Open Questions in Cosmology

In 1920 the Smithsonian Museum of Natural History hosted a debate between Harlow Shapley and Heber Curtis, two of the most prominent astronomers of the day, over the question of whether *anything* exists outside our galaxy. Less than a decade after that debate, Edwin Hubble showed that the universe is filled with distant galaxies—and that those galaxies are rushing away from each other in a universal, cosmic expansion.

In the ensuing century we learned how the hot, dense, expanding early universe gave rise to particles, then nuclei, then atoms, and eventually the stars and galaxies we see filling the universe today. We calculated in mathematical detail patterns of quantum fluctuations that arose during a span of perhaps 10^{-39} s almost 14 billion years ago, and used those calculations to accurately predict patterns in the cosmic microwave background radiation. Today our telescopes can see far enough to observe light from some of the earliest galaxies, formed a few hundred million years after the Big Bang. In short, almost all the information in Chapter 14, our introduction to modern cosmology, represents discoveries made since the Shapley–Curtis debate.

But despite the amazing theoretical and observational progress of the last century, many open questions remain. In this section we will briefly describe some of those questions, and the work that is currently being done to address them. Those questions include:

- What happened before the Big Bang?
- What were the details of how inflation occurred, and how it ended?
- How did the universe come to have more matter than antimatter?
- What is dark matter, and what is dark energy?
- How did the first stars form?
- Is our observable universe representative of the entire universe, or are we part of a "multiverse" whose regions have very different histories and properties?

Different cosmologists might come up with somewhat different lists, but we think most would agree that the questions presented here are among the main issues driving cosmology research today.

What Happened Before the Big Bang?

In Section 14.1 we emphasize that the Big Bang was *not* the beginning of the universe. Rather, the Big Bang was the moment, approximately 13.8 billion years ago, when the universe dropped below “Planck density” (10^{96} kg/m^3). We can formulate theories about how the universe has evolved from that moment until this, and we can test those theories observationally. Those theories are based on the most fundamental laws we know: quantum field theory, and general relativity. But we cannot formulate and test theories about the universe before the Big Bang, because the laws we understand do not apply to matter above Planck density.

We don’t say much in Chapter 14 about *why* our current known physical laws cannot be applied to such high-density material, but here’s a quick explanation.

Quantum mechanics says that a particle can exist in a superposition of different positions and momenta. General relativity (www.cambridge.org/felder-modernphysics/onlinesections) says that mass and energy affect the geometry of spacetime.

Taken together, these two theories imply that a particle in a superposition of different positions must cause spacetime to exist in a superposition of different geometries. But when we try to write a mathematical description of such a state, the math ends up giving us inconsistent, contradictory results.

Practically speaking, this does not limit our ability to use these theories to predict the results of experiments. General relativity works great on the very large scales of astronomy and cosmology, reduces correctly to Newtonian physics on ordinary scales, and is irrelevant at very small scales. Quantum mechanics works great on the small scales of particle physics, reduces correctly to Newtonian physics on ordinary scales, and is irrelevant on very large scales. So we can use quantum mechanics, Newtonian mechanics, and general relativity, each in its proper domain.

But above Planck density, quantum fluctuations of particles and fields should lead to non-negligible fluctuations of spacetime geometry. We have no idea what happens in such a case. Which brings us back to where we started: we have no way of describing or predicting the behavior of the universe before the Big Bang.

To resolve that puzzle, we need a self-consistent theory of quantum gravity. The best-known attempt at such a theory is string theory. Calculations suggest that string theory is self-consistent, reproduces the predictions of general relativity at large scales where quantum effects are negligible, and reproduces the predictions of quantum field theory at small scales where gravity is negligible.

So what does string theory predict about the behavior of matter above Planck density? We don’t know, because the math is so complicated that we can’t solve the equations to figure that out. And even if we find some theoretical answers, we won’t believe those answers until we can test them—and there’s no practical way to do that. We can’t create Planck-density matter, which would require the equivalent of packing a billion galaxies into an atomic nucleus. The only Planck-density matter we know of in the post-Big-Bang universe is at the center of black holes, but nothing can send a signal out from the center of a black hole to let us know what’s happening there.

So right now (2022) we have no good theories to test, and no way to test them even if we had them.

Many physicists and mathematicians are working to better understand the predictions of string theory. Others are working out different proposed theories of quantum gravity. Perhaps in the coming decade we’ll develop these theories well enough to determine the correct one, and then continue our extrapolations of the early universe back to earlier times.

What Were the Details of How Inflation Occurred, and How It Ended?

Section 14.6 explores three mysteries left open by the original Big Bang model: the large-scale homogeneity of the universe, the universe's large-scale Euclidean geometry, and the lack of magnetic monopoles and other large relic particles that could have been produced in the early universe. Section 14.7 describes how all three mysteries are resolved by inflation, a brief period of exponential expansion that occurred a fraction of a second after the Big Bang. Inflationary theory has made successful predictions, such as the mathematical pattern of inhomogeneities observed in the cosmic microwave background.

So there are a lot of reasons to believe that inflation actually happened. But there are a lot of unanswered questions about how and why it happened.

We believe the cause of inflation was a "scalar field" that dominated the energy of the early universe. But we have never observed a scalar field with the right properties to have been responsible for inflation. You will sometimes hear the phrase "inflaton field," but that's just a label that means "whatever that field was, that we don't understand, that caused inflation."

It's no great surprise that we have never observed a field like the inflaton. The Higgs field is the only scalar field we *have* actually observed. Most particle physicists believe that many other scalar fields exist, but that exciting those fields requires more energy than we can produce in an accelerator.

Because we have never observed the inflaton field, we know very little about its properties, and that limits our knowledge of the inflationary period of the universe. Many models of inflation have been proposed, each specifying different properties for the inflaton field. Until we know which (if any) of these models is correct, we can't answer basic questions such as "How long did inflation last?" or "How much did the universe expand during inflation?" Perhaps more importantly, knowing the properties of the inflation field would allow us to model how it decayed and produced the matter we see in the universe today.

Theorists are working on using models of inflation to make testable predictions, and observers are making ever-more accurate measurements to give us clues about these early processes. For example, some models of inflation predict that the decay of the inflaton would have produced strong gravitational waves, and experiments are being planned that might detect those waves.

Some of the most promising work involves measurements of the polarization of the cosmic microwave background. Essentially all models of inflation predict the same fluctuations in *intensity* for this radiation, but different models make different predictions about how the *polarization* of the CMB should vary across the sky. Measurements of that polarization have already ruled out some of the simplest models of inflation, and future experiments may constrain those models much more significantly.

In addition to teaching us about the early universe, knowledge of the inflaton field would be a great first hint about particle physics at higher energies than we can measure in accelerators. Many theories such as "grand unified theories" or "supersymmetry" predict the existence of fields that could only be seen with accelerators much more powerful than we can build. Using cosmological observations to figure out the properties of the inflaton could help us to test and refine some of those theories.

We should note that while most cosmologists believe inflation occurred in the early universe, there are skeptics who are actively pursuing alternative theories. Such theories have to reproduce the successful predictions inflation has made, including the detailed structure of the CMB fluctuations, but some theorists believe that other theories may be able to do that. If so, then further experiments will be needed to determine whether inflation or one of these other processes truly occurred.

How Did the Universe Come to Have More Matter Than Antimatter?

During inflation, the universe expanded by so much in such a short time that the density of all types of particles was driven to almost exactly zero. Any particles that existed before inflation presumably continued to exist afterwards, but the probability of one of those particles being in our observable universe is vanishingly small.

The only type of energy that did survive inflation with non-negligible density was that of the inflaton field itself. That means all of the particles in the observable universe—both matter and antimatter particles—came from the decay of the inflaton field. So far, so good; we see many processes where a field or type of particle can decay and produce other types of fields and particles.

If the universe had ended up with a balanced mix of matter and antimatter, then a moment later all of those particles would have annihilated. Because we know that the surviving universe (our universe) still has some matter, the decay of the inflaton must have created more matter than antimatter.

How much more? We can answer that by counting the number of matter particles and photons in the universe today. Our theories suggest that, following the decay of the inflaton field, there should have been a roughly equal number of photons and matter (or antimatter) particles. Every time a matter-antimatter pair annihilated, it produced two photons. In the universe today there are about a billion times more photons than matter particles, suggesting that there were about a billion annihilation events for every one particle that survived.[1]

The conclusion is that when matter was first created, roughly a billion and one matter particles were created for each billion antimatter particles. This is called "baryogenesis" because it left the universe with a net excess of baryons over antibaryons. (It also created a net excess of leptons, but baryons got to claim the name.)

But here's the rub: in every process we know of where particles are formed from other types of matter, exactly equal amounts of matter and antimatter are created. Nobody knows what caused the post-inflation asymmetry, or what process *could* have caused such an asymmetry. Understanding this process will require advances in our understanding of fundamental particle physics. And as we advance our particle physics theories, we will be able to test them by seeing how well they predict the results of baryogenesis.

1 Those photons, almost all of which are in the cosmic microwave background radiation, have so much less energy per particle than matter that they collectively make up about 0.01% of the energy in the observable universe, even though they make up virtually 100% of the particles.

What Is Dark Matter, and What Is Dark Energy?

These two questions are discussed in Section 14.5, so we will just briefly review them here: first dark matter, and then dark energy.

A galaxy is a disk-shaped cluster of stars and planets, and those stars and planets are in turn made of protons, neutrons, and electrons. You know, ordinary matter. But surrounding and permeating that disk of ordinary matter is a spherical halo of "dark matter." The name is quite literal: this type of matter does not emit, absorb, or reflect light, so it is invisible.

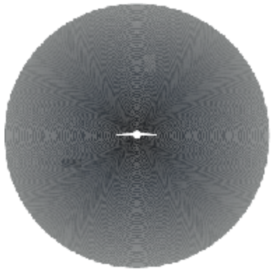

Figure 1 A galaxy (white) is surrounded by a dark matter halo more than five times wider than the disk of visible matter.

We know dark matter is out there because of its gravitational effects. Those effects have been measured and confirmed in numerous ways.

- Calculations of galactic rotations—essentially the same orbital calculations that Newton did for the solar system—imply the existence of far more mass inside each galaxy than all the stars we can see. It was this observation that led Vera Rubin to postulate the existence of dark matter in the 1970s. Her calculations showed that a typical galaxy contains roughly five times as much dark matter as regular (visible) matter, by mass.
- As the light from distant galaxies propagates toward us, the path of that light is bent by the gravitational pull of galaxies it passes on the way. This "gravitational lensing" was predicted by general relativity, and the degree of bending provides an independent measurement of galactic masses. The resulting masses do *not* match the stars that we see, but they *do* match the much higher masses that Vera Rubin predicted from her rotation curves.
- Galaxies are formed when regions of unusually high density pull more matter toward them. Computer simulations of this process predict the pattern of galaxies that we should find in the universe. Once again, these simulations do not match observation if we base them solely on visible matter, but they do make correct predictions if we build in the invisible matter predicted by Vera Rubin.

We believe that dark matter is a type of particle that interacts with normal matter very weakly, either only through gravity or through gravity and the weak force.[2] Many experiments are underway to directly detect dark matter, but as of this writing none have yet succeeded.

But while dark matter represents much more energy than ordinary matter, the majority of the energy in the observable universe is in yet another form: dark energy.

We know dark energy exists because the expansion of the universe is accelerating. That means something is pushing galaxies away from each other, and with greater force than their mutual gravitational attraction. What could cause such repulsive gravity? There are two leading theories.

---

2 There is much more dark matter than ordinary matter, but it is also more evenly spread out throughout the galaxy than ordinary matter is. A solar system is a clump of matter extraordinarily more dense than the average density of the galaxy, so within a solar system ordinary matter dominates over dark matter.

One theory is a scalar field. Accelerated expansion occurred in the early universe during inflation, and we believe a scalar field caused that period of acceleration. So it's quite possible that a *different* sort of scalar field could be causing our current acceleration. (It could conceivably be the same one, but that would require extremely fine-tuned parameters.)

Another possibility is that space itself may have an intrinsic energy density, sometimes called "vacuum energy" and sometimes a "cosmological constant."[3] In the equations of general relativity, such a vacuum energy would cause accelerated expansion. Such energy might occur because of ground state energies of random fluctuations of fields like the electromagnetic field. But when you use quantum field theory to calculate how big that ground state energy density should be, you get an answer 10^{120} times bigger than the observed density of dark energy! Oops. Something is clearly wrong with our understanding of how to do these calculations. Some theorists suspect that correctly predicting the vacuum energy density will require a theory of quantum gravity.

Right now we have no way to directly observe dark energy; we can only see the effect it has on the expansion of the universe. But we hope to learn more about it by studying in detail how that expansion has changed over the universe's history. If dark energy is a scalar field, then its density might have changed over time, which would lead to a different expansion history than a constant vacuum energy density. Measurements of that expansion history are underway.

The past century has seen tremendous progress in our understanding of the basic building blocks of the universe. From Thomson's 1897 discovery of the electron—the first indication that atoms were themselves made of smaller particles—we progressed down to the nucleus, and from there down to the standard model of particle physics (Section 13.2), which underlies all the particles we have created or measured. But dark matter is not built from the particles in the standard model, and neither is dark energy. So our attempts to unravel those mysteries constitute explorations of fundamental substances that today are only dimly understood.

How Did the First Stars Form?

Stars shine because of nuclear fusion. Lighter elements, primarily hydrogen, fuse into heavier elements in the core of a star. When there isn't enough nuclear fuel left the star ends its life and explodes, scattering these heavier elements into the space around it. Later generations of stars form from this debris, so each successive generation forms with a higher ratio of heavy to light elements. Our sun is a third generation star, and probably started with 1–2% elements heavier than helium.

Computer programs have simulated the process of star formation in great detail, and those simulations have been confirmed by observing stars at various stages of the formation process. But those simulations show that for a star to collapse enough to begin forming, it needs a certain fraction of heavy elements. Even a small fraction, such as our sun had initially, is enough to enable the star to form.

3 You may remember the cosmological constant as the term Einstein introduced into general relativity to try to allow for a static universe. It was later shown that the modification of the equations he proposed is exactly what you would get if you used his original equations but assumed a constant energy density of space itself, in addition to the energy of the stuff in that space.

But the earliest stars didn't have heavy elements, so they had to form essentially entirely from hydrogen and helium. The process by which that formation occurred is poorly understood. What we do know about it suggests that these early stars would have been extremely large and short-lived, so we won't find any of them still around in our galaxy.

By looking far enough away, however, we could in principle see the light cast by those first stars, over 13 billion years ago. As of this writing, the James Webb Space Telescope has just begun operation, and within the next few years it may be able to measure that ancient starlight and tell us about how the first stars came to be.

Is Our Universe Part of a Multiverse?

Because the Big Bang happened a finite time ago, and light travels at a finite speed, we can only see a limited distance around us. The region of the universe we can see is called our "observable universe" (see Section 14.3), and the edge of that region is called our "cosmological horizon" (or sometimes "particle horizon").

Within our observable universe, every (sufficiently large-scale) neighborhood is pretty much like every other one. The galaxies we see in one direction have roughly the same sizes, shapes, and compositions as the ones we see in any other direction. But does that uniformity hold throughout the entire universe? Some theories suggest that our observable universe might be an island of uniformity within a much larger and widely varied universe. Perhaps there are regions that have different compositions, or no matter at all, or even apparently different physical laws. Perhaps there are regions where the Big Bang occurred earlier or later than it did for us. Perhaps there are regions separated from us by other dimensions, or even not physically connected to our observable universe at all.

Such speculations have long been the realm of science fiction, but a number of current theories suggest that some or all of these things might actually be true in our universe. Just to give you a flavor of these theories, we'll describe one of them here.

Some theories like string theory suggest that our universe might have more than three spatial dimensions. So why do we perceive only three? One possible answer is that our 3D world is a surface floating in a higher dimensional space. You can imagine this by picturing a 2D plane inside a 3D space. If the 2D inhabitants of that plane can't move off the plane and can't see anything beyond it, they would think the entire universe was 2D. Now imagine that within that 3D space there's another 2D plane floating above the first one. From the point of view of the beings on that first 2D plane, this second one is effectively another universe, cut off from their own by an un-bridgeable gulf.

By analogy, we might live in a 3D world floating in a 4D space, and perhaps there are other such 3D worlds out there. How could we tell if that were the case? We might be able to produce a particle that, unlike us, can move off of our 3D surface. We would see an experiment in which the total energy of the particles we saw coming out was less than the total energy we put in. Physicists are actively looking for such signatures in accelerator experiments, but to date none have been found.

Final Thoughts

Of your two authors, one of us—Gary—actually researches and teaches about cosmology for a living. Part of his teaching work is a downloadable course on the early universe, part of the "Great Courses" series from Wondrium. If you want to explore some of the themes in this section in more depth, this series of lectures is available in audio and video formats at: https://www.thegreatcourses.com/courses/the-big-bang-and-beyond-exploring-the-early-universe Cosmology is still in its infancy. Just a century after we realized that we live in an expanding universe of galaxies, we now know a tremendous amount about the structure and history of the universe. But huge questions still remain. It is our fond hope that some of the students reading our book today will become the scientists answering those questions tomorrow.