On the origin of the long-period comets: competing theories

Sebastian Buhai
Utrecht University College
email: sbuhai@ucu.uu.nl

Abstract. More than ever the problem of the origin of the long-period comets seems to materialize in both inciting and exciting debates among astronomers and physicists. For more than 50 years scientists have been trying to solve the issue; nevertheless all the present acknowledged theories have their shortcomings and eventually do not provide a clear answer. Several theories accounting for the formation of the Oort cloud comets in the Uranus-Neptune planetary region with variations on the theme, as well as several competing theories designing interstellar molecular clouds as the correct origin of the long-period comet cloud, have been advanced. The present paper tries to present an overview and to critically discuss the main ideas together with their principal inconsistencies in an accessible and concise format.

1. Historical background: Newton, Halley, long and short-period comets

In an attempt to shed light on the existing knowledge of the origin of cometary objects, a short overview of the main historical track leading to capital developments in the cited domain is undoubtfully appropriate.

Leaving aside the early observation of the comets and their wrongful labeling as clouds of luminous gas high in the Earth's atmosphere, the first focus on the comets as such (let us call it "in a true scientific manner") was done in the 17th century by Isaac Newton. Using his then-new law of gravitation next to Kepler's existing laws of planetary motion, Newton pointed out that the Sun-grazing comet of 1680 had been moving through space along an apparent parabolic path, therefore not bound to the Sun, but extending far into interstellar space. Not long afterwards Edmond Halley corrected Newton, arguing that comets move in very long ellipses rather then parabolas, and thus they are bound to the Sun. Halley's theory can hardly be denied its revolutionary feature provided that his conclusion was based solely on very crude observations of comets which he compiled in the very first catalogue of 24 cometary orbits.

We notice that progress did not wait too long once the start has been enacted; the boost given by Newton's and Halley's discoveries was probably determinant. As early as the second half of the 18th century, based on the increasing empirical data (discovering and tracking more and more comets), astronomers could divide the comets into two groups: long-period comets and short-period comets. The main distinctive feature consisted in the time required for comets to complete an orbit around the Sun: short-period orbits were the orbits requiring less than 200 years, while the long-period orbits were the comets that needed more than 200 years to complete their orbits. Another important difference resided in the long-period comets tending to enter the inner solar system randomly from all directions, while the short-period comets have usually orbits inclined no more than 40 degrees to the ecliptic plane. An immediate consequence is that long-period comets might move retrograde around the Sun, while the short-period comets are moving prograde as our Earth does. In particular interest appeared concerning the origin and the cause of each of these types of comets. We will further concentrate on the long-period comets and will try to investigate and pronounce ourselves with regard to their origin.

Surprisingly enough, despite (or should we rather say "due to") the fast technological development and the substantial newly acquired data, confusion spread among scientists when trying to assess the origin of the comets. Concretely, in the beginning of the 19th century it became obvious that the periods of many long-period comets were not just long, but exceedingly wrong. It appeared that about one-third of the long-period comets were coming into the planetary region on trajectories not bound gravitationally to the solar system; this time they were thought to be hyperbolas. We can consider us fortunate as this time the astronomers realized quite fast that the hyperbolic trajectories were the result of the perturbations from the giant planets coupled with the very weak bounding of these comets to the Sun. Moreover, all the calculations were corrected for a reference to the center of mass of the entire solar system and not only to the Sun, thus obtaining so-called baryocentric rather than heliocentric orbits. Had these corrections been performed, all the initially hyperbolic orbits became elliptical and thus the way towards new discoveries explaining the origin and cause of the long-period comets was cleared again.

2. Setting the scene: Oort's contribution

It was not before mid-20th century that astronomers had their first "acceptable" theories accounting for the origin of the long-period comets. Adianus van Woerkom opened the stage by showing that the apparently broad and flat distribution of cometary orbital periods could be explained entirely by planetary perturbations, which tend to scatter the comets in a random manner to both larger and smaller orbits. Nevertheless a dilemma remained with reference to the pileup of comets at near-zero energy. It was here that the astronomer Jan Oort came in and recognized that exactly this spike had to represent the source of the long-period orbits. He imagined this source as a vast cloud of objects lying far beyond the planets and extending to the edge of the Sun's gravitational influence. Perturbations of cometary orbits by passing stars or molecular clouds are responsible for scattering comets into the inner solar system. As these Oort cloud comets enter the planetary system for the first time, their courses are sometimes considerably altered by the planet's influence. Those that gain orbital energy are shot out of the solar system becoming interstellar wanderers. On the other hand, comets that lose energy become more tightly bound to the Sun and thus fall among the flat distribution of orbital energies calculated by van Woerkom. Briefly, in the standard setting[1], a comet scattering off the giant planets returns repeatedly to the inner solar system until either something catastrophic (that could be ejection or collision with a planet for instance) happens or the orbit receives a large enough external perturbation that it no longer enters the inner solar system. In this sense, Oort's picture of long-period cometary cloud seemed highly reasonable; nevertheless some assumptions made by the astronomer as well as some characteristic values that he calculated proved to be wrong in later analysis.

At his time, Oort calculated that the long-period comets have orbits that extend out to between 100,000 and 300,000 AU from the Sun. Nowadays, having so much many accurate orbits for long-period orbits, we know that the average orbital aphelion of a "new" long-period comet is about 44, 000 AU. The explanation for this smaller distance is that in addition to perturbing random passing stars the Oort cloud is also perturbed by gravitational tides generated by stars both in the Milky way's disk and in the galactic core (the latter to a considerable lesser extent). It consequently turned out that the tide due to he galactic disk is stronger than the sole perturbations taken into account by Oort from random passing stars. The result is that comets with orbital aphelia beyond 200, 000 AU are easily lost to interstellar space; thus the Oort cloud extends only until approximately 200, 000 AU. Moreover the gravitational effects of the mass in the nucleus and in the disk of our galaxy lead to an ellipsoidal shape of the tenuous boundary of the Oort cloud. Hence a theoretical image of the Oort cloud exists even if the cometary cloud as such has never been seen.

---

[1] For our purpose a "standard setting" will mean a setting where we do not allow for planetary migration. If we include planetary migration computations as well, the model becomes much more complex (see Failed Oort clouds and planetary migration, by B. Hansen, astro-ph/0004058, 5 April 2000)

Another wrong assumption made by Oort consisted in the always “gentle” character of the phenomena within the Oort cloud. In his words the cometary cloud looked as “ a garden, gently raked by stellar perturbations”. In reality extremely violent processes might occur either by perturbing very close passing stars or by giant molecular clouds in the galaxy (the so-called GMC’s). It is not within our purpose in this paper to describe these particular violent phenomena although if they were to be considered on their scientific merits, there is no way they could be left out. We will instead focus on the consequences of these phenomena for the topic of the present paper. What is extremely relevant for the topic is that due to these GMC encounters as well as to other several phenomena, the population of the Oort cloud is assessed to have increased to about 6 trillion comets. Naturally the question was where had all these comets come from? And once the question was asked, many theories accounting for the origin of the Oort cloud were born.

3. Hypotheses for the origin of the Oort cloud- pros and cons

Pessimists might easily argue that the origin of the Oort cloud remains largely an unanswered question as all existing theories have their shortcomings. Nevertheless there is no doubt that many things are much more clear now and even if no theory can completely account for the origin of the Oort cloud, it is extremely interesting to consider and analyze the most challenging and competing ideas.

3.1. Theories accounting for creation of comets within planet formation

Generally speaking we can divide the bulge of theories trying to explain the origin of the long-period comets in theories of comet creation within planet formation and theories of comet creation within star formation or even earlier. Of course these are not clear-cut subdivisions, theories accounting for instance for interstellar origin and later capturing by our solar system being also advanced (however, as we shall see, these do not have a real scientific value). We begin our investigation by presenting the main theories that account for the comet creation during planet formation.

To give history what belongs to history, Oort speculated already in 1950 that the comets must have been ejected from the asteroid belt by the giant planets during the formation of the solar system. This was probably one of the shortest-living theories accounting for the origin of the long-period comets as Fred Whipple proved in the same year that comets were “dirty snowballs” and thus they should have been formed much further from the Sun, in locations cold enough for water to condense.

The most primordial theory at present assumes that the Oort cloud comets probably came from the Uranus-Neptune zone. Having smaller masses than the massive Jupiter or Saturn, Uranus and Neptune could not easily throw so many comets onto escape trajectories and consequently a larger fraction of the comets in their zones ended up in the Oort cloud. Doubt has been cast on this scenario, as the Uranus-Neptune region cannot account for sufficient source of energy to eject the cometary nuclei into the Oort cloud. Weissman estimated that the initial Oort cloud cometary population must have exceeded at least 80 M_{Earth} . In 1994 Bailey came with an estimation of the present cometary mass of at least 380 M_{Earth} with an upper limit of survival probability of 20%. That obviously means that the ejected mass of cometary population was at least several times higher than the sum of masses of Uranus and Neptune (less than 32 M_{Earth}), which are regarded as the main ejecting planets. So a contradiction appeared. The reason why this theory is still regarded as the primordial theory (despite its mentioned shortcomings) is that late dynamical studies suggested that also objects from the asteroid belt have been ejected in the Oort cloud (the old theory of Oort revived) and surprisingly their number is not at all negligible. Computations by H. Levinson led to an almost 2% ejected asteroids from the inner-planet region of the total Oort cloud population. No valid explanation exists nowadays for this relative huge number of asteroids in the Oort cloud. Hence it is not yet very clear whether computations are misleading, some additional element is missing or the theory cannot survive.

A slightly different theory from the primordial one, attempting to solve the dilemma of the energy needed to deliver the cometary nuclei into the Oort cloud, was proposed by Fernandez. He assumed that a considerable fraction of the nuclei was ejected into the cloud not only by Uranus and Neptune, but also by Jupiter and Saturn. However, this idea failed as well as the efficiency of the latter planets to place the nuclei into the cloud was assessed to be too small (the computed ratios of mass ejected to the comet cloud and interstellar space respectively were found to be only 0.03 for Jupiter and 0.16 for Saturn) and thus still insufficient to account for the huge Oort cloud population. Moreover, another observation came against Fernandez's idea with Bergin demonstrating that cometary nuclei were created in a cooler environment than the Jupiter-Saturn zone of protoplanetary discs: the initial temperature of the two most recent long-period comets, Hyakutake and Hale-Bopp was found to be between 25 to 30K and respectively 40K. Naturally we cannot generalize on an absolute scale but observational evidence is only supporting so far the latter view.

Next to mention, there are several theories ascribing interstellar origins for the Oort cloud and this comets being captured All this class of theories is presently heavily criticized as not being able to provide any reasonable mechanism of capture of interstellar comets by the solar system (one of the main critics in this sense is Torbett). Furthermore, even if we ignore by absurd this first problem, the interstellar origin theories fail to show any scientific merit in accounting for the high-density core of the Oort cloud. It is estimated that the density would have been much smaller if the comets had been captured by our solar system from the interstellar space and would not necessarily exhibit the present pattern (high-density core, extremely tenuous borderline).

3.2. Theories accounting for creation of comets in molecular interstellar clouds, before planetary formation

It has to be stated first of all that a certain weakness of all this class of theories cannot be disregarded: so far no specific physical mechanism of creation of cometary nuclei in an interstellar-cloud environment, no matter how dense, has been worked out in detail. However late empirical discoveries have revealed a certain number of unexpected phenomena[2] in the interstellar clouds and hence it is considered that it is only a matter of time until this weakness will be combated with solid arguments.

We begin with a theory that is not exactly belonging to the class of theories discussed under the above-mentioned heading, but it is a very influential theory in this context. One of the extremely interesting theories, the theory of creation in situ, was published by Hills in as early as 1982. Although a relatively old theory, it does not have so many or so important shortcomings as one would imagine. Practically Hills suggested that "pressure exercised by radiation from the Sun and from neighboring protostars may have forced the coagulation into comets of dust grains in collapsing layers of the protosun at distances from 1 to 5 thousand AU." If so far nothing can be brought against this reasoning, thinking about the implications of this theory, if such efficiency in comet creation is actually present, then the comets would have also further coagulated. Moreover, this coagulation would have happened at an even higher rate, in molecular interstellar clouds during the passages of the extremely luminous stars through the dense cloud regions. Not perfect, this theory still raises considerable interest. Variations of the same main idea, aiming to solve the present controversies, are equally important. A follow-up of Hills's idea in a new theory will be described in more details below.

---

[2] One of the most interesting example is the discovery of the evaporating gaseous globules in the M16 nebula with the Hubble Space Telescope. Such globules are overdense regions appropriate for a condensation of macroscopic bodies. For more details see Hestler et al. ,1996, AJ 111, 2349

Among the most recent and challenging views concerning the formation of the Oort cloud, the conception of the Oort cloud as a remnant of the protosolar nebula seems to raise particular and increasing interest. L. Neslusan is the author of several papers and articles describing very accurate terms this claim. There are many indications (some touched by presenting the above hypotheses) of similarities among cometary material and that of cool, dense interstellar clouds. On the other hand, it might be regarded as proven that the comets have been bounded to the solar system over all its existence (empirical evidence). If we combine these facts, the most appropriate birthplace for comets seems to be the molecular interstellar clouds. To be extremely critical, it does not follow directly, but in no other theory do we find exclusively deductive ways of reasoning. Once accepted this first assumption (see the weakness mentioned in the beginning of this section), then it is obvious that the cometary nuclei had to take part in the collapse of protostelar clouds. Neslusan proved that it is more than likely that the comets would have remained at Oort cloud distances from the center of the cloud and thus the comets in the Oort cloud may represent a remnant of the nebular stage of the solar system. Obviously this theory is based on assumptions that cannot really be proved (but counter-proof is also inexistent). First of all, it is assumed that the cometary nuclei were sufficiently large so as to move according to mechanic laws and not to obey the laws of hydrodynamics as the dust and the gas molecules. If this assumption is not so hard to accept as macroscopic bodies are subjected to mechanics rather then hydrodynamics, a second assumption is added, assuming that the change in gravitational potential during the collapse was not so strong as to lead to the collapse of cometary nuclei into the protoplanetary disc. Hence, what is implied is that the comets from the Oort cloud represent a remnant of protosolar nebula from a stage before its collapse into the protosun and protoplanetary disc. The entire demonstration is based on showing that a significant number of cometary nuclei remained somewhere at large distances from the center at the end of the collapse, if the peak of the presumed Maxwell-Boltzman distribution of the initial velocities of the nuclei was not lower than 100 m/s (an upper limit of a few hundred m/s also exists, but an excess of this limit was deemed improbable so it is not at all discussed). Unfortunately it is impossible to make a reliable quantitative estimate of this peak of velocity distribution. Nowadays it is aimed to prove the theory with the help of analysis of its consequences. To what extent this is going to happen in the near future, we cannot tell.

4. Conclusions

If we were to draw any conclusions we could only assess the steady if not increasing interest in the origin of the long-period comets. Albeit supporters of the first class of theories or the latter one, the scientists are looking for further proofs for their theories. Observational evidence is increasing but not yet to the extent that a final decision could be reached. Until then we will have to satisfy ourselves with what we have been offered: far-reaching but still incomplete competing theories aiming to explain the origin of the Oort cloud.

References

J.Beatty, C. Petersen, A.Chaiking, The New Solar System, 4th edition, Cambridge University Press, 1999

L. Neslusan, The Oort cloud as a remnant of the protosolar nebula, Astron. Astrophys. 361, 369-378 (2000)