CALCULATING THE COSMOS

How Mathematics Unveils the Universe

IAN STEWART

PROFILE BOOKS

First published in Great Britain in 2016 by
PROFILE BOOKS LTD
3 Holford Yard
Bevin Way
London
WC1X 9HD
www.profilebooks.com

Copyright © Joat Enterprises, 2016

10 9 8 7 6 5 4 3 2 1

Typeset in Sabon by MacGuru Ltd
Printed and bound in Great Britain by Clays, St Ives plc

The moral right of the author has been asserted.

All rights reserved. Without limiting the rights under copyright reserved above, no part of this publication may be reproduced, stored or introduced into a retrieval system, or transmitted, in any form or by any means (electronic, mechanical, photocopying, recording or otherwise), without the prior written permission of both the copyright owner and the publisher of this book.

A CIP catalogue record for this book is available from the British Library.

ISBN 978 1 78125 431 8
eISBN 978 1 78283 150 1
Export ISBN 978 1 78125 718 0

FSC
www.fsc.org
MIX
Paper from responsible sources
FSC® C018072

Contents

1

Attraction at a Distance

> Macavity, Macavity, there's no one like Macavity,
> He's broken every human law, he breaks the law of gravity.
>
> Thomas Stearns Eliot, *Old Possum's Book of Practical Cats*

WHY DO THINGS FALL down?

Some don't. Macavity, obviously. Along with the Sun, the Moon, and almost everything else 'up there' in the heavens. Though rocks sometimes fall from the sky, as the dinosaurs discovered to their dismay. Down here, if you want to be picky, insects, birds, and bats fly, but they don't stay up indefinitely. Pretty much everything else falls, unless something is holding it up. But up there, nothing holds it up – yet it doesn't fall.

Up there seems very different from down here.

It took a stroke of genius to realise that what makes terrestrial objects fall is the same thing that holds celestial objects up. Newton famously compared a falling apple to the Moon, and realised that the Moon stays up because, unlike the apple, it's also moving *sideways*.[1] Actually, the Moon is perpetually falling, but the Earth's surface falls away from it at the same rate. So the Moon can fall forever, yet go round and round the Earth and never hit it.

The real difference was not that apples fall and Moons don't. It was that apples don't move sideways fast enough to miss the Earth.

Newton was a mathematician (and a physicist, chemist, and mystic), so he did some sums to confirm this radical idea. He calculated the forces that must be acting on the apple and the Moon to make them follow their separate paths. Taking their different masses into account, the forces turned out to be identical. This convinced him that the Earth must be pulling both apple and Moon towards it. It was natural to suppose that the same type of attraction holds for any pair of bodies, terrestrial or celestial. Newton expressed those attractive forces in a mathematical equation, a law of nature.

One remarkable consequence is that not only does the Earth attract the apple: the apple also attracts the Earth. And the Moon, and everything else in the universe. But the apple's effect on the Earth is way too small to measure, unlike the Earth's effect on the apple.

This discovery was a huge triumph, a deep and precise link between mathematics and the natural world. It also had another important implication, easily missed among the mathematical technicalities: despite appearances, 'up there' is in some vital respects the same as 'down here'. The laws are identical. What differs is the context in which they apply.

We call Newton's mysterious force 'gravity'. We can calculate its effects with exquisite accuracy. We still don't understand it.

✦

For a long time, we thought we did. Around 350 BC the Greek philosopher Aristotle gave a simple reason why objects fall down: they are seeking their natural resting place.

To avoid circular reasoning, he also explained what 'natural' meant. He maintained that everything is made from four basic elements: earth, water, air, and fire. The natural resting place of earth and water are at the centre of the universe, which of course coincides with the centre of the Earth. As proof, the Earth doesn't move: we live on it, and would surely notice if it did. Since earth is heavier than water (it sinks, right?) the lowest regions are occupied by earth, a sphere. Next comes a spherical shell of water, then one of air (air is lighter than water: bubbles rise). Above that – but lower than the celestial sphere that carries the Moon – is the realm of fire. All other bodies tend to rise or fall according to the proportions in which these four elements occur.

This theory led Aristotle to argue that the speed of a falling body is proportional to its weight (feathers fall more slowly than stones) and inversely proportional to the density of the surrounding medium (stones fall faster in air than in water). Having reached its natural rest state, the body remains there, moving only when a force is applied.

As theories go, these aren't so bad. In particular, they agree with everyday experience. On my desk, as I write, there is a first edition of the novel *Triplanetary*, quoted in the epigram for Chapter 2. If I leave it alone, it stays where it is. If I apply a force – give it a shove – it moves a few centimetres, slowing down as it does so, and stops.

Aristotle was right.

And so it seemed for nigh on two thousand years. Aristotelian physics, though widely debated, was generally accepted by almost all intellectuals until the end of the sixteenth century. An exception was the Arab scholar al-Hasan ibn al-Haytham (Alhazen), who argued against Aristotle's view on geometric grounds in the eleventh century. But even today, Aristotelian physics matches our intuition more closely than do the ideas of Galileo and Newton that replaced it.

To modern thinking, Aristotle's theory has some big gaps. One is weight. *Why* is a feather lighter than a stone? Another is friction. Suppose I placed my copy of *Triplanetary* on an ice-skating rink and gave it the same push. What would happen? It would go further: a lot further if I rested it on a pair of skates. Friction makes a body move more slowly in a viscous – sticky – medium. In everyday life, friction is everywhere, and that's why Aristotelian physics matches our intuition better than Galilean and Newtonian physics do. Our brains have evolved an internal model of motion with friction built in.

Now we know that a body falls towards the Earth because the planet's gravity pulls it. But what is gravity? Newton thought it was a force, but he didn't explain how the force arose. It just *was*. It acted at a distance without anything in between. He didn't explain how it did that either; it just *did*. Einstein replaced force by the curvature of spacetime, making 'action at a distance' irrelevant, and he wrote down equations for how curvature is affected by a distribution of matter – but he didn't explain *why* curvature behaves like that.

People calculated aspects of the cosmos, such as eclipses, for millennia before anyone realised that gravity existed. But once gravity's role was revealed, our ability to calculate the cosmos became far more powerful. Newton's subtitle for Book 3 of the *Principia*, which described his laws of motion and gravity, was 'Of the System of the World'. It was only a slight exaggeration. The force of gravity, and the manner in which bodies respond to forces, lie at the heart of most cosmic calculations. So before we get to the latest discoveries, such as how ringed planets spit out moons, or how the universe began, we'd better sort out some basic ideas about gravity.

✦

Before the invention of street lighting, the Moon and stars were as familiar, to most people, as rivers, trees, and mountains. As the Sun went down, the stars came out. The Moon marched to its own drummer, sometimes appearing during the day as a pale ghost, but shining much more brightly at night. Yet there were patterns. Anyone observing the Moon even casually for a few months would quickly notice that it follows a regular rhythm, changing shape from a thin crescent to a circular disc and back again every 28 days. It also moves noticeably from one night to the next, tracing a closed, repetitive path across the heavens.

The stars have their own rhythm too. They revolve, once a day, round a fixed point in the sky, as if they're painted on the inside of a slowly spinning bowl. *Genesis* talks of the firmament of Heaven: the Hebrew word translated as 'firmament' means bowl.

Observing the sky for a few months, it also became obvious that five stars, including some of the brightest, don't revolve like the majority of 'fixed' stars. Instead of being attached to the bowl, they crawl slowly across it. The Greeks associated these errant specks of light with Hermes (messenger of the gods), Aphrodite (goddess of love), Ares (god of war), Zeus (king of the gods), and Kronos (god of agriculture). The corresponding Roman deities gave them their current English names: Mercury, Venus, Mars, Jupiter, and Saturn. The Greeks called them *planetes*, 'wanderers', hence the modern name planets, of which we now recognise three more: Earth, Uranus, and Neptune. Their paths were strange, seemingly unpredictable. Some moved relatively quickly, others were slower. Some even looped back on themselves as the months passed.

Most people just accepted the lights for what they were, in the same way that they accepted the existence of rivers, trees, and mountains. But a few asked questions. What are these lights? Why are they there? How and why do they move? Why do some movements show patterns, while others break them?

The Sumerians and Babylonians provided basic observational data. They wrote on clay tablets in a script known as cuneiform – wedge-shaped. Among the Babylonian tablets that archaeologists have found are star catalogues, listing the positions of stars in the sky; they date to about 1200 BC but were probably copies of even earlier Sumerian tablets. The Greek philosophers and geometers who followed their lead were more aware of the need for logic, proof, and theory. They were pattern-seekers; the Pythagorean cult took this attitude to extremes, believing that the entire universe is run by numbers. Today most scientists would agree, but not about the details.

The Greek geometer who had the most influence on the astronomical thinking of later generations was Claudius Ptolemy, an astronomer and geographer. His earliest work is known as the *Almagest*, from an Arabic rendering of its original title, which started out as 'The Mathematical Compilation', morphed into 'The Great Compilation', and then into '*al-majisti*' – the greatest. The *Almagest* presented a fully fledged theory of planetary motion based on what the Greeks considered to be the most perfect of geometric forms, circles and spheres.

The planets do not, in fact, move in circles. This wouldn't have been news to the Babylonians, because it doesn't match their tables. The Greeks went further, asking what would match. Ptolemy's answer was: combinations of circles supported by spheres. The innermost sphere, the 'deferent', is centred on the Earth. The axis of the second sphere, or 'epicycle', is fixed to the sphere just inside it. Each pair of spheres is disconnected from the others. It wasn't a new idea. Two centuries earlier, Aristotle – building on even earlier ideas of the same kind – had proposed a complex system of 55 concentric spheres, with the axis of each sphere fixed to the sphere just inside it. Ptolemy's modification used fewer spheres, and was more accurate, but it was still rather complicated. Both led to the question whether the spheres actually existed, or were just convenient fictions – or whether something entirely different was really going on.

✦

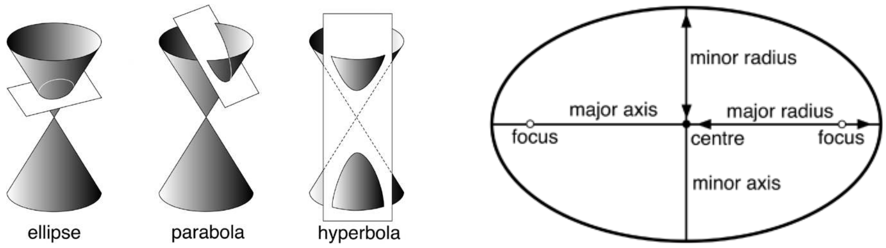
For the next thousand years and more, Europe turned to matters theological and philosophical, basing most of its understanding of the natural world on what Aristotle had said around 350 BC. The universe was believed to be geocentric, with everything revolving around a stationary Earth. The torch of innovation in astronomy and mathematics passed to Arabia, India, and China. With the dawn of the Italian Renaissance, however, the torch passed back to Europe. Subsequently, three giants of science played leading roles in the advance of astronomical knowledge: Galileo, Kepler, and Newton. The supporting cast was huge.

Galileo is famous for his invention of improvements to the telescope, with which he discovered that the Sun has spots, Jupiter has (at least) four moons, Venus has phases like the Moon's, and there's something strange about Saturn – later explained as its ring system. This evidence led him to reject the geocentric theory and embrace Nicolaus Copernicus's rival heliocentric theory, in which the planets and the Earth revolve round the Sun, getting Galileo into trouble with the Church of Rome. But he also made an apparently more modest, but ultimately more important, discovery: a mathematical pattern in the motion of objects such as cannonballs. Down here, a freely moving body either speeds up (when falling) or slows down (when rising) by an amount that is the same over a fixed, *small* period of time. In short, the body's acceleration is constant. Lacking accurate clocks, Galileo observed these effects by rolling balls down gentle inclines.

The next key figure is Kepler. His boss Tycho Brahe had made very accurate measurements of the position of Mars. When Tycho died, Kepler inherited his position as astronomer to Holy Roman Emperor Rudolph II, together with his observations, and set about calculating the true shape of Mars's orbit. After fifty failures, he deduced that the orbit is shaped like an ellipse – an oval, like a squashed circle. The Sun lies at a special point, the focus of the ellipse.

Ellipses were familiar to the ancient Greek geometers, who defined them as plane sections of a cone. Depending on the angle of the plane relative to the cone, these 'conic sections' include circles, ellipses, parabolas, and hyperbolas.

When a planet moves in an ellipse, its distance from the Sun varies. When it comes close to the Sun, it speeds up; when it's more distant, it slows down. It's a bit of a surprise that these effects conspire to create an orbit that has exactly the same shape at both ends. Kepler didn't expect this, and for a long time it persuaded him that an ellipse must be the wrong answer.



Left: Conic sections. Right: Basic features of an ellipse.

The shape and size of an ellipse are determined by two lengths: its major axis, which is the longest line between two points on the ellipse, and its minor axis, which is perpendicular to the major axis. A circle is a special type of ellipse for which these two distances are equal; they then give the diameter of the circle. For astronomical purposes the radius is a more natural measure – the radius of a circular orbit is the planet's distance from the Sun – and the corresponding quantities for an ellipse are called the major radius and minor radius. These are often referred to by the awkward terms semi-major axis and semi-minor axis, because they cut the axes in half. Less intuitive but very important is the eccentricity of the ellipse, which quantifies how long and thin it is. The eccentricity is 0 for a circle and for a fixed major radius it becomes infinitely large as the minor radius tends to zero.[2]

The size and shape of an elliptical orbit can be characterised by two numbers. The usual choice is the major radius and the eccentricity. The minor radius can be found from these. The Earth's orbit has major radius 149·6 million kilometres and eccentricity 0·0167. The minor radius is 149·58 million kilometres, so the orbit is very close to a circle, as the small eccentricity indicates. The plane of the Earth's orbit has a special name: the ecliptic.

The spatial location of any other elliptical orbit about the Sun can be characterised by three more numbers, all angles. One is the inclination of the orbital plane to the ecliptic. The second effectively gives the direction of the major axis in that plane. The third gives the direction of the line at which the two planes meet. Finally, we need to know where the planet is in the orbit, which requires one further angle. So specifying the orbit of the planet and its position within that orbit requires two numbers and four angles – six *orbital elements*. A major goal of early astronomy was to calculate the orbital elements of every planet and asteroid that was discovered. Given these numbers, you can predict its future motion, at least until the combined effects of the other bodies disturb its orbit significantly.

Kepler eventually came up with a set of three elegant mathematical patterns, now called his laws of planetary motion. The first states that the orbit of a planet is an ellipse with the Sun at one focus. The second says that the line from the Sun to the planet sweeps out equal areas in equal periods of time. And the third tells us that the square of the period of revolution is proportional to the cube of the distance.

✦

Newton reformulated Galileo's observations about freely moving bodies as three laws of motion. The first states that bodies continue to move in a straight line at a constant speed unless acted on by a force. The second states that the acceleration of any body, multiplied by its mass, is equal to the force acting on it. The third states that every action produces an equal and opposite reaction. In 1687 he reformulated Kepler's planetary laws as a general rule for how heavenly bodies move – the law of gravity, a mathematical formula for the gravitational force with which any body attracts any other.

Indeed, he *deduced* his force law from Kepler's laws by making one assumption: the Sun exerts an attractive force, always directed towards its centre. On this assumption, Newton proved that the force is inversely proportional to the square of the distance. That's a fancy way to say that, for example, multiplying the mass of either body by three also trebles the force, but multiplying the distance between them by three reduces the force to one ninth of the amount. Newton also proved the converse: this 'inverse square law' of attraction implies Kepler's three laws.

Credit for the law of gravity rightly goes to Newton, but the idea wasn't original with him. Kepler deduced something similar by analogy with light, but thought gravity pushed planets round their orbits. Ismaël Bullialdus disagreed, arguing that the force of gravity must be inversely proportional to the square of the distance. In a lecture to the Royal Society in 1666, Robert Hooke said that that all bodies move in a straight line unless acted on by a force, all bodies attract each other gravitationally, and the force of gravity decreases with distance by a formula that 'I own I have not discovered'. In 1679 he settled on an inverse square law for the attraction, and wrote to Newton about it.[3] So Hooke was distinctly miffed when exactly the same thing appeared in *Principia*, even though Newton credited him, along with Halley and Christopher Wren.

Hooke did accept that only Newton had deduced that closed orbits are elliptical. Newton knew that the inverse square law also permits parabolic and hyperbolic orbits, but these aren't closed curves, so the motion doesn't repeat periodically. Orbits of those kinds also have astronomical applications, mainly to comets.

Newton's law goes beyond Kepler's because of one further feature, a prediction rather than a theorem. Newton realised that since the Earth attracts the Moon, it seems reasonable that the Moon should also attract the Earth. They're like two country dancers, holding hands and whirling round and round. Each dancer feels the force exerted by the other, tugging at their arms. Each dancer is held in place by that force: if they let go, they will spin off across the dance floor. However, the Earth is much more massive than the Moon, so it's like a fat man dancing with a small child. The man seems to spin in place as the child whirls round and round. But look carefully, and you'll see that the fat man is whirling too: his feet go round in a small circle, and the centre about which he rotates is slightly closer to the child than it would have been if he were spinning alone.

This reasoning led Newton to propose that *every* body in the universe attracts every other body. Kepler's laws apply to only two bodies, Sun and planet. Newton's law applies to any system of bodies whatsoever, because it provides both the magnitude and the direction of *all of the forces that occur*. Inserted into the laws of motion, the combination of all these forces determines each body's acceleration, hence velocity, hence position at any moment. The enunciation of a universal law of gravity was an epic moment in the history and development of science, revealing hidden mathematical machinery that keeps the universe ticking.

✦

Newton's laws of motion and gravity triggered a lasting alliance between astronomy and mathematics, leading to much of what we now know about the cosmos. But even when you understand what the laws are, it's not straightforward to apply them to specific problems. The gravitational force, in particular, is 'nonlinear', a technical term whose main implication is that you can't solve the equations of motion using nice formulas. Or nasty ones, for that matter.

Post-Newton, mathematicians got round this obstacle either by working with very artificial (though intriguing) problems, such as three identical masses arranged in an equilateral triangle, or by deriving approximate solutions to more realistic problems. The second approach is more practical, but actually a lot of useful ideas came from the first, artificial though it was.

For a long time, Newton's scientific heirs had to perform their calculations by hand, often a heroic task. An extreme example is Charles-Eugène Delaunay, who in 1846 started to calculate an approximate formula for the motion of the Moon. The feat took over twenty years, and he published his results in two volumes. Each has more than 900 pages, and the second volume consists entirely of the formula. In the late twentieth century his answer was checked using computer algebra (software systems that can manipulate formulas, not just numbers). Only two tiny errors were found, one a consequence of the other. Both have a negligible effect.

The laws of motion and gravity are of a special kind, called differential equations. Such equations specify the rate at which quantities change as time passes. Velocity is the rate of change of position, acceleration the rate of change of velocity. The rate at which a quantity is currently changing lets you project its value into the future. If a car is travelling at ten metres per second, then one second from now it will have moved ten metres. This type of calculation requires the rate of change to be constant, however. If the car is accelerating, then one second from now it will have moved more than ten metres. Differential equations get round this problem by specifying the instantaneous rate of change. In effect, they work with very small intervals of time, so that the rate of change can be considered constant during that time interval. It actually took mathematicians several hundred years to make sense of that idea in full logical rigour, because no finite period of time can be instantaneous unless it's zero, and nothing changes in zero time.

Computers created a methodological revolution. Instead of