Newton's Lunar Theory

by Michael Nauenberg

Nicholas Kollerstrom, *Newton's Forgotten Lunar Theory: His Contribution to the Quest for Longitude*, (Green Lion Press 2000)

In 1687 Newton revolutionized the age old kinematical approach to planetary and lunar motion by demonstrating in his *Principia* that this motion can be determined from a mathematical theory based on gravitational forces. Some years later he published a small booklet entitled *Theory of the Moons Motion (TMM)* [1] where he gave rules for finding the position of the Moon without, however, including any justification for these rules based on his gravitational theory. In his book, *Newton's Forgotten Lunar Theory*, Kollerstrom deals primarily with this less well known work which presumably explains the word "forgotten" which appears in the title. A major topic of this book is an assesment of the numerical accuracy of Newton's lunar theory, which has been the subject of disagreements in the past. For example, the 17-th century Astronomer Royal John Flamsteed, who provided Newton with the best data on lunar positions, and Edmond Halley, who was also familiar with Newton's lunar theory, held opposite views on this subject, and other views have been expressed up to the present time. By implementing a computer program which incorporates Newton's rules in *TMM*, and comparing the results with those obtained from the modern lunar theory Kollerstrom has resolved some of these conflicts by showing the extent to which these rules work well in practice. But his analysis of Newton's work misses the vital connection between the Greek geometrical- kinematical tradition in astronomy which was applied to the moon's motion until Newton's time, and Newton's new dynamical approach based on his theory of gravitation. This has led Kollerstrom to a seriously flawed evaluation of the origin of Newton's lunar theory.

The anomalies or inequalities in the orbit of the moon around the earth have been a major challenge to astronomers since Antiquity. Already by the second century B.C., Hipparchus pointed out that the moon's position at quadrature deviated in longitude by over two and a half degrees from the predictions of the Greek model of epicyclic motion, although this model accounted for its position both at conjunction and opposition from the sun. Subsequently Ptolemy proposed the first mechanism for this anomaly [2], which later became known as the evection, but this mechanism also predicted a near doubling of the apparent size of the Moon which is not observed. Nevertheless, Ptolemy's lunar model was not challenged until the 15-th century when the Arab astronomer Ibn-al-Shatir developed an alternative mechanism [3][4] which was adopted later by Copernicus [5]. Their model accounted for the evection without introducing the large unobserved variations of the lunar size in Ptolemy's model. In the 17-th century alternative models where developed by Tycho Brahe and by Johannes Kepler [6] who incorporated his empirical law of areas for planetary motion into his lunar model. Finally, in 1640 a young astronomer, Jeremiah Horrocks [6], refined Kepler's model further predicting correctly the inequalities in the distance of the moon from the earth which was being determined at that time by a micrometer. By setting the center of the moons elliptic orbit on an epicycle, Horrock's model gave rise to an oscillating eccentricity and line of apses. Moreover, for the angle of rotation of this center, Horrocks adopted Kepler's choice of twice the angle between the earth-sun axis and the mean line of apses of the orbit of the moon. Since Greek times, the corresponding angle chosen to describe Ptolemy's evection inequality was twice the elongation of the moon from the sun, but it is the Kepler-Horrocks angle which turns out to be correct as shown by Newton's in his dynamical formulation of this inequality [7].

In the *Principia* Newton demonstrated that the regular motion of the Moon around the Earth was caused by the gravitational attraction between these two bodies, but that the above mentioned inequalities in this motion were due to perturbations by the gravitational force of the Sun. Detailed aspects of the calculations which involve the notoriously difficult three-body problem, however, remained a major challenge to astronomers after Newton's pioneering work. In the preface to the first edition of the *Principia* Newton remarked that " after I began to work on the inequalities of the motion of the moon, and then also began to explore other aspects of the laws and measure of gravity... I thought that publication should be put off to another time, so that I might investigate these other things and publish all my results together... ". A younger contemporary of Newton, John Machin, reported that he had told him that " his head never ached but with his study on the moon" [8].

Over the past centuries Newton's lunar work has been received with immense admiration by those who have been able to understand the profound mathematical innovations in his theory. An early reviewer of the second edition of the *Principia* stated that " the computation made of the lunar motions from their own causes, by using the theory of gravity, the phenomena being in accord, proves the divine force of intellect and the outstanding sagacity of the discoverer"[9]. The French astronomer Laplace asserted that the sections of the *Principia* dealing with the motion of the moon are " one of the most profound parts of this admirable work", and the British physicist and Astronomer Royal George Airy regarded it "as the most valuable chapter that has ever been written on physical science" [9]

In his book, Kollerstrom dissents with these views claiming instead that "Newton could not get his gravity theory to work and had to reach to an earlier kinematical theory. Later mythologizing about Newton led to forgetting of what he had really constructed". As if to underscore this thesis, the cover of his book displays a model of lunar motion which is not due to Newton, as one would expect, but it is an illustration of Horrock's lunar theory by William Crabtree [10]. Kollerstrom concludes that " an unbridgeable gulf seemed to loom between the dynamics of an emerging gravity theory and the kinematic series of circular motions required by a lunar theory that would work" (p. 228). But what is the basis for this profound change in the historical evaluation of Newton's lunar work? Kollerstrom's book indicates that this originated from an influential article written by D.T. Whiteside in 1976 who concluded that Newton's lunar theory " was a retrogresive step back to an earlier kinematic tradition which he had hoped once to transcend, and to a limited Horrocksian model which was not even his own invention" [11]. Subsequently this view was echoed by other Newtonian scholars, e.g. R.S. Westfall, who wrote in his biography of Newton, *Never at Rest*, that the most important of the corrections which Newton made in the second edition of the *Principia* was " a kinematic theory of the motion of the center of the moon's orbit which had no foundation in gravitational dynamics" [12]. These views, however, are due to a lack of understanding of the close relation that exists between the *kinematical* model of Horrocks, and the *dynamical* lunar theory of Newton [13]- [15]. Ironically, this relation was already understood in the 18-th century. For example, the great mathematician Leonhard Euler asserted that Horrocks model had been the inspiration for his well known method of variation of parameters - the basis for the modern formulation of Newton's lunar perturbation theory.

According to Newton's theory, in the absence of the solar perturbation, the orbit of the Moon would be an ellipse with the Earth at one of the foci satisfying the area law or what is now called conservation of angular momentum. It should be remembered, however, that such elliptical motion had been proposed earlier by Kepler [6], who followed the *kinematical* tradition of the ancient Greeks. By fitting Tycho Brahe's observational data, Kepler found that an ellipse gave the best results for the orbit of the planet Mars, and naturally he assumed that such an orbital curve applied also to other planets. One of Newton's important achievements in the *Principia* was his demonstration that this orbital motion was a consequence of a gravitational force which varies inversely as the square of the distance between the Sun and the planets. Likewise, by a veritable tour de force, Newton also demonstrated that Horrock's model was a consequence of the perturbing effects due to the gravitational force of the sun. In the *Principia* Newton presented most of his arguments in a qualitative manner in some of the 22 corollaries following Proposition 66. Some of the mathematical methods underlying his arguments can be found, although in a rather succint form, in Cor. 3 and 4 following Proposition 17, in Book 1. Fortunately, details of his perturbation theory appear in a manuscript in the Portsmouth collection of Newton papers which is has been published by Whiteside [16]. This manuscript shows that Newton's method anticipated the perturbation methods of Euler and Laplace [14] [15]

In his book, however, Kollerstrom does not mention any of these important corollaries in the *Principia* or acknowledges the existence of the Portsmouth manuscript which has been shown to be essential to an understanding of Newton's lunar work, and its relation to the model of Horrocks. Therefore, it is not surprising that having missed these connections, Kollerstrom is then led to commit some serious errors in his discussion of novel aspects of Newton's lunar theory. As indicated previously, some of Newton's explanations for the rules in *TMM* had been presented in the first edition of the *Principia*(1687). Subsequently in the second (1713) and third (1726) editions he added a new Scholium to Proposition 35 of Book 3 where some the contents of his 1702 booklet are included, and the gravitational basis of his rules are described. But Kollerstrom states that, on the contrary, " *TMM*, written by the Master of the Mint, surveyed the periods and inequalities of lunar motion and described a kinematical model basically that of Horrocks. It thus represents a diametric antithesis to the *Principia'* endeavor of 1687. The latter was a work of theory, of zero practical utility as far as lunar predictions are concerned. The former contains no theory ... it is as if the hope expressed in early 1695 had been extinguished, in that no theory was present, and its author had regressed to a kinematical approach, with the old epicycles and deferent still there"(p. 32). Actually, the lunar theory in the *Principia* and the rules for lunar tables presented in *TMM* are intimately related. Although Newton's lunar method, which we now characterize as a theory to lowest order in the perturbing force, was inadequate to obtain the *magnitude* of some of the lunar inequalities, the *periods* of these inequalities were not simply "surveyed", as alleged by Kollerstrom, but calculated from first principles.

Newton was evidently aware of shortcomings of his theory, and for the purpose of providing a useful source for tables of lunar positions, he judiciously adjusted amplitudes to obtain agreement with the excellent lunar data provided to him by Flamsteed. Unfortunately, he admitted only occasionally to this procedure causing thereby confusion to his readers. For example, in discussing the amplitude of the semiannual inequality of the line of apsides in the Scholium to Prop. 35, Newton stated that " it comes to about 12^0 $18'$ as nearly as I could determine from the phenomena". It can be readily verified, however, that Newton's approximate perturbation theory gives only about 8^0, which leads to a disagreement with the observed magnitude of the evection well known since Greek times. The Porstmouth manuscript shows that Newton was able to do this calculation, but evidently he suppressed this result. In fact, the amplitude of other inequalites, which Newton claimed to have obtained "by the theory of gravity" were also adjusted to fit observation, e.g. the annual inequality which Newton claimed to have found " by the theory of gravity" to be $11'$ $49''$, which is close to the modern value [17], is actually found to be $13'$ according to his lowest order theory [14] [15]. But these adjustments made by Newton to obtain better agreement with observations do not justify the claim in the opening pages of Kollerstrom's book that Newton " was faced with abject failure..."(p. 27), an incorrect assessment which is repeated throughout the entire book. Newton was aware of the difficulties in carrying out higher order calculations of his lunar equations, and it took another century of arduous labor by mathematical astronomers of the caliber of Clairut, Lagrange, Euler, Laplace and Hill before these calculations were carried out successfully.

In addition to deriving the lunar inequalities which were known from observations, Newton deduced four new inequalities from his gravitational theory which he included in *TMM* although these had not been recorded previously. One of these inequalities, referred by Kollerstrom as the "sixth equation", gives us also further evidence that Newton derived from his dynamical theory the model introduced by Horrocks. In Newton's approximate theory, the radius of Horrocks' epicycle for the center of the lunar orbit is proportional to the average of the *difference* between the solar gravitational force acting on the moon and on the earth. Hence this radius depends inversely on the cube of the distance between the earth and the sun, and because of the eccentricity of the earth's orbit around the sun, this leads to an annual variation of this radius. For example, in the summer when the earth is farther from the sun, the radius of the epicycle decreases relative to its value during the winter. Newton took into account this effect by adding an epicyle to the Horrocksian model, which he described in detail in the Scholium after Proposition 35, Book 3. It is this epicycle which give rise to Newton's sixth equation. Kollerstrom reports that the inclusion of this equation gives an improvement after a wrong sign for this effect given in *TMM* is corrected, as Newton did in the second edition *Principia*. But Kollerstrom fails to understand the origin of this equation. Thus, he devotes an entire section (chapter 12, section 9) to compare the results obtained by including both Newton's new epicycle and the sixth equation, without realizing that these two are equivalent. In particular, this *double counting* presented in a graph on page 181 of his book does not make any sense.

Subtle connections which often exist between theoretical ideas before and after the occurrence of a scientific revolution are very important to our understanding of the history and philosophy of science. As mentioned earlier, Ptolemy's model of an eccentric motion for planetary motion which is governed in time by an equant, is mathematically equivalent to the epicycle models of Ibn-al-Shatir and Copernicus introduced later to replace this equant. Likewise, Kepler's elliptical orbit for the planets satisfying the area law is also equivalent to these models to first order in the eccentriciy. These various models could not be distinguished until the accuracy of astronomical observations exceeded about ten minutes of arc. This amount was necessary to observe effects due to quadratic terms in the eccentricity of Mars which was first achieved by Tycho Brahe, and culminated in Kepler's discovery of his three empirical laws. Applying his gravitational theory, Newton then derived Kepler's laws. His metamorphosis of Horrock's model, previously developed within the Greek kinematical tradition, into a dynamical model based on gravitational theory is another remarkable example of deep connections between these two seemingly unrelated traditions in science. In conclusion, it is worthwhile to quote Newton's 1694 reply to Flamsteed who evidently also misunderstood Newton's approach:

> I believe you have a wrong notion of my method in determining the Moons motions. For I have not been about making such corrections as you seem to suppose, but about getting a general notion of all the equations on which her motions depend and considering how afterwards I shall go to work with least labour and most exactness to determine them. For the vulgar way of approaching by degrees is bungling and tedious. The method which I propose to my self is first to get a general notion of the equations to be determined and then by accurate observations to determine them. If I can compass the first part of my designe I do not doubt but to compass the second... And to go about the second work till I am master of the first would be injudicious... [18]

Acknowledgments

I would like to thank J.B. Brackenridge and I.B. Cohen for valuable comments.

References

[1] Isaac Newton, *Theory of the Moon's Motion (1702).* With a bibliographical and historical introduction by I. Bernard Cohen (Dawson, 1975)

[2] *Ptolemy's Almagest.* Translated and annotated by G.J. Toomer with a foreword by Owen Gingerich (Princeton University Press, 1998) pp. 173-216

[3] V. Roberts, *The Solar and Lunar Theory of Ibn ash-Shatir: A Pre-Copernican Copernican Model*, Isis 48, 1957, pp. 428-432

[4] G. Saliba, *A history of Arabic Astronomy* (New York Univ. Press, 1994) pp. 291-305

[5] Nicholas Copernicus, *On the Revolutions.* Translation and Commentary by Edward Rosen (The Johns Hopkins Univ. Press, 1978) pp. 191-195

[6] C. Wilson, *Predictive Astronomy in the Century after Kepler*, in Planetary astronomy from the Renaissance to the rise of astrophysics, Part A, edited by R. Taton and C. Wilson (Cambridge University Press 1989) pp. 161- 206

[7] Isaac Newton, *The Principia* A new translation and guide by I. Bernard Cohen and Anne Whitman (University of California Press, 1999)

[8] R. S. Westfall, *Never at Rest* A biography of Isaac Newton (Cambridge Univ. Press, 1980) p. 541

[9] ref. [1] p.41

[10] ref. [6] p. 200

[11] D.T. Whiteside, *Newton's Lunar Theory: From High Hope to Disenchantment*, Vistas in Astronomy 19, pp.317-328.

[12] ref. [8] p. 547

[13] C. Wilson, *Newton on the Moon's Variation and Apsidal Motion: The Need for a Newer "New Analysis"*, in "Isaac's Newton Natural Philosophy", edited by Jed. Z. Buchwald and I. Bernard Cohen (The M.I.T. Press 2000) pp. 139-188

[14] M. Nauenberg, *Newton's Perturbation Methods for the three-Body Problem and their application to Lunar motion*, in "Isaac's Newton Natural Philosophy", edited by Jed. Z. Buchwald and I. Bernard Cohen (The M.I.T. Press 2000) pp. 189-224

[15] M. Nauenberg, *Newton's Portsmoth Perturbation Theory and its application to Lunar Motion*, in the Foundations of Newtonian scholarship, edited by R.H. Dalitz and M. Nauenberg (World Scientific, 2000) pp. 167-194

[16] *The Mathematical Papers of Isaac Newton*, 1684-1691, edited by D.T. Whiteside (Cambridge University Press, 1974) pp. 508-537.

[17] In his conclusion, Kollerstrom quotes D'Alembert (p. 230) who "doubted whether this derivation of the annual equation was really sound "

> Il en est quelques-unes que M. Newton did avoir calculées par la Theorie de la gravitation, mais san nous apprendre le chemin qu'il a pris pour y parvenir. Telles sont celles de 11′ 49″ qui dépend de l'équation du centre du soleil

[18] ref. [8] p. 545