Scientific Insights into the Evolution of the Universe and of Life
Pontifical Academy of Sciences, Acta 20, 2009
www.pas.va/content/dam/accademia/pdf/acta20/acta20-hawking.pdf

THE ORIGIN OF THE UNIVERSE

STEPHEN HAWKING

Early accounts of the origin of the world were attempts to answer the questions we all ask: Why are we here? Where did we come from? The answer generally given was that humans were of comparatively recent origin, because it must have been obvious, even at early times, that the human race was improving in knowledge and technology. So it can't have been around that long, or it would have progressed even more. On the other hand, the physical surroundings, like mountains and rivers, change very little in a human lifetime. They were therefore thought to be a constant background, and either to have existed forever as an empty landscape, or to have been created at the same time as the humans.

Not everyone however, was happy with the idea that the universe had a beginning. For example, Aristotle, the most famous of the Greek philosophers, believed the universe had existed forever. Something eternal is more perfect than something created. He suggested the reason we see progress, was that floods, or other natural disasters, had repeatedly set civilization back to the beginning.

If one believed that the universe had a beginning, the obvious question was, what happened before the beginning? What was God doing before He made the world? Was He preparing Hell for people who asked such questions? The problem of whether or not the universe had a beginning was a great concern to the German philosopher Immanuel Kant. He felt there were logical contradictions, or Antimonies, either way. If the universe had a beginning, why did it wait an infinite time before it began? He called that the thesis. On the other hand, if the universe had existed forever, why did it take an infinite time to reach the present stage? He called that the antithesis. Both the thesis, and the antithesis, depended on Kant's assumption, along with almost everyone else, that time was Absolute. That is to say, it went from the infinite past, to the infinite future, independently of any universe that might or might not exist in this background.

This is still the picture in the minds of many scientists today. However, in 1915, Einstein introduced his revolutionary General Theory of Relativity. In this, space and time were no longer absolute, no longer a fixed background to events. Instead, they were dynamical quantities that were shaped by the matter and energy in the universe. They were defined only within the universe, so it made no sense to talk of a time before the universe began. It would be like asking for a point south of the South Pole: it is not defined.

If the universe was essentially unchanging in time, as was generally assumed before the 1920s, there would be no reason that time should not be defined arbitrarily far back. Any so-called beginning of the universe would be artificial, in the sense that one could extend the history back to earlier times. Thus, it might be that the universe was created last year, but with all the memories and physical evidence to look like it was much older. This raises deep philosophical questions about the meaning of existence. I shall deal with these by adopting what is called the positivist approach. In this, the idea is that we interpret the input from our senses in terms of a model we make of the world. One cannot ask whether the model represents reality, only whether it works. A model is a good model, if first it interprets a wide range of observations, in terms of a simple and elegant model. And second, if the model makes definite predictions that can be tested, and possibly falsified, by observation.

In terms of the positivist approach, one can compare two models of the universe. One in which the universe was created last year, and one in which the universe existed much longer. The Model in which the universe existed for longer than a year can explain things like identical twins, that have a common cause more than a year ago.

On the other hand, the model in which the universe was created last year, cannot explain such events. So the first model is better. One cannot ask whether the universe *really* existed before a year ago, or just appeared to. In the positivist approach, they are the same.

In an unchanging universe, there would be no natural starting point. The situation changed radically however, when Edwin Hubble began to make observations with the hundred-inch (2.5m) telescope on Mount Wilson, in the 1920s.

Hubble found that stars are not uniformly distributed throughout space, but are gathered together in vast collections called galaxies.

By measuring the light from galaxies, Hubble could determine their velocities. He was expecting that as many galaxies would be moving towards us, as were moving away. This is what one would have in a universe that was unchanging with time. But to his surprise, Hubble found that nearly all the galaxies were moving away from us. Moreover, the further galaxies were from us, the faster they were moving away. The universe was not unchanging with time, as everyone had thought previously: it was expanding. The distance between distant galaxies was increasing with time.

The expansion of the universe was one of the most important intellectual discoveries of the 20th century, or of any century. It transformed the debate about whether the universe had a beginning: if galaxies are moving apart now, they must have been closer together in the past. If their speed had been constant, they would all have been on top of one another, about 15 billion years ago. Was this the beginning of the universe?

Many scientists were still unhappy with the universe having a beginning, because it seemed to imply that physics broke down. One would have to invoke an outside agency, to determine how the universe began. They therefore advanced theories in which the universe was expanding at the present time, but didn't have a beginning. One was the Steady State theory, proposed by Bondi, Gold, and Hoyle in 1948.

In the Steady State theory, as galaxies moved apart, the idea was that new galaxies would form from matter that was supposed to be continually being created throughout space. The universe would have existed forever, and would have looked the same at all times. This last property had the great virtue, from a positivist point of view, of being a definite prediction that could be tested by observation. The Cambridge radio astronomy group, under Martin Ryle, did a survey of weak radio sources in the early 1960s. These were distributed fairly uniformly across the sky, indicating that most of the sources lay outside our galaxy. The weaker sources would be further away, on average.

The Steady State theory predicted the shape of the graph of the number of sources, against source Strength. But the observations showed more faint sources than predicted, indicating that the density of sources was higher in the past. This was contrary to the basic assumption of the Steady State theory, that everything was constant in time. For this, and other reasons, the Steady State theory was abandoned.

Another attempt to avoid the universe having a beginning was the suggestion that there was a previous contracting phase, but because of rotation and local irregularities, the matter would not all fall to the same point. Instead, different parts of the matter would miss each other, and the universe would expand again, with the density remaining finite. Two Russians, Lifshitz and Khalatnikov, actually claimed to have proved that a general contraction without exact symmetry, would always lead to a bounce, with the density remaining finite. This result was very convenient for Marxist-Leninist dialectical materialism, because it avoided awkward questions about the creation of the universe. It therefore became an article of faith for Soviet scientists.

When Lifshitz and Khalatnikov published their claim, I was a 21-year-old research student, looking for something to complete my PhD thesis. I didn't believe their so-called proof, and set out with Roger Penrose to develop new mathematical techniques to study the question. We showed that the universe couldn't bounce. If Einstein's General Theory of Relativity is correct, there will be a singularity, a point of infinite density and space-time curvature, where time has a beginning.

Observational evidence to confirm the idea that the universe had a very dense beginning came in October 1965, a few months after my first singularity result, with the discovery of a faint background of microwaves throughout space. These microwaves are the same as those in your microwave oven, but very much less powerful. They would heat your pizza only to minus 271.3°C, not much good for defrosting the pizza, let alone cooking it. You can actually observe these microwaves yourself. Set your television to an empty channel. A few percent of the snow you see on the screen will be caused by this background of microwaves. The only reasonable interpretation of the background is that it is radiation left over from an early very hot and dense state. As the universe expanded, the radiation would have cooled until it is just the faint remnant we observe today.

Although the singularity theorems of Penrose and myself predicted that the universe had a beginning, they didn't say how it had begun. The equations of General Relativity would break down at the singularity. Thus, Einstein's theory cannot predict how the universe will begin, but only how it will evolve once it has begun. There are two attitudes one can take to the results of Penrose and myself. One is that the way the universe began is not within the realm of science. The other interpretation of our results, which is favoured by most scientists, is that it indicates that the General Theory of Relativity breaks down in the very strong gravitational fields in the early universe. It has to be replaced by a more complete theory. One would expect this anyway, because General Relativity does not take account of the small-scale structure of matter, which is governed by quantum theory. This does not matter normally, because the scale of the universe is enormous compared to the microscopic scales of quantum theory. But when the universe is the Planck size, a billion trillion trillionth of a centimetre, the two scales are the same, and quantum theory has to be taken into account.

In order to understand the Origin of the universe, we need to combine the General Theory of Relativity with quantum theory. The best way of doing so seems to be to use Feynman's idea of a sum over histories. Richard Feynman was a colourful character, who played the bongo drums in a strip joint in Pasadena, and was a brilliant physicist at the California Institute of Technology. He proposed that a system got from a state A to a state B by every possible path or history.

Each path, or history, has a certain amplitude or intensity, and the probability of the system going from A to B is given by adding up the amplitudes for each path. There will be a history in which the moon is made of blue cheese, but the amplitude is low, which is bad news for mice.

The probability for a state of the universe at the present time is given by adding up the amplitudes for all the histories that end with that state. But how did the histories start? This is the Origin question in another guise. Is the initial state of the universe determined by a law of science?

In fact, this question would arise even if the histories of the universe went back to the infinite past. But it is more immediate if the universe began only 15 billion years ago. The problem of what happens at the beginning of time is a bit like the question of what happened at the edge of the world, when people thought the world was flat. Is the world a flat plate, with the sea pouring over the edge? I have tested this experimentally: I have been round the world, and I have not fallen off.

As we all know, the problem of what happens at the edge of the world was solved when people realized that the world was not a flat plate, but a curved surface. Time, however, seemed to be different: it appeared to be separate from space, and to be like a model railway track. If it had a beginning, there would have to be someone to set the trains going.

Einstein's General Theory of Relativity unified time and space as space-time, but time was still different from space, and was like a corridor which either had a beginning and end, or went on for ever. However, when one combines General Relativity with Quantum Theory, Jim Hartle and I realized that time can behave like another direction in space under extreme conditions. This means one can get rid of the problem of time having a beginning, in a similar way in which we got rid of the edge of the world. Suppose the beginning of the universe was like the south pole of the earth, with degrees of latitude playing the role of time. The universe would start as a point at the South Pole. As one moves north, the circles of constant latitude, representing the size of the universe, would expand. To ask what happened before the beginning of the universe would become a meaningless question, because there is nothing south of the south pole.

Time, as measured in degrees of latitude, would have a beginning at the South Pole, but the South Pole is much like any other point, at least so I have been told. I have been to Antarctica, but not to the South Pole.

The same laws of Nature hold at the South Pole, as in other places. This would remove the age-old objection to the universe having a beginning that it would be a place where the normal laws broke down. The beginning of the universe would be governed by the laws of science.

The picture Jim Hartle and I developed of the spontaneous quantum creation of the universe would be a bit like the formation of bubbles of steam in boiling water.

The idea is that the most probable histories of the universe would be like the surfaces of the bubbles. Many small bubbles would appear, and then disappear again. These would correspond to mini universes that would expand, but would collapse again while still of microscopic size. They are possible alternative universes, but they are not of much interest since they do not last long enough to develop galaxies and stars, let alone intelligent life. A few of the little bubbles, however, will grow to a certain size at which they are safe from recollapse. They will continue to expand at an ever-increasing rate, and will form the bubbles we see. They will correspond to universes that would start off expanding at an ever-increasing rate. This is called inflation, like the way prices go up every year.

The world record for inflation was in Germany after the First World War: prices rose by a factor of ten million in a period of 18 months. But that was nothing compared to inflation in the early universe: the universe expanded by a factor of million trillion trillion in a tiny fraction of a second. Unlike inflation in prices, inflation in the early universe was a very good thing. It produced a very large and uniform universe, just as we observe. However, it would not be completely uniform. In the sum over histories, histories that are very slightly irregular will have almost as high probabilities as the completely uniform and regular history. The theory therefore predicts that the early universe is likely to be slightly non-uniform. These irregularities would produce small variations in the intensity of the microwave background from different directions. The microwave background has been observed by the Map satellite, and was found to have exactly the kind of variations predicted. So we know we are on the right lines.

The irregularities in the early universe will mean that some regions will have slightly higher density than others. The gravitational attraction of the extra density will slow the expansion of the region, and can eventually cause the region to collapse to form galaxies and stars. So look well at the map of the microwave sky: it is the blueprint for all the structure in the universe. We are the product of quantum fluctuations in the very early universe. God really does play dice.

We have made tremendous progress in cosmology in the last hundred years. The General Theory of Relativity and the discovery of the expansion of the universe shattered the old picture of an ever existing, and ever lasting universe. Instead, general relativity predicted that the universe, and time itself, would begin in the big bang. It also predicted that time would come to an end in black holes. The discovery of the cosmic microwave background and observations of black holes support these conclusions. This is a profound change in our picture of the universe and of reality itself.

Although the General Theory of Relativity predicted that the universe must have come from a period of high curvature in the past, it could not predict how the universe would emerge from the big bang. Thus, general relativity on its own cannot answer the central question in cosmology, Why is the universe the way it is? However, if general relativity is combined with quantum theory, it may be possible to predict how the universe would start. It would initially expand at an ever-increasing rate. During this so-called inflationary period, the marriage of the two theories predicted that small fluctuations would develop and lead to the formation of galaxies, stars, and all the other structure in the universe. This is confirmed by observations of small non-uniformities in the cosmic microwave background with exactly the predicted properties. So it seems we are on our way to understanding the origin of the universe, though much more work will be needed.

Despite having had some great successes, not everything is solved. We do not yet have a good theoretical understanding of the observations that the expansion of the universe is accelerating again, after a long period of slowing down. Without such an understanding, we cannot be sure of the future of the universe. Will it continue to expand forever? Is inflation a law of Nature? Or will the universe eventually collapse again? New observational results, and theoretical advances are coming in rapidly. Cosmology is a very exciting and active subject. We are getting close to answering the age-old questions: ‘Why are we here?’ ‘Where did we come from?’ I believe these questions can be answered within the realm of science.

Thank you for listening to me.