Implications and rejection: black holes and the expanding universe

Reference webpages:

http://skyserver.sdss.org/dr1/en/astro/universe/universe.asp
http://cosmology.berkeley.edu/Education/BHfaq.html
and Chapter 3 in Thorne. For far more details about cosmology, at an undergraduate major level, my lectures are at http://www.astro.umd.edu/∼miller/teaching/astr422/

Questions to keep in mind are:

1. Is the cosmos infinite?
2. Some current theories (e.g., superstrings) imply lots of other wild things, e.g., many extra dimensions. How much should we believe this?
3. How much evidence do we need about a theory to believe its currently unobserved predictions? One point of view is that we can never believe any unobserved predictions, the other is that beyond some threshold of success it is reasonable to accept such predictions.

The fourth of our debates will be on the last of these topics.

Introduction

Last time we discussed general relativity, which was driven by largely philosophical concerns but had comparatively little observational support until the last few decades. Nonetheless, a (small!) band of physicists explored various consequences of the theory, and it became generally accepted by physicists after Eddington's 1919 eclipse expedition. Among the consequences of the theory were that the universe had to be dynamic (expanding or contracting as a whole), and that black holes were a possibility. Remarkably, Einstein resisted both conclusions, but for different reasons. To understand his point of view, we need to reach back in history to explore pre-Einsteinian concepts and to determine whether the universe is finite or infinite.

Newton and the infinite universe

When Newton came up with with his concept of universal gravitation, he realized after some contemplation that there was a potentially serious problem. Suppose that the universe is finite, and let's consider a star on the outer edge of the universe. Then there is a gravitational acceleration that pulls it in, but none that pulls it out. As a result, we would expect matter to collect towards the center. Given that the strength of gravity increases as matter gets closer, you might think that this would continue, meaning that in a finite time all the matter in the universe would be collected in one spot. This hasn't happened, so there is a contradiction.

We do have to be careful, though. Consider our galaxy, the Milky Way. Suppose that this was all there is. Would it collapse? Certainly not in a short time. Our galaxy rotates, and this helps support it against collapse. Investigations long after Newton have revealed that if you wait long *enough*, the system isn't completely stable. Ultimately (and by that we mean more than 10^{20} years!), the outer parts will move away and the inner parts will grow ever denser.

Perhaps guided by such an instinct, Newton felt that the better solution was to imagine that although on local scales the universe is clearly not homogeneous (each region the same as any other region of the same volume) or isotropic (all directions look the same), that on large scales this is the way it is, and that the universe is infinite. These are, incidentally, assumptions made by modern-day cosmologists to simplify their predictions. To understand why this might appear to solve the problem, note that if you are in an infinite, homogeneous universe then all directions are the same and therefore the pull is the same in all directions as well. As a result, you will feel no net force and therefore there is no propensity for matter to cluster. You can think of this as a variant of the classic Buridan's Ass thought experiment: a donkey exactly between two equally enticing piles of hay would starve to death because it could not decide which way to go.

The problem is that, just as with the donkey, the solution is unstable. Suppose the donkey sneezes, and that is enough to move it slightly closer to one pile of hay than the other. Then the difference is small, so the donkey might make only slow, tenuous steps towards the nearer one, but as it moved the decision would become clearer and thus it would eat after all. In a similar way, unless the universe is mathematically, perfectly homogeneous, a given parcel of mass will always be attracted slightly more in one direction than another. Therefore, even in an infinite universe that is close to homogeneous, matter will collapse and form structure.

This by itself is therefore not a convincing reason why the universe has to be infinite. We need to determine what can be said observationally.

Why is the sky dark at night?

To address whether the universe is infinite, we can appeal to an argument called Olbers' paradox, which is so named because (as my advisor Sterl Phinney liked to say) Olbers was the fourth person to state the paradox and the second to give the wrong answer!

The question can be phrased simply: why is the sky dark at night? Basically, if the universe were infinite, then every line of sight would eventually hit a star, and thus the entire sky would have the same surface brightness (meaning energy per time per area per solid angle) as the Sun. We'd notice that! In the remainder of this lecture we will explore this paradox quantitatively.

First we'll start with the fundamental idea.

Ask class: Suppose that the universe is populated randomly with stars, but homogeneously, so that a typical volume will contain a typical number of stars and this number does not change as you go far away. The flux (energy per area per time) from a given star dies away as the inverse square of the distance: $F \propto 1/r^2$. If we consider a spherical shell of radius r from us and thickness $h \ll r$, then the volume of the shell is roughly $4\pi r^2 h$. The number of stars in that volume is proportional to the volume, so $N \propto 4\pi r^2 h$. Thus the total energy per area per time we get from this shell is equal to the number of stars in the shell times the flux per star. When you multiply these together, you see that the r^2 cancels the $1/r^2$, and we get a flux $F_{\rm shell} \propto h$. Thus every shell contributes equally. If we can go out to infinite distance, then, the flux should be infinite, in rather stark contrast to what we see!

One way out, of course, is that the universe is not infinite, not eternal, or both. If we could conclude this, it would represent a fantastically important inference about the cosmos. We are therefore obliged to think about other possible ways out, and how to test them, before we can draw any confident conclusions. What are some *other* possible ways out for us to explore?

As one try to resolve the paradox, we will consider the suggestion of Olbers himself. He suggested that since there exists gas and dust between the stars, the gas and dust would absorb and block the radiation. Unfortunately, this does not work if the observable universe is not only infinite, but eternal. This is because the gas and dust would heat up to the temperature of the illuminating stars after a finite time, so they would radiate just as much. You can't sweep this problem under an absorbing rug!

Another way might be to eliminate the homogeneity of the universe. If at greater distances the universe has more and larger holes, maybe you don't add up to unlimited light. But here observations come into play, and tell us that the universe does actually get more homogeneous at larger distances.

We can try another couple of possibilities, but eventually we can convince ourselves that the best solution is that the universe is not infinite. Note, though, that there are two different ways that this could happen. The universe might not be spatially infinite, or it might not have existed for infinite time. If the universe is spatially finite but eternal, we're back to Newton's concern. If the universe is spatially infinite but limited in age, however, we do get around Olbers' paradox because the finite speed of light means that we can only see out to a certain distance. That might be the situation we are actually in. It is also possible that the universe is finite in both space and time.

The expanding universe

Now let's return to Einstein. After he published the final version of his equations in 1915, various mathematicians and physicists got to work on them. It was pointed out within a year that his equations suggested that the universe had to be dynamic (either expanding or collapsing). He asked astronomers about the evidence, and they told him that there was no evidence for either. As a result, in 1917 he introduced his famous "cosmological constant" into his equations, erroneously thinking that this would allow a static universe (it doesn't; it is like balancing a sharpened pencil on its tip, where any slight push will cause it to go one way or another). But what would better observations say?

As it turns out, evidence that would tip the scales had already begun to accumulate when Einstein put in the cosmological constant. In the 1800s people discovered that when sources of light move away from an observer, the light appears to shift to longer wavelengths. In visible light, the longer wavelengths are redder, hence this is called a "redshift" and for the same reason a source coming towards us is "blueshifted". It is, of course, easiest to see redshifts if you know the precise wavelength the source would have at rest, so the detection of lines with spectroscopes was necessary. As early as 1912, the remarkably-named Vesto Slipher detected such lines from galaxies, and over the next several years it became clear that, surprisingly, almost all galaxies were redshifted (the few exceptions are all close).

What could cause this? Are we so offensive that all other galaxies flee from us? The key addition to this information came from measurements of distances to those galaxies. This is a *lot* more difficult than you might at first think. The problem is that many methods of distance measurement are simply not applicable to other galaxies. Direct measurement (laying down a ruler) is clearly out of the question. Even parallax (farther things appear to shift less due to the Earth's orbit) can't be measured with nearly enough precision at those distances.

To the rescue came a discovery made by Henrietta Swan Leavitt, who worked on the Harvard College Observatories photographic plates. She noticed that a particular type of star called a Cepheid variable, which changes its brightness periodically, had the property that the longer it took to change its brightness, the brighter it was intrinsically. Therefore, if an observer could identify a Cepheid in a distant galaxy, they could compare its apparent brightness with its actual brightness and use that to get the distance to the galaxy. This is still a standard method today.

Using redshifts and distances and better telescopes, Edwin Hubble (after whom the telescope is named) managed to establish in 1929 that the apparent speed at which a galaxy goes away from us (thus producing its redshift) is simply proportional to its distance. This, as it turns out, is easily explained if the universe is expanding as a whole, just as Einstein's theory predicted but Einstein himself did not trust. Einstein called his lack of nerve and introduction of the cosmological constant the biggest blunder of his life, but now we think that something like a cosmological constant really does exist! Maybe some day I can make such a blunder...

Perspective: why did Einstein lose his nerve?

Why did Einstein not say, when the astronomers told him the universe seemed static, that his theory predicted that it was dynamic? He had shown equally strong confidence in his theories before. For example, there is an apocryphal story that after his light bending prediction was confirmed by Eddington, a reporter asked him how he would have felt if it had been disproved. His reaction, supposedly, was "I would have felt sorry for the good Lord. The theory is correct." (!!) I do not have a good answer for this. It could be that even for such a revolutionary as Einstein, the concept that the entire universe was expanding or contracting was too much for him, and he balked. We can't really blame him for that, and besides, his cosmological constant is at the forefront of much research today.

What is surprising is that he thought the cosmological constant would allow for a static universe. It is really not that difficult to show that such a universe would be unstable. Basically, a cosmological constant acts as something that pushes spacetime apart, and there is a constant amount of it per unit volume. It would act against gravity, which pulls things together. Therefore, to balance the two you would need a very precise amount of volume for a given amount of mass. More than that, and the cosmological constant would push more than gravity pulls. Less, and gravity would pull more than the cosmological constant pushes. Either way, once one starts to win, it wins more and more with time, just like in the case of Buridan's Ass. Even Einstein was fallible!

Black holes

The other featured prediction of Einstein's theory that he rejected is the star of this course: black holes. We encountered an early version of these when we discussed John Michell and his thoughts of "dark stars". The modern concept dates from 1916, when Karl Schwarzschild read Einstein's publication and, from a trench at the Russian front of World War I, derived a simplified solution to the Einstein equations. He investigated the situation outside a spherically symmetric star of mass M. At a distance r from the center of the star, the metric reads

$$ds^2 = -c^2(1 - 2GM/(rc^2))dt^2 + dr^2/(1 - 2GM/(rc^2)) + r^2(d\theta^2 + \sin^2 d\phi^2) \ . \qquad (1)$$

Here θ and ϕ are the normal spherical coordinates (the first is like a latitude, the second like a longitude), t is the time interval measured by someone at a very large distance, and r is actually the *circumferential* radius, that is, the circumference of a circle at that radius divided by 2π (!). Yes, in these situations you have to resort to convoluted definitions like that.

So what? It's a little complicated, sure, but you can use this to get verifiable predictions about light bending, gravitational redshifts, and the precession of elliptical orbits. In fact, for more than two decades no one worried too much about the weirder implications.

Eventually, though, and for reasons that we will discuss more fully in the next lecture, more and more people started to point out that when $r = 2GM/c^2$ (about 3 km for the Sun) this looked really bad. The coefficient of the dt^2 term goes to zero, and that of the dr^2 term goes to infinity! Yikes. On the other hand, maybe this is just because of what the coordinates do. Think about using latitude and longitude on the Earth, and traveling to the South Pole. When you get there, you find that all the lines of longitude converge, so that you can do silly tricks like "walking around the Earth" in a few steps :). It may be cold there, but no spacetime rips await you.

However the problem at $r = 2GM/c^2$ isn't quite that innocuous. Investigations by various scientists showed that if you were a long way away and dropped something in to an object with that radius, then the redshift would increase without bound. Recalling that a redshift means a longer wavelength and thus a lower frequency, this means that if you watched a clock drop towards that distance, it would slow down indefinitely. A consequence of this is that, in fact, you would never see anything cross the $r = 2GM/c^2$ line. For example, suppose a star collapsed; its collapse would initially appear rapid to you, but the collapse would slow down and it would never appear to go inside $r = 2GM/c^2$. Huh??? This radius is called the "event horizon" because events that happen at smaller radii are not visible to us, just as objects beyond the horizon are not.

To Einstein and most of his contemporaries, this meant that "Schwarzschild singularities", as they were called then instead of black holes, implied stars hovering just outside the $r = 2GM/c^2$ limit. But as Einstein showed, this was not possible. The two ways you could imagine matter hovering were (1) it had particles in rapid orbits, or (2) it was supported by internal pressure. But he demonstrated that any orbit with a radius smaller than $r = 3GM/c^2$ required orbital speeds faster than light, so that doesn't work. He also showed that anything more compact than $r = 9GM/4c^2$ could not be supported even by infinite internal pressure; the problem is that in general relativity all forms of energy (including pressure!) gravitate, so that the pressure that would support the matter adds to the gravity. His conclusions were published in a paper in which he indicated that this was a clear physical demonstration that Schwarzschild singularities do not exist.

This paper was more than three decades after Einstein had published his original equations. His calculations were correct, but we know now that his interpretation was not. He had indeed showed that nothing that compact *and outside a radius of* $2GM/c^2$ could exist, but he had not taken the intuitive leap to the possibility that the collapse is *not* stopped. Instead, once a star reaches a certain compactness, collapse is inevitable, and you have a black hole.

Why did he not have this realization? It may be in part for the same reason that he did not stand by the prediction of the expanding universe: the consequences were too strange for even Einstein. There is also an important difference, which is that although an expanding or contracting universe is *inevitable* given general relativity, you might imagine that black holes are prevented from forming due to various astrophysical effects. We'll explore this possibility in more detail next lecture.

However, a strong reason why Einstein and others did not believe in black holes is that the Schwarzschild metric was misinterpreted. What it says is that a *distant* observer would never see anything fall closer than $r = 2GM/c^2$. It is not as obvious what would happen from the perspective of the thing that is falling, but in 1958 David Finkelstein showed (using different coordinates) that if you fell into a black hole you would, from your perspective, definitely cross inside this barrier and to your squishy doom at the center. This new mathematical point of view (which does not contain anything physically new; you can derive it from the Schwarzschild metric as well) freed people's minds about black holes, but sadly it happened three years after Einstein died. Incidentally, to forestall a possible misconception, even as a distant observer you would *not* expect to see frozen surprised aliens just outside of black holes. The redshift goes to infinity, remember, so any light from things that fell in would be redshifted into undetectablity, and this would happen very fast (microseconds for black holes of a few times the Sun's mass, up to hours for the supermassive black holes at the centers of galaxies).

The door was thus open for black holes. But do they really exist, or does nature have preventive measures against them? To understand this, next time we will explore white dwarfs and neutron stars.