama suggested that the three families were each the remnants of a larger asteroid that had fractured. This resurrected, in modified form, the theories of Thomas Wright and Wilhelm Olbers, in the 18th and 19th centuries. They both believed that there had been a planet between the orbits of Mars and Jupiter that had broken up.

In the 19th century, Eros had been discovered with a perihelion of 1.13 AU. In 1932, another asteroid, now called Amor, was found that had an orbit that came even closer to that of the Earth than Eros. Then, just 6 weeks later, the first asteroid, now called Apollo, whose orbit crossed that of the Earth, was discovered. The names of Amor and Apollo have now been given to families of asteroids with similar orbital characteristics.

7.14 Comets

Huggins had shown in the 19th century that there were hydrocarbon compounds in the heads of comets, but he was not able to specify exactly which hydrocarbons were involved. Molecular carbon, C_2, was first identified in the head of a comet just after the turn of the century, and by the mid-1950s C_3, CH, CN, OH, NH, and NH_2, had been found in the heads of comets.

Molecular bands were observed in the tail of Daniel's comet by Deslandres, Bernard, and Evershed in 1907 and in the tail of Morehouse's comet by Deslandres and Bernard the following year. These bands were later identified by Alfred Fowler as those of ionized carbon monoxide, (CO^+) and $N_2{}^+$. Later $CO_2{}^+$ as also found in the tail of a comet.

In the 1930s, Karl Wurm observed that many of the molecules found in comets were chemically very active, and so they cannot have been present there for very long. He suggested, instead, that they had come from the more stable so-called parent molecules $(CN)_2$, H_2O, and CH_4 (methane). In 1948, Pol Swings, in his study of Encke's comet, concluded that the parent molecules were water, methane, ammonia (NH_3), nitrogen, carbon monoxide and carbon dioxide, all of which had been in the form of ice before being heated by the Sun.

In 1950 and 1951, Fred Whipple proposed his icy-conglomerate model (better known as his dirty snowball theory) in which the nucleus is composed of ices, such as methane, with meteoric material embedded within it. Unfortunately, some of the parent molecules were highly volatile. But in 1952 Delsemme and Swings suggested that these highly volatile elements would be able to resist solar heating better if they were trapped within the crystalline structure of water ice, in what are known as clathrate hydrates.

It was difficult to determine the orbits of long-period comets because they were only observed for the fraction of their orbit when they were close to the Sun. However, a survey of about 400 cometary orbits observed up to 1910 showed that only a tiny minority appeared to be hyperbolic. Strömgren and Fayet then showed that none of these comets had hyperbolic orbits before they passed Saturn or Jupiter on their approach to the Sun. So the long-period comets appeared to be members of the solar system.

In 1932, Ernst Öpik concluded, from an analysis of stellar perturbations, that comets could remain bound to the Sun at distances of up to 10^6 AU. Some years later, Adrianus Van Woerkom showed that there must be a continuous source of new, near-parabolic comets to explain the relative numbers observed. Then in 1950 Jan Oort showed that the orbits of 10 comets, with near parabolic orbits, had an average **aphelion** distance of about 100,000 AU. As a result, he suggested that all long-period comets originate in what is now called the Oort cloud about 50,000 to 150,000 AU from the Sun.

7.15 The Origin of the Solar System

In the early decades of the 20th century, theories of the origin of the solar system generally focused on the effect of collisions, and close encounters of another star to the Sun. But all the theories were found to have significant problems, so Laplace's theory of a condensing nebula was reconsidered.

Laplace's theory had been rejected in the 19th century because the original solar nebula did not appear to have had enough angular momentum. However, in the 1930s, McCrea showed that this would not be a problem if the original nebula had been turbulent.

In 1943, Carl von Weizsäcker produced a theory where cells of circulating convection currents, or vortices, formed in the solar nebula after the Sun had condensed. These vortices produced planetesimals that grew to form planets by accretion. Unfortunately, as Chandrasekhar and Kuiper showed, the vortices would not be stable enough to allow condensation to take place. Kuiper then produced his own theory, as did Safronov and others, with the common theme of planetesimals merging to form planets, but none was fully satisfactory.

Bibliography

Hoskin, M., ed. (1997). "Cambridge Illustrated History of Astronomy." Cambridge Univ. Press, Cambridge, England.

Hufbauer, K. (1993). "Exploring the Sun; Solar Science Since Galileo." Johns Hopkins Univ. Press, Baltimore.

Koestler, A. (1990). "The Sleepwalkers: A History of Man's Changing Vision of the Universe." Penguin Books.

Leverington, D. (2003). "Babylon to Voyager and Beyond: A History of Planetary Astronomy." Cambridge Univ. Press, Cambridge, England.

North, J. (1995). "The Norton History of Astronomy and Cosmology." Norton, New York.

Pannekoek, A. (1961). "A History of Astronomy." Interscience, New York (Dover reprint 1989).

Taton, R., and Wilson, C., eds. (1989). "The General History of Astronomy, Vol. 2, Planetary Astronomy from the Renaissance to the Rise of Astrophysics: Part A, Tycho Brahe to Newton." Cambridge Univ. Press, Cambridge, England.

Taton, R. and Wilson, C., eds. (1995). "The General History of Astronomy, Vol. 2, Planetary Astronomy from the Renaissance to the Rise of Astrophysics: Part B, The Eighteenth and Nineteenth Centuries." Cambridge Univ. Press, Cambridge, England.