Paths of Discovery
Pontifical Academy of Sciences, Acta 18, Vatican City 2006
www.pas.va/content/dam/accademia/pdf/acta18/acta18-rees.pdf

FROM CONTROVERSY TO CONSENSUS IN COSMOLOGY

MARTIN J. REES

1. An Accelerating Universe

Will our universe go on expanding for ever? Will distant galaxies move ever-further away from us? Or could these motions ever reverse, so that the entire firmament eventually recollapses to a 'big crunch'? Until five years ago, the future of our expanding universe was an unsettled question, widely disputed among cosmologists. But progress on several different fronts has allowed us to delineate more confidently what its fate might be.

We used to believe that the answer just depended on the 'competition' between the force of gravity that tends to attract objects to each other, and the expansion energy that tends to move them apart. The Universe will recollapse – gravity will eventually 'defeat' the expansion, unless some other force intervenes – if the density exceeds a definite 'critical' value. This critical density can be computed easily: it corresponds to about 5 atoms in each cubic metre. That does not seem much: but if all the stars in galaxies were dismantled and their material was spread uniformly through space that would amount to only 0.2 atoms per cubic meter – 25 times less than the critical density.

This may seem to imply perpetual expansion, by a wide margin. We have learnt however that the universe contains not just atoms but dark matter – probably some species of particles made in the Big Bang which help to gravitationally bind together stars and galaxies – but even when this is included, the sum seems no more than 0.3 of the critical density. If there were nothing in the universe apart from atoms and dark matter, the expansion would be decelerating, but not enough to ever bring it to a halt.

But observational advances have led to a perplexing twist. A new form of energy – a new force – is latent even in completely empty space. This unexpected force exerts a repulsion that overwhelms the pull of gravity, and causes the cosmic expansion to be speeding up. The evidence came initially from careful study of exploding stars several billion light years away.

Some stars, in their death-throes, explode as supernovae, and for a few days blaze nearly as bright as an entire galaxy containing many billions of ordinary stars. A distinctive type of supernova, technically known as Type 1a, signals a sudden nuclear explosion in the centre of a dying star, when its burnt-out core gets above a particular threshold of mass and goes unstable. It is, in effect, a nuclear bomb with a standard calculable yield. The physics is fairly well understood, but the details need not concern us. What is important is that Type 1a supernovae are pretty close to being 'standard candles' – bright enough to be detected at great distances, and standardised enough in their intrinsic luminosity.

From how bright these supernovae appear, it should be possible to infer reliable distances, and thereby (by measuring the redshift as well) to relate the expansion speed and distance at a past epoch. Cosmologists expected that such measurements would reveal the expected slowdown-rate.

However, the measurements instead provided evidence for an accelerating Universe!

2. Energy Latent in Space: Einstein's Cosmological Constant?

An acceleration in the cosmic expansion implies something remarkable and unexpected about space itself: there must be an extra force that causes a 'cosmic repulsion' even in a vacuum. This force would be indiscernible in the Solar System; nor would it have any significance within our galaxy; but it could overwhelm gravity in the still more rarefied environment of intergalactic space. We normally think of the vacuum as 'nothingness'. But if one removes from a region of interstellar space the few particles that it contains, and even shields it from the radiation passing through it, and cools it to the absolute zero of temperature, the emptiness that is left may still exert some residual force, tending to push things apart.

Einstein himself conjectured this. As early as 1917, soon after he had developed general relativity, he began to think how his theory might apply to the universe. At that time, astronomers only really knew about our own Galaxy – a disc-like swarm of stars; the natural presumption was that the universe was static – neither expanding nor contracting. Einstein found that a universe that was 'set up' static would immediately start to contract, because everything in it attracts everything else. A universe could not persist in a static state unless an extra force counteracted gravity. So he added to his theory a new number which he called the cosmological constant, and denoted by the Greek letter lambda. Einstein's equations then allowed a static universe where, for a suitable value of lambda, a cosmic repulsion exactly balanced gravity. This universe was finite but unbounded: any light beam you transmitted would eventually return and hit the back of your head.

Einstein's reason for inventing lambda has been obsolete for 70 years. But that does not discredit the concept itself. On the contrary, lambda now seems less contrived and 'ad hoc' than Einstein thought it was. Sometimes in science, a new way of looking at a problem 'inverts' it. The question used to be: why should empty space exert a force? Now we ask: why is there not a colossal amount of energy latent in space? Why is the force so small? Empty space, we now realise, is anything but simple. Any particle, together with its antiparticle, can be created by a suitable concentration of energy. On an even tinier scale, empty space may be a seething tangle of strings, manifesting structures in extra dimensions.

All this activity involves energy. Indeed, from this perspective the puzzle is: Why don't all the complicated processes that are going on, even in empty space, have a net effect that is much larger? Why is space not as 'dense' as an atomic nucleus or a neutron star (in which case it would close up on itself within 10-20 kilometres)? Indeed, if lambda represents the energy latent in space, which we realise has intricate structure on subatomic scales, the best theoretical guess is that it should induce a cosmic repulsion 120 powers of ten stronger than is actually claimed!

Most physicists suspected that some process, not yet understood, made the resultant vacuum energy exactly zero, just as other features of our universe – for instance its net electric charge – are believed to be.

But the vacuum energy turns out to be not zero, but it is very, very, small. Why? There clearly has to be some impressive cancellation, but why should this be so precise that it leads to a row of 119 zeros after the decimal point, but not 120 or more? Lambda is remarkably small, but clearly any theory that offers a deep reason why it is exactly zero is wrong too.

A slightly different idea is that the repulsion is due to some all-pervasive field that has negative pressure, and therefore exerts a gravitational repulsion, but which dilutes and decays during the expansion, so that it is by now guaranteed to be very small. This mysterious substance has been dubbed 'quintessence' or dark energy. Yet another possibility, of course, is that Einstein's equations might need modification for some unsuspected reason on the scale of the entire cosmos.

(Just one technical comment. If there is energy in empty space [equivalent, as Einstein taught us, to mass, through his famous equation $E=mc^2$], why does it have the opposite effect on the cosmic expansion from the atoms, the radiation and the dark matter, all of which tend to slow down the expansion? The answer depends on a feature of Einstein's theory that is far from intuitive: gravity, according to the equations of general relativity, depends not just on energy [and mass] but on pressure as well. And a generic feature of the vacuum is that if its energy is positive, then its pressure is negative [in other words, it has a 'tension', like stretched elastic]. The net effect of vacuum energy is then to accelerate the cosmic expansion. It has got a negative pressure and so, according to Einstein's equations, it pushes rather than pulls).

Other independent evidence now supports the case for dark energy. This comes from measurements of the faint microwave radiation that pervades all of space – the afterglow of the hot dense beginning.

Starting out hot and dense, this radiation has cooled and diluted; its wavelength has stretched bringing it into the microwave band. The temperature of this radiation which is lingering on today is only 3 degrees above absolute zero, and it fills the entire Universe. This background radiation is not completely uniform across the sky: there is a slight patchiness in the temperature, caused by the ripples that evolve into galaxies and clusters.

The WMAP spacecraft. is a million miles away, at the Lagrangian point beyond the moon. It is a marvellously sensitive instrument, conceived and designed by a group led by David Wilkinson at Princeton. Wilkinson died in 2002, but fortunately lived long enough to learn of the project's success and to see the early data. It scans the microwave background precisely enough to detect differences of a few micro-degrees between the temperatures in different directions. These variations are imprinted by the precursors of the structures like galaxies and clusters that we see today.

Theory tells us that the most conspicuous waves in the universe – those with the biggest amplitude – are concentrated at particular wavelengths.

There are, as it were, 'resonances' at particular frequencies, just as music in a small room can be distorted by the way particular notes resonate. We can calculate the wavelength of the dominant ripples. But how large they appear on the sky – whether, for instance, they are one degree across or only half a degree across – depends on the geometry of the universe, which in turn depends on the total content of the universe.

If there were nothing in the universe apart from atoms and dark matter with 0.3 of the critical density, the geometry would be what mathematicians call hyperbolic – if you draw a triangle on a saddle-shaped surface, its three angles add up to less than 180 degrees. Light rays travelling across the universe would behave as in a diverging lens; distant objects would then look smaller than they do in Euclidean space where the angles of a triangle add up exactly to 180 degrees.

The WMAP observations have pinned down the angular scale of this amplitude peak: it lies within a few percent of where it ought to be if the Universe were flat. If there were nothing else in the universe beyond the dark matter, we would expect an angle smaller by a factor of 2 – definitely in conflict with observations.

3. Some New Controversies

Cosmology used to be a subject with few facts. But cosmologists now speak of 'concordance'. We now have a convincingly-established framework for interpreting observations of distant objects, and for modelling how the fluctuations in the early universe evolve into the first gravitationally-bound cosmic structures, within which galaxies, stars and planets eventually emerge.

There is indeed remarkable consistency between several independent methods of measuring the key numbers describing our universe. It seems that the universe is flat – in harmony with theoretical prejudices. But there is a new and quite unexpected puzzle. Our universe contains an arbitrary-seeming mix of strange ingredients. Ordinary atoms (baryons), in stars, nebulae, and diffuse intergalactic gas, provide just 4 percent of the mass-energy; dark matter provides 23 percent; and dark energy the rest (i.e. 73 percent). These at least are the values that fit the data best. The expansion accelerates because dark energy (with negative pressure) is the dominant constituent. Of the atoms in the universe, only about half are in galaxies and the rest are diffusely spread through intergalactic space. The most conspicuous things in the cosmos, the stars and glowing gas in galaxies, are less than 2 percent of the universe's total budget of mass-energy – an extraordinary turnaround from what would have been the natural presumption at the start of the 20th century.

There is also firmer evidence for a hot dense 'beginning'. The extrapolation back to a stage when the Universe had been expanding for a few seconds (when the helium formed) deserves to be taken as seriously as, for instance, what geologists or palaeontologists tell us about the early history of our Earth: their inferences are just as indirect (and less quantitative). Several discoveries might have been made over the last thirty years, which would have invalidated the hypothesis, but these have not been made – the big bang theory has lived dangerously for decades, and survived.

But as always in science, each advance brings into sharper focus a new set of perplexing issues: in particular, why was the universe 'set up' expanding the way it is, with such a perplexing mix of ingredients?

Most cosmologists suspect that the uniformity, and the special-seeming expansion rate, are legacies of something remarkable that happened in the first trillionth of a trillionth of a trillionth of a second. The expansion would then have been exponentially accelerated, so that everything in the presently visible part of our universe could have inflated, homogenised, and established the 'fine tuned' balance between gravitational and kinetic energy when that entire domain was only a few centimetres across.

This concept of 'inflation', depending on assumptions about physics far beyond the regime where we have experimental tests, plainly has unsure foundations, but it is not just metaphysics: one can test particular variants of the idea. For instance, the seeds for galaxies and clusters could have been tiny quantum fluctuations, imprinted when the entire universe was of microscopic size, and stretched by inflationary expansion.

The details of the fluctuations depend, in a calculable way, on the physics of 'inflation'. The microwave background, a relic of the pregalactic era, should bear the imprint of these fluctuations. The European Space Agency's Planck/Surveyor spacecraft will, within a few years, yield precise enough data to settle many key questions about cosmology, the early universe, and how galaxies emerged. Such observations will therefore be able to test various possible assumptions about the currently-uncertain physics that prevailed under the extreme conditions of inflation, and thereby at least narrow down the range of options. We will surely learn things about 'grand unified' physics that cannot be directly inferred from ordinary-energy experiments.

4. Beyond 'Our' Big Bang

Another tantalising possibility is that physical reality could be far more extensive than what we have traditionally called our universe. We can only observe the aftermath of 'our' big bang. But there could be an infinity of 'big bangs' within an eternally expanding substratum. Many three-dimensional universes can be packed, so that they do not overlap each other, in a space with 4 or more dimensions. Bugs crawling on a large sheet of paper – their two-dimensional universe – would be unaware of other bugs on a separate sheet of paper. Likewise, we would be unaware of our counterparts on another space-time just a millimetre away, if that millimetre were measured in a 4th spatial dimension, and we are imprisoned in just three.

And the big bangs may all be different. 'Are the laws of physics unique?' is a prosaic paraphrase of Einstein's famous question: 'Did God have any choice when he created the universe?'. Perhaps the ingredients of our universe, and the fluctuations that are the 'seeds' for galaxies, are 'environmental contingencies', imprinted in the immediate aftermath of our big bang, rather than given uniquely by some magic formula. Perhaps, in this enlarged perspective, what we have traditionally called the laws of nature – even those of Einstein and the quantum – could be mere parochial bylaws in our local cosmic patch. There may still be a 'final' theory, at a deeper level, that holds sway over an infinite 'ecological variety' of big bangs.

As an analogy, consider the form of snowflakes. Their ubiquitous six-fold symmetry is a direct consequence of the properties and shape of water molecules. But snowflakes display an immense variety of patterns because each is moulded by its micro-environments: how each flake grows is sensitive to the fortuitous temperature and humidity changes during its growth. The fundamental theory should tell us which aspects of nature are direct consequences of the bedrock theory (just as the symmetrical template of snowflakes is due to the basic structure of a water molecule) and which are (like the distinctive pattern of a particular snowflake) environmental contingencies peculiar to 'our' big bang.

This line of thought is an extrapolation of a perspective-shift that we have already made on a more modest scale – that of stars and planets. We have learnt that millions of stars each have planetary system. So it is unsurprising to find some planets like our rare Earth – planets with the optimum size, temperature, orbit to allow a biosphere. What we have traditionally called 'the universe' may be the outcome of one big bang among many, just as our Solar System is merely one of many planetary systems in the Galaxy. We look for beautiful mathematics in nature, but we do not always find it. Kepler thought that planetary orbits were related by beautiful mathematics. We now know that the Earth's orbit is the outcome of messy and complex dynamics – but happens to end up at a habitable distance from the Sun. The quest for exact formulae for what we normally call the constants of nature may consequently be as doomed as was Kepler's quest for the exact numerology of planetary orbits. And other big bangs will become part of scientific discourse, just as 'other planets' now are.

We still await a 'battle tested' fundamental theory, corroborated by measurements we can actually make, that tells us whether there could have been many 'big bangs' rather than just one, and (if so) how much variety they might display. Until then the epistemological status of the other big bangs is of course precarious.

5. Back to Earth

Humans on Earth are the outcome of four billion years of Darwinian selection. But our solar system is barely middle-aged: the Sun has been shining on the Earth for 4.5 billion years, The unfolding of complexity is just beginning – perhaps humans are still an early stage in this process, rather than (as we sometimes seem to believe) its culmination. There is an unthinking tendency to imagine that humans will be around in 6 billion years, watching the Sun flare up and die. But any life and intelligence that exists then could be as different from us as we are from a bacterium.

Do any general arguments set limits to evolution and complexity – or is the potential literally infinite? The science of complexity probably offers even greater challenges than the 'fundamental' sciences of the cosmos and the microworld.