
Introduction

Cosmology is the study of the universe, or cosmos, regarded as a whole. Attempting to cover the study of the entire universe in a single volume may seem like a megalomaniac’s dream. The universe, after all, is richly textured, with structures on a vast range of scales; planets orbit stars, stars are collected into galaxies, galaxies are gravitationally bound into clusters, and even clusters of galaxies are found within larger superclusters. Given the complexity of the universe, the only way to condense its history into a single book is by a process of ruthless simplification. For much of this book, therefore, we will be considering the properties of an idealized, perfectly smooth, model universe. Only near the end of the book will we consider how relatively small objects, such as galaxies, clusters, and superclusters, are formed as the universe evolves. It is amusing to note in this context that the words *cosmology* and *cosmetology* come from the same Greek root: the word *kosmos*, meaning harmony or order. Just as cosmetologists try to make a human face more harmonious by smoothing over small blemishes such as pimples and wrinkles, cosmologists sometimes must smooth over small “blemishes” such as galaxies.

A science that regards entire galaxies as being small objects might seem, at first glance, very remote from the concerns of humanity. Nevertheless, cosmology deals with questions that are fundamental to the human condition. The questions that vex humanity are given in the title of a painting by Paul Gauguin (Figure 1.1): “Where do we come from? What are we? Where are we going?” Cosmology grapples with these questions by describing the past, explaining the present, and predicting the future of the universe. Cosmologists ask questions such as “What is the universe made of? Is it finite or infinite in spatial extent? Did it have a beginning some time in the past? Will it come to an end some time in the future?”

Cosmology deals with distances that are very large, objects that are very big, and timescales that are very long. Cosmologists frequently find that the standard SI units are not convenient for their purposes: the meter (m) is awkwardly

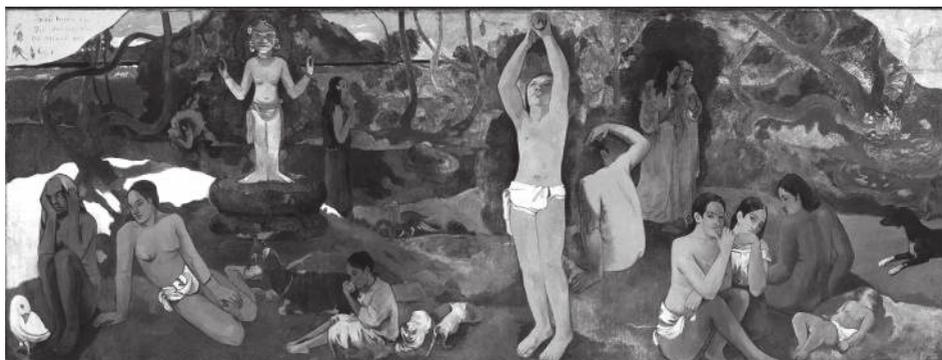


Figure 1.1 *Where Do We Come From? What Are We? Where Are We Going?* Paul Gauguin, 1897–98. [Museum of Fine Arts, Boston]

short, the kilogram (kg) is awkwardly tiny, and the second (s) is awkwardly brief. Fortunately, we can adopt the units that have been developed by astronomers for dealing with large distances, masses, and times.

One distance unit used by astronomers is the astronomical unit (AU), equal to the mean distance between the Earth and Sun; in metric units, $1 \text{ AU} = 1.50 \times 10^{11} \text{ m}$. Although the astronomical unit is a useful length scale within the solar system, it is small compared to the distances between stars. To measure interstellar distances, it is useful to use the parsec (pc), equal to the distance at which 1 AU subtends an angle of 1 arcsecond; in metric units, $1 \text{ pc} = 3.09 \times 10^{16} \text{ m}$. For example, we are at a distance of 1.30 pc from Proxima Centauri (a small, relatively cool star that is the Sun's nearest neighboring star); we are at a distance of 8500 pc from the center of our galaxy, the Milky Way Galaxy. Although the parsec is a useful length scale within our galaxy, it is small compared to the distances between galaxies. To measure intergalactic distances, we use the megaparsec (Mpc), equal to 10^6 pc , or $3.09 \times 10^{22} \text{ m}$. For example, we are at a distance of 0.76 Mpc from M31 (otherwise known as the Andromeda galaxy) and 15 Mpc from the Virgo cluster (the nearest big cluster of galaxies).

The standard unit of mass used by astronomers is the solar mass (M_{\odot}); in metric units, the Sun's mass is $1 M_{\odot} = 1.99 \times 10^{30} \text{ kg}$. The total mass of our galaxy is not known as accurately as the mass of the Sun; in round numbers, though, it is $M_{\text{gal}} \sim 10^{12} M_{\odot}$. The Sun, incidentally, also provides the standard unit of power used in astronomy. The Sun's luminosity (that is, the rate at which it radiates away energy in the form of light) is $1 L_{\odot} = 3.83 \times 10^{26} \text{ watts}$. The total luminosity of our galaxy is not known as accurately as the luminosity of the Sun; a good estimate, though, is $L_{\text{gal}} \approx 3 \times 10^{10} L_{\odot}$.

For times much longer than a second, it is convenient to use the year (yr) as a unit of time, with $1 \text{ yr} \approx 3.16 \times 10^7 \text{ s}$. In a cosmological context, a year is frequently an inconveniently short period of time, so cosmologists often use

megayears (Myr), with $1 \text{ Myr} = 10^6 \text{ yr} = 3.16 \times 10^{13} \text{ s}$. Even longer timescales call for use of gigayears (Gyr), with $1 \text{ Gyr} = 10^9 \text{ yr} = 3.16 \times 10^{16} \text{ s}$. For example, the age of the Earth is more conveniently written as 4.57 Gyr than as $1.44 \times 10^{17} \text{ s}$.

In addition to dealing with very large things, cosmology also deals with very small things. Early in its history, as we shall see, the universe was very hot and dense, and some interesting particle physics phenomena were occurring. Consequently, particle physicists have plunged into cosmology, introducing some terminology and units of their own. For instance, particle physicists tend to measure energy units in electron volts (eV) instead of joules (J). The conversion factor between electron volts and joules is $1 \text{ eV} = 1.60 \times 10^{-19} \text{ J}$. The rest energy of an electron, for instance, is $m_e c^2 = 511\,000 \text{ eV} = 0.511 \text{ MeV}$, and the rest energy of a proton is $m_p c^2 = 938.27 \text{ MeV} = 1836.1 m_e c^2$.

When you stop to think of it, you realize that the units of meters, megaparsecs, kilograms, solar masses, seconds, and gigayears could only be devised by ten-fingered Earthlings obsessed with the properties of water. An eighteen-tentacled silicon-based lifeform from a planet orbiting Betelgeuse would probably devise a different set of units. A more universal, less culturally biased system of units is the Planck system, based on the universal constants G , c , and \hbar . Combining the Newtonian gravitational constant, $G = 6.67 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$, the speed of light, $c = 3.00 \times 10^8 \text{ m s}^{-1}$, and the reduced Planck constant, $\hbar = h/(2\pi) = 1.05 \times 10^{-34} \text{ J s} = 6.58 \times 10^{-16} \text{ eV s}$, yields a unique length scale, known as the Planck length:

$$\ell_P \equiv \left(\frac{G\hbar}{c^3} \right)^{1/2} = 1.62 \times 10^{-35} \text{ m}. \quad (1.1)$$

The same constants can be combined to yield the Planck mass,¹

$$M_P \equiv \left(\frac{\hbar c}{G} \right)^{1/2} = 2.18 \times 10^{-8} \text{ kg}, \quad (1.2)$$

and the Planck time,

$$t_P \equiv \left(\frac{G\hbar}{c^5} \right)^{1/2} = 5.39 \times 10^{-44} \text{ s}. \quad (1.3)$$

Using Einstein's relation between mass and energy, we can also define the Planck energy,

$$E_P = M_P c^2 = 1.96 \times 10^9 \text{ J} = 1.22 \times 10^{28} \text{ eV}. \quad (1.4)$$

By bringing the Boltzmann constant, $k = 8.62 \times 10^{-5} \text{ eV K}^{-1}$, into the act, we can also define the Planck temperature,

$$T_P = E_P/k = 1.42 \times 10^{32} \text{ K}. \quad (1.5)$$

¹ The Planck mass is roughly equal to the mass of a grain of sand a quarter of a millimeter across.

When distance, mass, time, and temperature are measured in the appropriate Planck units, then $c = k = \hbar = G = 1$. This is convenient for individuals who have difficulty in remembering the numerical values of physical constants. However, using Planck units can have potentially confusing side effects. For instance, many cosmology texts, after noting that $c = k = \hbar = G = 1$ when Planck units are used, then proceed to omit c , k , \hbar , and/or G from all equations. For instance, Einstein's celebrated equation, $E = mc^2$, becomes $E = m$. The blatant dimensional incorrectness of such an equation is jarring, but it simply means that the rest energy of an object, measured in units of the Planck energy, is equal to its mass, measured in units of the Planck mass. In this book, however, I will retain all factors of c , k , \hbar , and G , for the sake of clarity.

Here we will deal with distances ranging from the Planck length to 10^4 Mpc or so, a span of some 61 orders of magnitude. Dealing with such a wide range of length scales requires a stretch of the imagination, to be sure. However, cosmologists are not permitted to let their imaginations run totally unfettered. Cosmology, I emphasize strongly, is based ultimately on observation of the universe around us. Even in ancient times, cosmology was based on observations; unfortunately, those observations were frequently imperfect and incomplete. Ancient Egyptians, for instance, looked at the desert plains stretching away from the Nile valley and the blue sky overhead. Based on their observations, they developed a model of the universe in which a flat Earth (symbolized by the earth god Geb in Figure 1.2) was covered by a solid dome (symbolized by the sky goddess Nut). Underneath the sky dome, the disk of the Sun was carried from east to west by the sun god Ra. Greek cosmology was based on more precise and sophisticated observations. Ancient Greek astronomers deduced, from their observations, that the Earth and

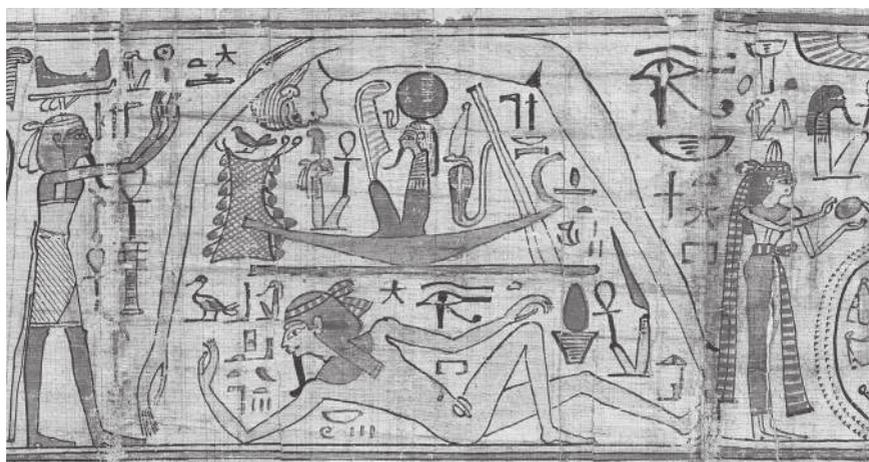


Figure 1.2 The ancient Egyptian view of the cosmos: the sky goddess Nut arches over the earth god Geb, while the sun god Ra travels between them in a reed boat. (Book of the Dead of Nespakashuty, ca. 1000 BC) [Musée du Louvre, Paris]

Moon are spherical, that the Sun is much farther from the Earth than the Moon is, and that the distance from the Earth to the stars is much greater than the Earth's diameter. Based on this knowledge, Greek cosmologists devised a “two-sphere” model of the universe, in which the spherical Earth is surrounded by a much larger celestial sphere, a spherical shell to which the stars are attached. Between the Earth and the celestial sphere, in this model, the Sun, Moon, and planets move on their complicated apparatus of epicycles and deferents.

Although cosmology is ultimately based on observation, sometimes observations temporarily lag behind theory. During periods when data are lacking, cosmologists may adopt a new model for aesthetic or philosophical reasons. For instance, when Copernicus proposed a new Sun-centered model of the universe, to replace the Earth-centered two-sphere model of the Greeks, he didn't base his model on new observational discoveries. Rather, he believed that putting the Earth in motion around the Sun resulted in a conceptually simpler, more appealing model of the universe. Direct observational evidence didn't reveal that the Earth revolves around the Sun, rather than vice versa, until the discovery of the aberration of starlight in the year 1728, nearly two centuries after the death of Copernicus. Foucault didn't demonstrate the rotation of the Earth, another prediction of the Copernican model, until 1851, over *three* centuries after the death of Copernicus. However, although observations sometimes lag behind theory in this way, every cosmological model that isn't eventually supported by observational evidence must remain pure speculation.

The current standard model for the universe is the “Hot Big Bang” model, which states that the universe has expanded from an initially hot and dense state to its current relatively cool and tenuous state, and that the expansion is still going on today. To see why cosmologists have embraced the Hot Big Bang model, let us turn, in the next chapter, to the fundamental observations on which modern cosmology is based.

2

Fundamental Observations

Some of the observations on which modern cosmology is based are highly complex, requiring elaborate apparatus and sophisticated data analysis. However, other observations are surprisingly simple. Let's start with an observation that is deceptive in its extreme simplicity.

2.1 The Night Sky is Dark

Step outside on a clear, moonless night, far from city lights, and look upward. You will see a dark sky, with roughly two thousand stars scattered across it. The fact that the night sky is dark at visible wavelengths, instead of being uniformly bright with starlight, is known as *Olbers' paradox*, after the astronomer Heinrich Olbers, who wrote a scientific paper on the subject in 1823. As it happens, Olbers was not the first person to think about Olbers' paradox. As early as 1576, Thomas Digges mentioned how strange it is that the night sky is dark, with only a few pinpoints of light to mark the location of stars.¹

Why should it be paradoxical that the night sky is dark? Most of us simply take for granted the fact that daytime is bright and nighttime is dark. The darkness of the night sky certainly posed no problems to the ancient Egyptians or Greeks, to whom stars were lights stuck to a dome or sphere. However, the cosmological model of Copernicus required that the distance to stars be very much larger than an astronomical unit; otherwise, the parallax of the stars, as the Earth goes around on its orbit, would be large enough to see with the naked eye. Moreover, since the Copernican system no longer requires that the stars be attached to a rotating celestial sphere, the stars can be at different distances from the Sun. These

¹ The name "Olbers' paradox" is thus a prime example of what historians of science jokingly call the law of misonomy: nothing is ever named after the person who really discovers it. The law of misonomy is also known as "Stigler's law," after a statistician who admits that he (of course!) didn't discover it.

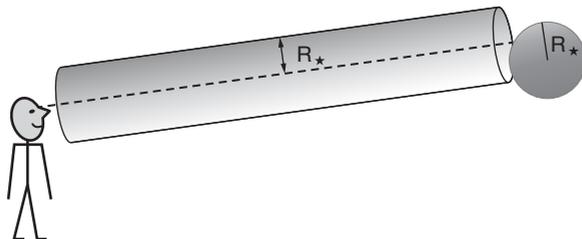


Figure 2.1 A line of sight through the universe eventually encounters an opaque star.

liberating realizations led Thomas Digges, and other post-Copernican astronomers, to embrace a model in which stars are large, opaque, glowing spheres like the Sun, scattered throughout infinite space.

Let's compute how bright we expect the night sky to be in an infinite universe. Let n_* be the number density of stars in the universe; averaged over large scales, this number is $n_* \sim 10^9 \text{ Mpc}^{-3}$. Let R_* be the typical radius of a star. Although stars have a range of sizes, from dwarfs to supergiants, we may adopt the Sun as a typical mid-sized star, with $R_* \sim R_\odot = 7.0 \times 10^8 \text{ m} = 2.3 \times 10^{-14} \text{ Mpc}$. Consider looking outward in some direction through the universe. If you draw a cylinder of radius R_* around your line of sight, as shown in Figure 2.1, then if a star's center lies within that cylinder, the opaque star will block your view of more distant objects. If the cylinder's length is λ , then its volume is $V = \lambda\pi R_*^2$, and the average number of stars that have their centers inside the cylinder is

$$N = n_* V = n_* \lambda \pi R_*^2. \quad (2.1)$$

Since it requires only one star to block your view, the typical distance you will be able to see before a star blocks your line of sight is the distance λ for which $N = 1$. From Equation (2.1), this distance is

$$\lambda = \frac{1}{n_* \pi R_*^2}. \quad (2.2)$$

For concreteness, if we take $n_* \sim 10^9 \text{ Mpc}^{-3}$ and $\pi R_*^2 \sim \pi R_\odot^2 \sim 10^{-27} \text{ Mpc}^2$, then you can see a distance

$$\lambda \sim \frac{1}{(10^9 \text{ Mpc}^{-3})(10^{-27} \text{ Mpc}^2)} \sim 10^{18} \text{ Mpc} \quad (2.3)$$

before your line of sight intercepts a star. This is a very large distance; but it is a *finite* distance. We therefore conclude that in an infinite universe (or one that stretches at least 10^{18} Mpc in all directions), the sky will be completely paved with stars.

What does this paving imply for the brightness of the sky? If a star of radius R_* is at a distance $r \gg R_*$, its angular area, in steradians, will be

$$\Omega = \frac{\pi R_*^2}{r^2}. \quad (2.4)$$

If the star's luminosity is L_* , then its flux measured at a distance r will be

$$f = \frac{L_*}{4\pi r^2}. \quad (2.5)$$

The surface brightness of the star, in watts per square meter of your pupil (or telescope mirror) per steradian, will then be

$$\Sigma_* = \frac{f}{\Omega} = \frac{L_*}{4\pi^2 R_*^2}, \quad (2.6)$$

independent of the distance to the star. Thus, the surface brightness of a sky paved with stars will be equal to the (distance-independent) surface brightness of an individual star. We therefore conclude that in an infinite universe (or one that stretches at least 10^{18} Mpc in all directions), the entire sky, night and day, should be as dazzlingly bright as the Sun's disk.

This is utter nonsense. The surface brightness of the Sun is $\Sigma_{\odot} \approx 5 \times 10^{-4}$ watts m^{-2} arcsec $^{-2}$. By contrast, the surface brightness of the dark night sky is $\Sigma \sim 5 \times 10^{-17}$ watts m^{-2} arcsec $^{-2}$. Thus, my estimate of the surface brightness of the night sky ("It's the same as the Sun's") is wrong by a factor of 10 trillion.

One (or more) of the assumptions that went into my estimate of the sky brightness must be wrong. Let's scrutinize some of the assumptions. One assumption that I made is that space is transparent over distances of 10^{18} Mpc. This might not be true. Heinrich Olbers himself tried to resolve Olbers' paradox by proposing that distant stars are hidden from view by interstellar matter that absorbs starlight. This resolution does not work in the long run, because the interstellar matter is heated by starlight until it has the same temperature as the surface of a star. At that point, the interstellar matter emits as much light as it absorbs, and glows as brightly as the stars themselves.

A second assumption that I made is that the universe is infinitely large. This might not be true. If the universe extends to a maximum distance $r_{\text{max}} \ll \lambda$, then only a fraction $F \sim r_{\text{max}}/\lambda$ of the night sky will be covered with stars. This result will also be found if the universe is infinitely large, but is devoid of stars beyond a distance r_{max} .

A third assumption, slightly more subtle than the previous ones, is that the universe is infinitely old. This might not be true. Because the speed of light is finite, when we look farther out in space, we are looking farther out in time. Thus, we see the Sun as it was 8.3 minutes ago, Proxima Centauri as it was 4.2 years ago, and M31 as it was 2.5 million years ago. If the universe has a finite age, $t_0 \ll \lambda/c$, then we are not yet able to see stars at a distance greater than $r \sim ct_0$,

2.2 The Universe is Isotropic and Homogeneous

9

and only a fraction $F \sim ct_0/\lambda$ of the night sky will be covered with stars. This result will also be found if the universe is infinitely old, but has only contained stars for a finite time t_0 .

A fourth assumption is that the surface brightness of a star is independent of distance, as derived in Equation 2.6. This might not be true. The assumption of constant surface brightness would have seemed totally innocuous to Olbers and other nineteenth-century astronomers, who assumed that the universe was static. However, in an expanding universe, the surface brightness of distant light sources is decreased relative to what you would see in a static universe. (In a contracting universe, the surface brightness would be increased, which would only make the problem of a bright night sky even worse.)

Thus, the infinitely large, eternally old, static universe that Thomas Digges and his successors pictured simply does not hold up to scrutiny. This is a textbook, not a suspense novel, so I'll tell you right now: the primary resolution to Olbers' paradox comes from the fact that the universe has a finite age. The stars beyond some finite distance, called the horizon distance, are invisible to us because their light hasn't had time to reach us yet. A particularly amusing bit of cosmological trivia is that the first person to hint at the correct resolution of Olbers' paradox was Edgar Allan Poe.² In his essay "Eureka: A Prose Poem," completed in 1848, Poe wrote, "Were the succession of stars endless, then the background of the sky would present us an [*sic*] uniform density... since there could be absolutely no point, in all that background, at which would not exist a star. The only mode, therefore, in which, under such a state of affairs, we could comprehend the voids which our telescopes find in innumerable directions, would be by supposing the distance of the invisible background so immense that no ray from it has yet been able to reach us at all."

2.2 The Universe is Isotropic and Homogeneous

What does it mean to state that the universe is isotropic and homogeneous? Saying that the universe is *isotropic* means that there are no preferred directions in the universe; it looks the same no matter which way you point your telescope. Saying that the universe is *homogeneous* means that there are no preferred locations in the universe; it looks the same no matter where you set up your telescope. Note the very important qualifier: the universe is isotropic and homogeneous *on large scales*. In this context, "large scales" means that the universe is only isotropic and homogeneous on scales of roughly 100 Mpc or more.

² That's right, the "Nevermore" guy. Poe was an excellent student at the University of Virginia (before he fell into debt and withdrew). He was then an excellent student at West Point (before he was court-martialed and expelled).

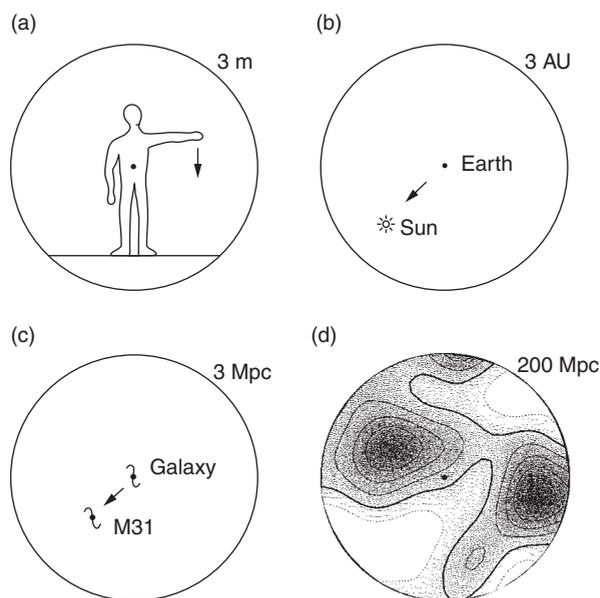


Figure 2.2 (a) A sphere 3 m in diameter, centered on your navel. (b) A sphere 3 AU in diameter, centered on your navel. (c) A sphere 3 Mpc in diameter, centered on your navel. (d) A sphere 200 Mpc in diameter, centered on your navel. Shown is the smoothed number density of galaxies. The heavy contour is drawn at the mean density; darker regions represent higher density. [Dekel *et al.* 1999, *ApJ*, **522**, 1]

The isotropy of the universe is not immediately obvious. In fact, on small scales, the universe is blatantly anisotropic. Consider, for example, a sphere 3 m in diameter, centered on your navel (Figure 2.2a). Within this sphere, there is a preferred direction; it is the direction commonly referred to as “down.” It is easy to determine the vector pointing down. Just let go of a small dense object. The object doesn’t hover in midair, and it doesn’t move in a random direction; it falls down, toward the center of the Earth.

On significantly larger scales, the universe is still anisotropic. Consider, for example, a sphere 3 AU in diameter, centered on your navel (Figure 2.2b). Within this sphere, there is a preferred direction; it is the direction pointing toward the Sun, which is by far the most massive and most luminous object within the sphere. It is easy to determine the vector pointing toward the Sun. Just step outside on a sunny day, and point to that really bright disk of light up in the sky.

On still larger scales, the universe is *still* anisotropic. Consider, for example, a sphere 3 Mpc in diameter, centered on your navel (Figure 2.2c). This sphere contains the Local Group of galaxies, a small cluster of about a hundred galaxies. By far the most massive and most luminous galaxies in the Local Group are our own galaxy and M31, which together contribute about 86 percent of the total luminosity within the 3 Mpc sphere. Thus, within this sphere, our galaxy and