What do we know about the origin of our Moon?

The search for the origin of the Moon has continued unabated for almost a number of decades. During this time, many excellent researchers have proposed various hypotheses and theories about the origin of the Moon, even though today they seem to be more or less flawed to a certain extent with the launch of NASA's Apollo manned missions to the Moon and the bringing of samples of lunar rocks back to Earth between 1969 and 1972. At the same time, other aspects of the Moon also attracted a great deal of interest. This can be seen in the fact that its name was the same as the moons of the other planets until Galileo discovered the moons of the other planets, especially finding four moons orbiting Jupiter in 1610, or in the Apollo exploration program, it was discovered that the Moon in fact has a thin and weak atmosphere, but not enough to protect it from solar radiation and meteorite impacts, or perhaps in the latest, in October 2020, when NASA's Stratospheric Infrared Observatory confirmed the first time the presence of water on the sunlit surface of the moon, which indicates that water may be distributed across the lunar surface, and not limited to the shadowed area. In this paper, the discussion will focus on the origin of the Moon, and how they are explained by three different theories and one of the most representative, including Fission theory, Capture theory, Co-accretion theory, and Giant-impact hypothesis. Meanwhile, the flaws of each will also be discussed.

The theory that the Moon originated from the fission of the Earth was first proposed by Darwin in 1879 and explained in detail by Wise in 1963. According to Darwin's conjecture on the origin of the Moon, it can be learned that the Moon was formed due to the rapid rotation rate and unstable rotation state of the Earth during the formation of the mantle, which led to the separation of the Moon from the Earth due to the unstable nature of this generation. (Ahrens 07 October, 2021). Based on Darwin's view, Wyss refined it in 1963 and proposed that the formation of the Earth's core could be interpreted as a rapidly rotating Earth in which the increasing centrifugal force induced by the increasing rate of rotation of the Earth would free it from the Earth's gravitational pull near the equator. (Ahrens 07 October, 2021). It is also added that the Earth is in a low rotational state at the beginning of the process, so that the shape of the Earth is initially an oblate spheroid, and that as time passes, internal subsidence, coalescence, compression or collapse reduces the moment of inertia and thus leads to the conservation of angular momentum and an increase in the rate of rotation. (Ahrens 07 October, 2021). Subsequently, as the rate of rotation increases, the shape of the Earth flattens out and forms an unstable rotational state since the gradually ellipticized shape allows the Earth to start rotating around its elliptical short axis as the centre. (Ahrens 07 October, 2021). Finally, it is suggested that if there is sufficient rotational acceleration, the Earth's shape will continue to flatten and eventually form a shape similar to a pear. (Ahrens 07 October, 2021). At this point in time, it is believed that the stem of this pear-like shape is the Moon, and that a sufficiently rapid rate of rotation caused this part of the Earth to separate and eventually reshape itself under the influence of each other's gravitational pull to form the present Earth-Moon system. (Ahrens 07 October, 2021). According to Darwin and Wyss, even though much of the evidence today has grown against this view since the beginning of the Apollo explorations, there is some evidence prior to the Apollo explorations that does support its validity. One such piece of evidence that supports the idea of fission is the discovery by British geologist and clergyman Osmund Fisher in 1882 of a huge hole in the Pacific Basin and a gap in the Atlantic Basin that he attributed to the breakup of the Earth as a result of the separation of the Moon from the Earth. (Cummings 26 September, 2019). This is because it explains why the American continent bears a rough resemblance to the Old World. (Cummings 26 September, 2019). An even more significant support comes from the fact that the Moon has an average density of 3.34 g/cm3, which is very close to the density of the Earth's upper layers of material, which is measured at about 3.32 g/cm3 (Cummings 26 September, 2019). At the same time, isotopes found on the surface of the Moon that are similar to those found on the Earth continue to support the idea that the Moon was separated from the Earth by rotation. (Cummings 26 September, 2019). In addition, from the discovery of isotopes similar to those found in the discovery that the far side of the Moon contains a lighter remnant of primordial Earth material and that the near side is composed of mantle material. (Cummings 26 September, 2019). However, the theory of rotational fission was later challenged by Harold Jeffries. This was because, after using his own calculations on the two-site cylindrical model, he found that the theory only worked if the model of the Earth was not homogeneous. Therefore, the Earth's central core must then be considered. And when the central core of the Earth was taken into account as a factor in the desirability of the theory, he considered the possibility that the tidal flow of the outer shell through the central core might create friction that would impose an upper limit on the rate of rotation of the Earth, and in his estimation of this friction he finally concluded that it was sufficiently strong to invalidate the theory.

Compared with the fission theory, the capture theory proposed by chemist Harold Ulay is also a conjecture about the origin of the moon. Based on what he believed to be a conjecture, he described the formation of the Moon as the following events through his suggestion that the Moon was formed primarily from the accumulation of material that had undergone heating and fractionation of heavy and light elements. (Cummings 26 September, 2019). First, due to the rotation of the sun, it has a chance to create a nebula when it contracts. The nebula gradually cools as the sun is blocked by the surrounding gas and dust, and some of the gas is separated from the nebula in the process. (Cummings 26 September, 2019). Afterwards, the solid materials and silicate dust that collide with each other due to the gradual cooling state will be carried into space together. He believed that the moon was formed in such a way, but the difference was that he believed that during the formation process, the moon escaped the fragmentation and accumulation processes that could form planets, and collected materials similar to that of the earth. (Cummings 26 September, 2019). The material was then captured in its orbit. (Cummings 26 September, 2019). Therefore, from his point of view, the composition of the moon is between the composition of the earth and meteorites, and his theory is based on the assumption that the earth and the moon were formed near the solar accretion disk at different times and in different regions. In fact, his hypothesis can also be accepted by Hannes Olaf Gosta Alfvin of Sweden and Viktor Sergeevich Savronov, a Russian astronomer and planetary scientist. (Cummings 26 September, 2019). This is because Hannes Olaf Gosta Alvervin discovered in 1963 that the moon has a lower density than Mercury, Venus, Earth and other high-density planets. (Cummings 26 September, 2019). So, he speculated that the nebula that formed the moon was different from the nebula that formed Mercury, Venus and the earth, and used a diagram to divide the moon into one group and other planets as a basis for supporting Urey's view. (Cummings 26 September, 2019). Meanwhile, he pointed out that because Triton is moving in a retrograde orbit around Neptune, he took advantage of Triton to point out the possibility that the moon may be captured by the earth. (Cummings 26 September, 2019). However, one issue must be considered in terms of the constitutionality of the capture theory, which is whether the Moon's kinetic energy was lost as it encountered the Earth and was successfully captured by the Earth. For this question, an explanation was proposed by a German school teacher named Horst Gerstenkorn. In his explanation, the capture of the moon by the Earth can be explained by the fact that the moon had a high kinetic energy at the beginning of its formation and travelled around the Earth in a retrograde trajectory. (Cummings 26 September, 2019). Due to the retrograde trajectory, the tidal friction formed by the universal gravitational force between the Earth and the Moon to continuously consume its kinetic energy and be captured by the Earth at a distance of 26 Earth radii. (Cummings 26 September, 2019). At the same time, as the kinetic energy of the moon was being lost during this process, its angular momentum decreases with decreasing orbital velocity and, according to the law of conservation of angular momentum, this decrease in the Moon's angular momentum causes the Earth's angular momentum to increase in this process. Thus, it was argued that the high rate of rotation of the Earth during the initial period of the Moon's capture would have made it 4.6 hours after one day, and that normous tidal friction due to the dissipation of the tidal friction would have caused the Earth to be heated in such a way that no forms of life have existed on the Earth during the period of the Moon's capture. (Cummings 26 September, 2019). This exemplifies the importance of the first law of thermodynamics, which states that energy cannot be destroyed or built upon and can only be transformed from one form to another. In this case, kinetic energy was converted into thermal energy, and demonstrated as an indication of how entropy increases as the thermal energy increases during this process according to the second law of thermodynamics. It also explains that tidal forces increase as the distance between the Moon and the Earth decreases because they satisfy the force of gravity so that their distance exhibits a power law relationship with the magnitude of the force. Similarly, one the variants of his explanation has found that, based on the law of gravity, when the Moon is captured at the minimum distance closest to the Earth would have been less than the Roche's limit (Cummings 26 September, 2019), which further implies that the Moon once lost half of its original mass at the Roche limit based on what Hannes Alfvén believes in his point of view.

In the mid-1960s, the theory that the moon's origin was thought to be co-formed with the earth came into being, called the Co-accretion theory. This theory is based on the idea that a slowly rotating cloud of dust and gas first formed the Sun, then the planets after flattening into a disk called an accretion disk, and later separated into pairs of two different interpretations of this theory. (Cummings 26 September, 2019). One is based on Schmidt's thought of how the planets formed in the solar system. (Cummings 26 September, 2019). Although he believed in the early days that there were planetary embryos in the circumsolar clouds, and that these embryos gradually became planets, in his later research with his colleagues L. E. Gurevich and A. I. Lebedinsky concluded that whether planetary embryos really exist or not, the evolution of planets Collisions and agglomeration between particles must occur during the process, and this process is described as first, the relative speed of the particles is reduced due to collision, which increases the collision frequency and density. (Cummings 26 September, 2019). Later, when the density of the particles reaches a certain critical level, , because the system cannot maintain its original state and gradually gathers together under the action of gravity, finally these aggregates collide again. (Cummings 26 September, 2019). However, as the volume increases, in most cases once these aggregates reach a certain volume, they begin to retain peripheral fragments in their gravitational field under the action of their own gravity, and at the same time make these fragments become a particle swarm if not absorbed by the aggregate. (Cummings 26 September, 2019). Eventually, these particle groups will continuously accumulate to form an independent individual and evolve into satellites when these particle groups form an almost circular orbit. (Cummings 26 September, 2019). This developmental process is a particular example of gravity's ability to turn a disordered system into an orderly system through the emergence of patterns. This is also similar to the idea that gravity contributes to a homogenous system that can be transformed from a homogenous system into an inhomogeneous system around the time of the Big Bang explosion, as mentioned in the book Deep Simplicity because each of these examples demonstrates how gravity breaks up the chaotic state of the system and involves to operate in a regular form of the present-day motion of the moon around the Earth, or as an analogy, satellite around its planet. However, even though his thoughts were among the strongest at the time in explaining the origin of the Moon, there were still objections to the theory from pro-capture scientist Harold Urey. One of his objections is that the density of the moon is much smaller than that of the earth because the iron content may be much smaller than that of the earth. This goes against the fact that the moon and the earth were formed together because if the Earth and Moon accumulated material from the same region of the primordial solar nebula, their internal structures should be similar. (Cummings 26 September, 2019). This can be demonstrated by the fact that the moon's values are particularly significantly lower than those of the Earth's in a piece of paper published in 1966 by Ted Linwood and William Mason Kaula, where Kaula specifically gave estimates of the iron-nickel content of the Moon Earth and the other terrestrial planets as 65.0% on Mercury, 26.5% on Venus, 31.5% on Earth, 6.0% on the Moon, and 19.0% on Mars. (Cummings 26 September, 2019). The second interpretation of this theory is a variant of the first one and the moon's lower density than the Earth and differences in chemical composition were also taken into account. The idea was developed by Alfred Edward "Ted" Ringwood of the Australian National University and suggests that the early Earth's atmospheric temperatures exceeded 1,500 degrees Celsius, causing some of the silicate in the Earth's metallic iron to fractionate and sublimate into the atmosphere and escape from the Earth. (Cummings 26 September, 2019). Later, as these gases expanded and cooled, they condensed to form fragments that gathered around the Earth and formed a ring of sediments. (Cummings 26 September, 2019). Finally, the outflowing gases and the condensed sediments fly off into the distance under the influence of mutual adhesion to form the moon. (Cummings 26 September, 2019). Through his explanation, it can be realised that this process is similar to air convection. For example, air convection demonstrates the process by which a substance changes from solid to a gas as the temperature increases and then changes back as solid as it cools. The difference, however, is that convection does not figure prominently in his explanation of the formation of the moon, as it does not explicitly mention whether convection could have formed in the interior of the early Earth when it was first formed.

Relative to the nature of the other three theories that had objections at the time, the giant impact hypothesis remains the dominant idea to explain the origin of the moon until now. This hypothesis can be traced back to the start of the Apollo missions. From the samples mined from the moon and returned to the earth, it can be known that there is indeed no metallic iron core or only a trace amount of iron core on the moon, and it was found that the moon has been completely covered by a deep magma ocean. (Genda 01 January, 2018). Shortly after analysing these samples, Hartmann and Davis proposed in 1975 the idea that the Moon originated from an oblique impact that ejected material into Earth's orbit. (Genda 01 January, 2018). At first, although this idea was not convinced by the public, since some people thought that such catastrophic events and extreme results seemed unlikely to occur, 10 years later, after the "Moon Origin Conference" held in Kona, Hawaii in 1984, the big impact hypothesis thus became the dominant hypothesis. (Genda 01 January, 2018). This is since it was discovered through dynamic simulation that this theory naturally explains the lack of metallic iron in the moon because when an asteroid hits the Earth, which accounts for about 10% of the Earth's mass, the ejected material is able to form a "protolunar disk" around the Earth, and the phenomenon of huge amounts of energy being released during large impacts supports the conclusion that the moon has a magma ocean after analysing samples obtained from Apollo's mining of the moon. (Genda 01 January, 2018). At the same time, it can also be found that the proportion of stable isotopes in the moon is almost the same as that of the earth, especially after comparing the ratio between Cr-53 and Cr-52 in the Apollo samples with the Earth, the average of the Cr-53 value for the Earth was found to be 0.013 ± 0.014 ‰, which was close to the 0.0116± 0.015 ‰ analysed from the collected samples. (Cole 14 September, 2017). However, a 2007 paper by Stevenson and his then-colleague Kaveh Pahlevan struck down the possibility of this theory. They argued that the asteroid that hit the Earth would necessarily be just a micro-planet, a fraction of the Earth's mass, and so would be formed from material in a relatively narrow band around the planet's orbit. (Clery 11 October, 2013. But, an Earth-sized planet would have cleared material from a much wider disk across the orbits of Mars and Mercury, where planet-forming material has very different isotopic ratios. (Clery 11 October, 2013. This idea threw studies of the Moon's origin into a panic even though Stevenson and Pahlevan later patched up the scenario that isotopes could be produced on collision and pointed out that the heat from a huge impact would have produced an Earth and a debris disk made of melted and evaporated magma, and that later, over a cooling time of 100 to 1,000 years, enough turbulent mixing and Earth-formation could have occurred to bring them into equilibrium between the disk and the Earth, resulting in a similar material of Earth and Moon. (Clery 11 October, 2013. This is because very quickly, some people became sceptical of this scenario, particularly Melosh, who argued that in order to achieve a homogeneous composition, the Earth and the disk would have had to exchange so much material that the disk would have collapsed during the process. (Clery 11 October, 2013. Even more so, Matija Ćuk and Sarah Stewart of Harvard University have proposed that the impacting asteroid would have been much smaller than thought, and would have been able to produce a homogeneous Earth and Moon through a head-on collision even at high rotational speeds, but only if the Earth-Moon system after the impact had too much angular momentum and no longer needed the Sun's evection resonance with the Moon's orbit around the Earth to gradually reduce its angular momentum in order to become today‘s state. (Clery 11 October, 2013). According to this view, Canup argues that even if a disk with the correct composition were produced, it would still need the evection resonance to slow it down to account for its evolution after impact, but then there is the view of Jack Wisdom of the Massachusetts Institute of Technology, Cambridge, that when the Moon is in evection resonance, its orbit becomes more elongated, and that such a change produces extreme tidal forces that heat the Moon up and change its physical properties enough to end the resonance before the Moon has time to exhaust enough angular momentum from the system (Clery 11 October, 2013), so again it appears that this resonance is less important to the impact theory in explaining the origin of the Moon. Overall, the idea of an impact theory to explain the origin of the Moon is changeable because, over time, different researchers have been able to find flaws in their current explanations and disprove them or revise them based on different interpretations. This highlights the fact that any theory in science can be disproved by new ideas, and that this process is ongoing. This is because at the conference where Stevenson presented his ideas in 2007, many researchers argued that the reason that explaining the origin of the Moon through the impact theory was complicated was that there was no model that can explain the process between the impact of the Earth and the formation of the Moon when it comes to the mixing and platooning of turbulence that occurs during this process. Even if there is a way to get out of the situation, testing it is not easy. This can be demonstrated by the assumption that if it is found to be wrong there are significant variations in isotope ratios in the solar system, then the reasoning that there would be similarities in isotopes between the Earth and the Moon can arise naturally. (Clery 11 October, 2013). But again, the modelling about Mars cannot hold this assumption to be true because the characteristic that Mars possesses a mass of only 10% of the Earth’s has led to the assumption that it could have been formed elsewhere completely different in the solar system, but moved to where it is today. (Clery 11 October, 2013). Hence, this could also be demonstrated that all scientific hypothesis are not rigorous if there is not sufficient and strong evidence to temporarily eliminate possible refutations, and scientific modelling is even more impossible because it requires specific operational definitions or quantitative forms to describe how a scientific system works, and in this case, the specific process of the Earth's impact to form the Moon has not been implemented with enough simplicity.

What can be gleaned from these four different theories and hypotheses is that the origin of the moon has not been conclusively established until now. Even though the Giant-Impact hypothesis is the most credible one since it explains the discovery of similar isotopes between the Moon and the Earth after the Apollo exploration program, the insufficient scientific modelling to explain the process of the Moon's formation eventually puts this hypothesis in doubt. In addition, the other three theories are also controversial due to their respective flaws, such as the fission theory does not take into account that the fraction caused by tidal forces, the capture theory does not consider the similar isotopes between the Earth and the Moon, as well as why the Moon has lower iron density than the Earth about the Co-accretion theory, if it holds true at all. Therefore, in the future, the astronomical community may need more new ideas and evidence to continue research on this topic because from these four known theories and hypotheses, they more or less have the ability to compensate for each other's flaws to a greater or lesser extent. After that, perhaps a new theory will be able to better explain the origin of the Moon if the new ideas and evidence are collaborated well. The origin of the moon continues to fascinate us to this day, and it will likely remain a source of wonder and inspiration for generations to come.

References

Ahrens, Catlin. 07 October, 2021. “Moon: Origin, Alternative Theories.” Edited by Brian Cudnik. In *Encyclopedia of Lunar Science*, 1-4. Houston, USA: Springer Link. https://doi.org/10.1007/978-3-319-05546-6.

Clery, Daniel. 11 October, 2013. “Impact Theory Gets Whacked.” In *Science*, 183-185. Vol. 342. Washington, D.C., United States: American Association for the Advancement of Science. 10.1126/science.342.6155.183.

Cole, Devon B. 14 September, 2017. “Chromium Isotopes.” In *Encyclopedia of Geochemistry*, 1-6. Ithaca, United States: Springer Link. https://doi.org/10.1007/978-3-319-39193-9.

Cummings, Warren D. 26 September, 2019. “Pre-Apollo Theories About the Origin of the Moon.” In *Evolving Theories on the Origin of the Moon*, 51-75. 1st ed. USRA, Columbia: Springer Link. https://doi.org/10.1007/978-3-030-29119-8.

Genda, Hidenori. 01 January, 2018. “Giant Impact Hypothesis.” In *Encyclopedia of Geochemistry*, 617-620. 1st ed. Cornell University, Ithaca: Springer Link. https://doi.org/10.1007/978-3-319-39312-4.