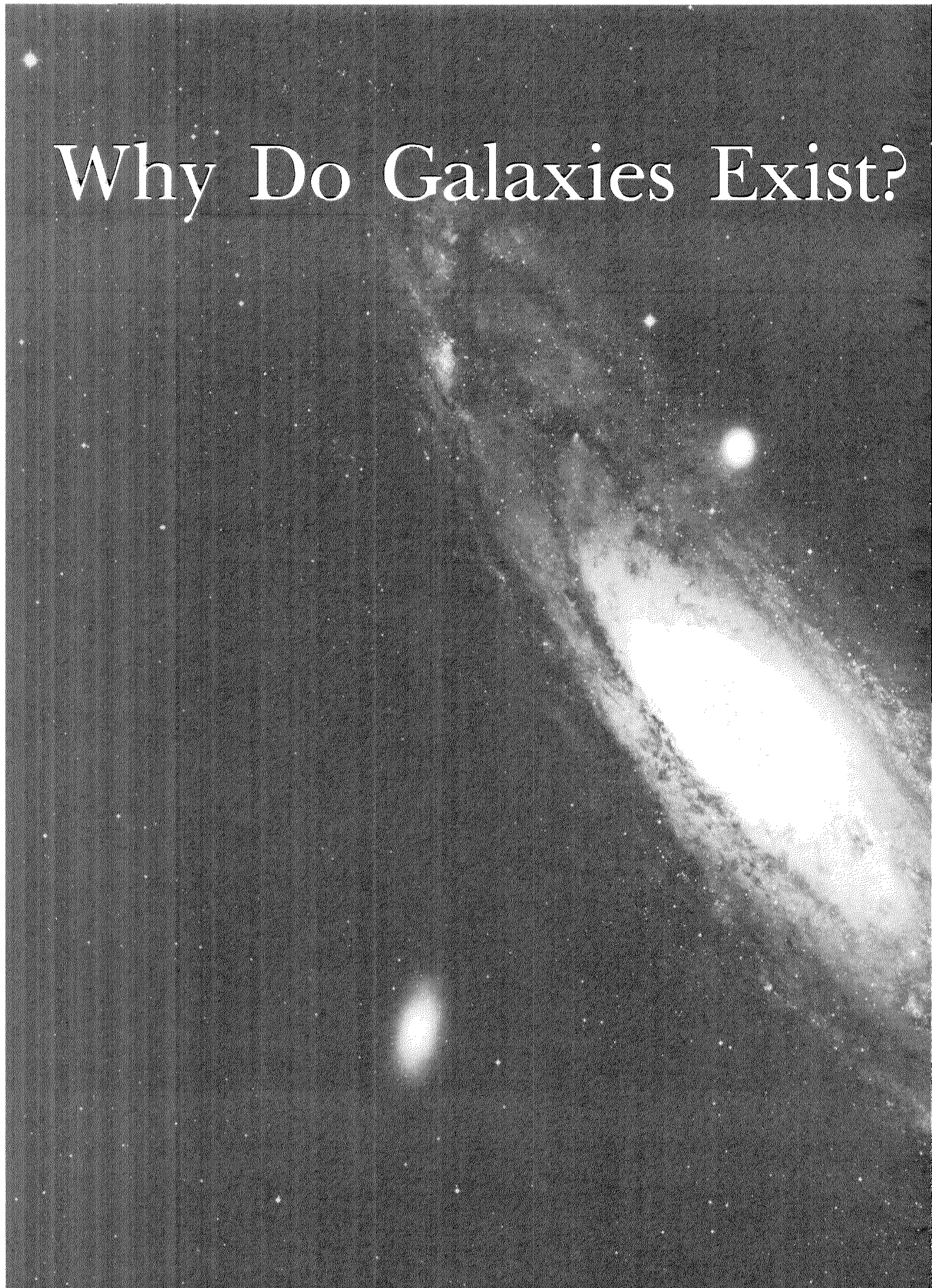
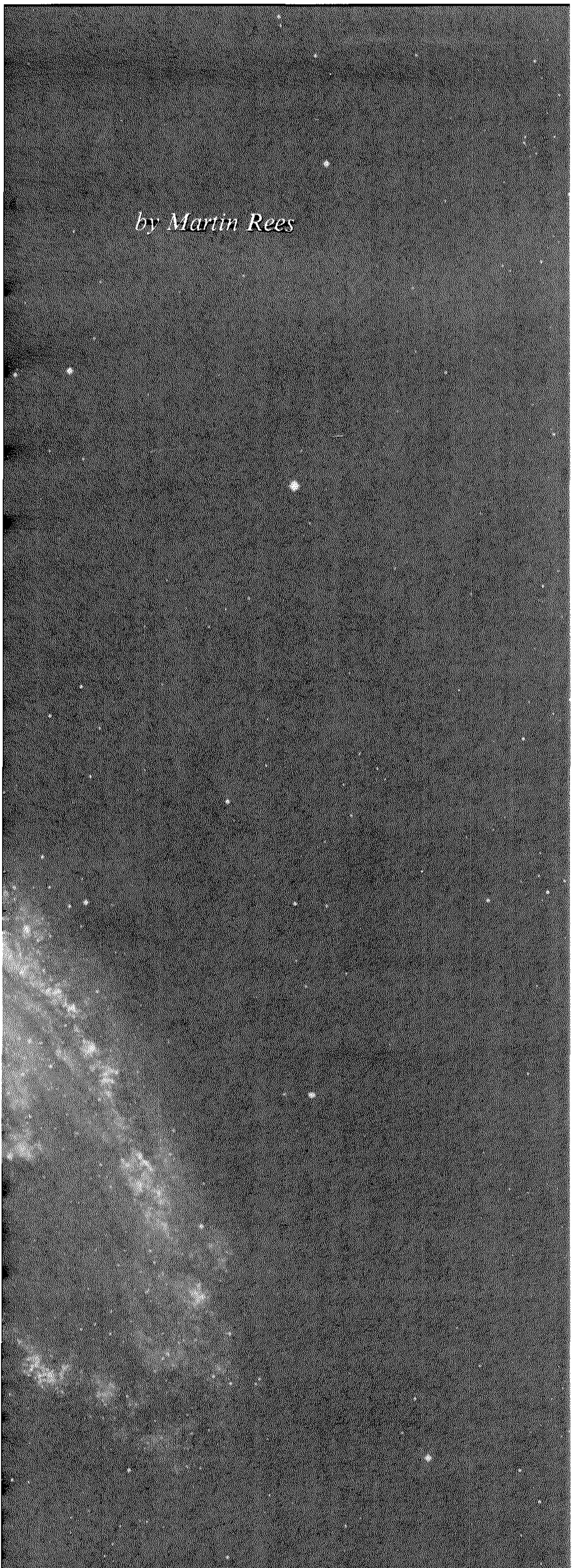
Why Do Galaxies Exist?



The great spiral galaxy in Andromeda (M31), the closest galaxy to our own, is flanked by two elliptical galaxies—the fuzzy spots above (M32) and below (NGC 205).

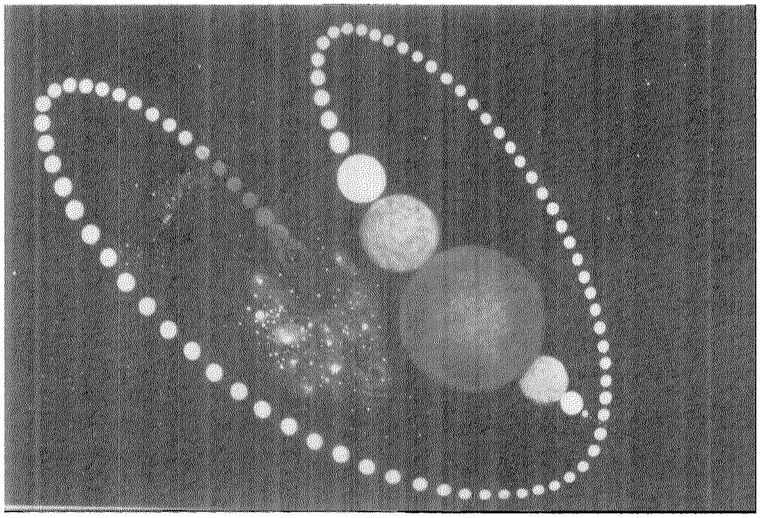


THE UNIVERSE HAS NO CENTER, but the universe of astronomers does. Ever since Hubble's time that center has been here in Pasadena, and it's therefore daunting, although an honor, to be invited from the periphery to speak to an audience here. I am further abashed because the honest answer I must give to the question of why galaxies exist is: I don't know. Nobody knows. But I'll try to describe why this isn't a presumptuous question to pose and summarize what current research programs tell us about what galaxies are made of, how they evolve and what may happen to them, and, more speculatively, what special features of the physical universe are necessary for their emergence.

In doing this I'll try to illustrate three of the intrinsic motives for doing astronomy. The first is just discovery—to find out what's there, be it in the solar system or in the remotest extragalactic realm. This vicarious exploration is something that a wide public can share. But to the astrophysicist it's preliminary to the second motive of understanding and interpreting what we see and setting our earth and our solar system in an evolutionary scheme traceable right back to the so-called Big Bang from which our entire universe emerged. Physicists have a third motive: the cosmos allows us to study how material behaves under far more extreme conditions than we can simulate terrestrially, and thereby to test the laws of nature to their limits and perhaps even find new ones.

Let's start, though, with something that's fairly well understood. The life cycle of a star like our sun begins as the sun condenses by gravitation from an interstellar cloud. It then contracts until its center gets hot enough to ignite nuclear reactions; fusion of hydrogen to helium then releases enough energy to keep the sun burning steadily. It's been going for about 4.5 billion years and has about another 5 billion to go before the hydrogen in its core is used up. It will then swell up to become a red giant, engulfing the inner planets, and ultimately settle down to a quiet demise as a white dwarf. Most stars we see are evolving in this way. Stars are so long-lived compared to astronomers that we only have in effect a single snapshot of each. But the fact that there are so many of them makes up for this, and we can check our theories, just as you can infer the life-cycle of a tree by one day's observation in a forest.

But not everything in the cosmos happens all that slowly. Stars heavier than the sun



The life cycle of a star like our sun is illustrated here as a series of time-lapse pictures with 100 million years between successive frames. When the hydrogen in the core is used up, it will swell into a red giant and then die away quietly as a white dwarf.

evolve faster, and some expire violently as supernovae. Supernova explosions signify the violent end point of stellar evolution, when a star too massive to become a white dwarf exhausts its nuclear fuel and then faces an energy crisis. Its core implodes, releasing so much energy gravitationally that the outer layers are blown off. Nearby supernovae are rare, and the astronomical event of 1987, which rated an eight-page cover story in *Time* magazine, was a supernova in the southern sky, the nearest and brightest by far of modern times. Its evolution is at the moment being closely monitored by all observational techniques.

Supernovae, even the nearest ones, may seem remote and irrelevant to our origins. But, on the contrary, it's only by studying the births of stars and the explosive way some of them die, that we can tackle such an everyday question as where the atoms we are made of came from. The respective abundances of the chemical elements can be measured in the solar system and inferred spectroscopically in stars and nebulae. And the proportions in which the elements occur display regularities from place to place—a fact that demands some explanation. Complex chemical elements are an inevitable by-product of the nuclear reactions that provide the power in stars. In fact, a massive star develops a kind of onion structure, where the inner, hotter shells are "cooked" further up the periodic table. The final explosion then ejects most of this processed material. All the carbon, nitrogen, oxygen, and iron on the earth could have been manufactured in stars that exhausted their fuel and exploded before the sun formed. The solar system could have condensed from gas contaminated by this ejected debris, and this process can account for the observed proportions of the different elements—why oxygen and iron are common, but gold and uranium are rare—and how they came to be in the solar system.

This concept of stellar nucleosynthesis originated with Sir Fred Hoyle and Willy Fowler (Nobel laureate and Institute Professor of Physics, Emeritus). Its detailed development is one of the outstanding scientific triumphs of the last 40 years. The work was spearheaded here in the Kellogg Lab, so it's perhaps appropriate to celebrate it on an occasion dedicated to the Lauritsens. This idea sets our solar system in a kind of ecological perspective involving the entire Milky Way Galaxy. The mix of elements we see around us isn't ad hoc but the outcome of transmutation and recycling processes, whose starting point is a young galaxy containing just the lightest elements. Each atom on earth can be traced back to stars that died before the solar system formed. Imagine a carbon atom, forged in the core of a massive star and ejected when it explodes as a supernova. This atom may spend hundreds of millions of years wandering in interstellar space before finding itself in a dense cloud that contracts into a new generation of stars. Then once again it could be in a stellar interior, or it could be out on the boundary of a new solar system in a planet, and maybe eventually in a human cell. As Willy Fowler likes to remind us, we are quite literally the ashes of these long-dead stars.

Theoretical studies of stars and their life cycles were stimulated by the challenge of observations. It's interesting that the properties of stars could have been deduced by a physicist who lived on a perpetually cloud-bound planet—or indeed by an *English* astronomer. He could have posed the question: Can one have a gravitationally bound fusion reactor, and what would it be like? And he'd reason as follows.

Gravity is extremely feeble on the atomic scale. In a hydrogen molecule, for instance, consisting of two protons neutralized by two electrons, the gravitational binding energy between the protons is feebler than the electrical energy by a factor of 10^{36}. But any macroscopic object—an asteroid or a lump of rock—contains almost equal numbers of positive and negative charges, so that the electrical forces tend to cancel out. But there's no such cancellation of gravity. Everything has the same gravitational charge and attracts everything else, so gravitation becomes more significant for larger objects. Gravitational binding energy rises as mass divided by radius, and that means it rises a hundredfold for each thousandfold increase in mass. So gravity wins out over electrical energy on sufficiently large scales.

How large? Imagine that we were to assemble a set of bodies containing successively 10, 100, 1,000 atoms, and so on. The 24th of these would be the size of a sugar lump—about one cubic centimeter. The 39th would be like a kilometer of rock. Gravity starts off with a handicap of 10^{36}; it gains as a two-thirds power, and when we get to our 54th object (54 being three halves of 36), gravity will have caught up with electrical energy. We then have an object the mass of Jupiter, and anything bigger than Jupiter will start getting crushed by gravity. To be squeezed by gravity and heated to the point where nuclear fusion could ignite, a body would have to be well over 10^{54} atoms.

So gravitationally bound fusion reactors must be massive because gravity is so weak. And having inferred this, the physicist could in principle calculate the entire life cycle of stars. In fact, Sir Arthur Eddington was the first person to express this argument clearly. He then said that "when we draw aside the veil of clouds beneath which our physicist is working and let him look up at the sky, there he will find a thousand million globes of gas, nearly all with masses [in this calculated range]."

Let's now enlarge our horizons to the extragalactic realm. It's been clear since the 1920s that our Milky Way is just one galaxy similar to millions of others visible to large telescopes. Galaxies are held in equilibrium by a balance between gravity, which tends to make the stars hold together, and the countervailing effect of the stellar motions, which, if gravity didn't act, would make the galaxies fly apart. In some galaxies, our own among them, the stars move in nearly circular orbits in giant disks. In others, the less photogenic ellipticals, we see stars swarming around in more random directions, each feeling the integrated gravitational pull of all the others.

Galaxies are the most conspicuous features of the large-scale cosmic scene. Self-gravitating assemblages tens of thousands of light years across, they typically contain about a hundred billion stars. Unfortunately we don't yet have an accepted explanation of what's special about their dimensions in the same sense that we do for stars. But there is a scenario that accounts qualitatively for why there are two basic types of galaxy—disks and ellipticals. Let's suppose that a galaxy starts life as a huge, turbulent, clumpy, slowly spinning gas cloud, contracting under its own gravitation and gradually fragmenting into stars. The collapse of such a gas cloud is highly dissipative in the sense that any two of the clumps that collide will radiate their relative energy by shock waves and will merge. The end result of the collapse of a rotating gas cloud will be a disk—the lowest energy state it can get to if it conserves its angular momentum.

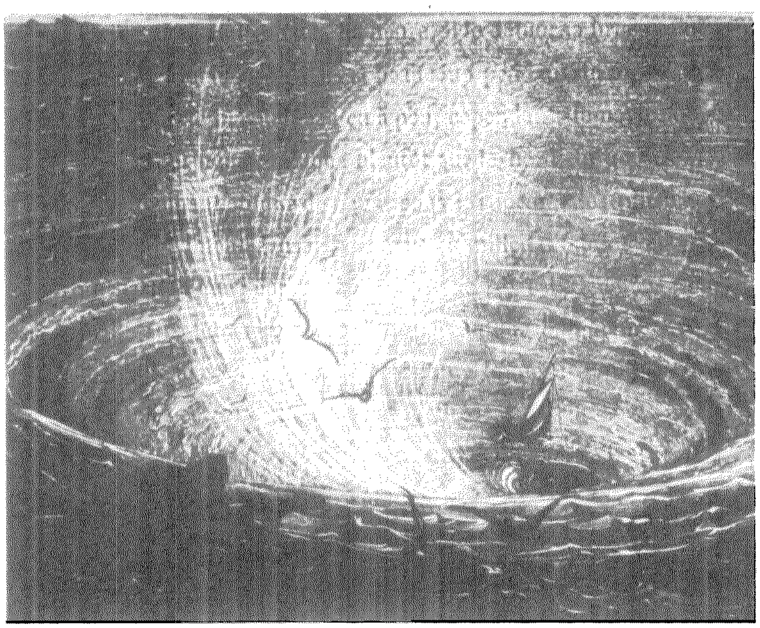
Stars, on the other hand, don't collide with each other, and can't dissipate energy in the same way as gas clouds. This suggests that the rate of condensation of gas into stars is the crucial feature determining the type of galaxy that results. Ellipticals will be those in which the conversion is fast, so that most of the stars have already formed before the gas has a chance to form a disk. The disk galaxies are those of slower metabolism, where star formation is delayed until the gas has settled into a disk. The origin of these giant gas clouds is a mystery—a cosmological question. But given these clouds, the physics determines that galactic morphology is nothing more exotic than Newtonian gravity and gas dynamics.

Some peculiar galaxies, though, which harbor intense superstellar activity in their centers, are much more than just a pile of stars. The most extreme are the so-called quasars, in which a small region no bigger than the solar system outshines the entire surrounding galaxy. In these objects the central power output exceeds a million supernova explosions going off in unison. It seems that gas and stars have accumulated in the center until some kind of runaway catastrophic collapse occurs. Gravity overwhelms all other forces, and a black hole forms. Here we do need somewhat more highbrow physics to know what's going on, in particular Einstein's theory of general relativity—that matter tells space how to curve and space tells matter how to move. Indeed, ever since such active galaxies were discovered, relativity specialists have been (in the words of Cornell cosmologist Tommy Gold) "not merely magnificent cultural ornaments, but actually relevant to astrophysics."

The rate of research progress over the 25 years since the phenomenon of active galactic nuclei was first recognized seems disappointingly slow. Sometimes we've had the illusion that it's rapid, but it's really been a rather slow advance with "saw-tooth" fluctuations imposed on it as fashions have come and gone. But there is now a fair consensus that the central prime mover in active galaxies involves a spinning black hole, as massive as perhaps a hundred million suns, fueled by capturing gas or even entire stars from its surroundings. This captured debris swirls in a flow resembling a whirlpool down into a central hole, carrying magnetic fields with it and moving at nearly the speed of light. At least 10 percent of the rest-mass energy of the infalling material could be radiated, and still more power could be extracted from the hole's spin. Some of us are hopeful that these ideas can be put on a firm basis just as our theories of stellar evolution have been, but we still have a long way to go.

If we can do this, we would have an opportunity to learn from a safe distance whether black holes really behave as theory predicts. In the vicinity of black holes space and time are thought to behave in highly nonintuitive ways. Time would stand still for an observer who managed to hover or orbit just outside a hole, and that observer could see the entire future of the universe in what was to him quite a short time. Even stranger things might happen inside the hole, but anyone who ventures there is trapped. So keep your distance unless your Faustian urge is overwhelmingly strong. No one inside the hole can transmit signals, so you would learn nothing by sending a student, even if you had an expendable one. If you really want to explore a black hole, then you should at least pick one of the monster ones in galactic nuclei. They are as large as an entire solar system and so, having passed inside, you'd at least have several hours for leisured if not relaxed observation before being discomforted by tidal forces and crushed in the central singularity.

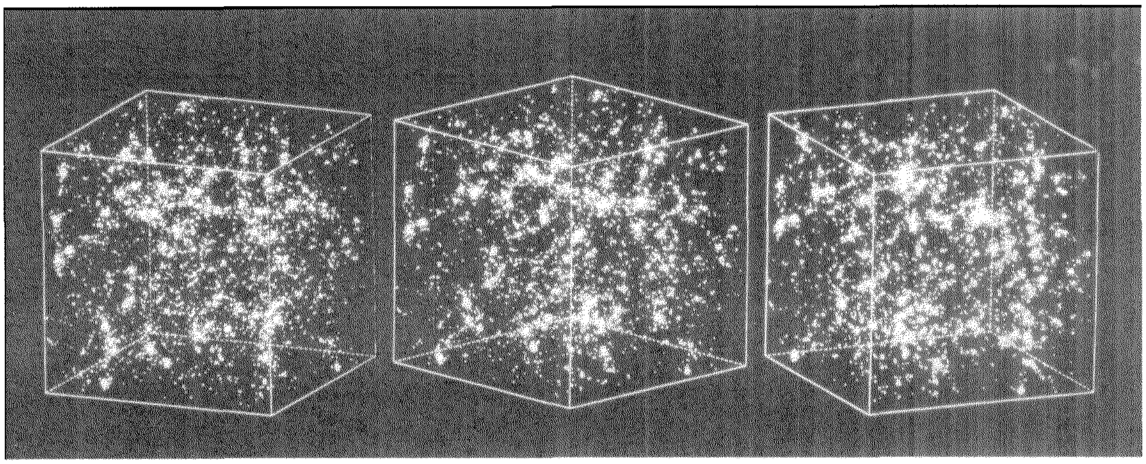
The flow swirling into a black hole resembles a whirlpool like this illustration to Edgar Allen Poe's "Descent into the Maelstrom." As scientists grope for the right theory for galactic nuclei, they also have only a crude cartoon of what conditions are really like.



Active galaxies are a bafflingly varied zoo, and we need more data on lots of samples to clarify their taxonomy. Violent central activity, whether it's a quasar or a strong radio source, is thought to be a relatively brief phase in a typical galaxy's life history. Dead quasars, massive black holes now starved of fuel, may lurk in the nuclei of many nearby galaxies. Recently the inner parts of some of these, including the Andromeda galaxy, have been studied accurately enough to infer that the stars very near the center are orbiting around a dark compact object, which at least answers the description of a black hole. As such it would be almost quiescent, but not quite. Now and again a star would wander so close to it that the tidal forces would rip the star apart. You would then see a flare persisting for as long as it took the debris from the star to be swallowed or expelled from the hole.

Let's turn now to cosmology, the description of our universe as a single dynamical entity. Cosmology is the study of a unique object and a unique event, by definition. No physicist would happily base his theory on a single unrepeatable experiment, and no biologist would formulate theories of animal behavior by looking at just one rat. But we can't check our cosmological ideas by applying them to other universes. Despite having all these things stacked against it though, scientific cosmology has proved possible because the observed universe in its large-scale structure turns out to be simpler than we had any right to expect.

It's natural to start off by making simplifying assumptions about symmetry, etc. And cosmologists did this. But what is surprising is that these models remain relevant and the simplifying assumptions have been vindicated. The intergalactic scales of distance are vast. To the cosmologist even entire galaxies are just markers or test particles scattered through space, which indicate how the material content of the universe is distributed and is moving. Galaxies are clustered; some are in small groups, like our local group of



On a large scale the universe seems roughly homogeneous. A box with sides 100 million light years in length would hold quite similar contents no matter where in the universe it is plunked down.

which the Milky Way and Andromeda are the dominant members. Others are in big clusters with hundreds of members. But on the really large scale the universe genuinely does seem smooth and simple. If you imagine a box whose sides are a hundred million light years, then the contents will be more or less the same wherever you plunk the box down. In other words, there's a well-defined sense in which the observable universe is roughly homogeneous above this scale.

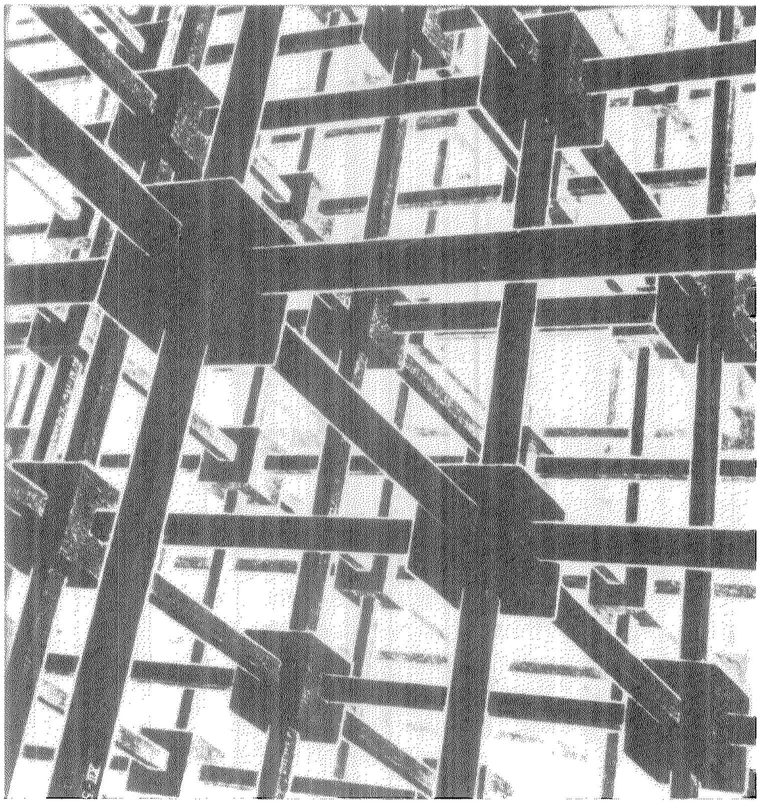
When we look out at the nearest 2,000 galaxies (that's out to a distance of 2 or 3 hundred million light years), they appear fairly uniform over the sky. And as we look at still fainter galaxies, probing greater distances, clustering becomes even less evident. This tells us that we are not in the kind of universe with clusters of clusters of clusters ad infinitum. Such a universe would look equally lumpy over the sky whatever depth you probed it to. So unless we are "anti-Copernican" and assign ourselves some sort of central position, the isotropy all around *us* implies that the universe must be roughly isotropic around *any* galaxy, that the universe is homogeneous, and that all parts have had more or less the same history.

The fox knows many things, but the hedgehog knows one big thing. Cosmologists are the most hedgehog-like of all scientists because their subject boasts very few firm facts, though each has great ramifications. The first such fact emerged in 1929 when Hubble enunciated his famous law that galaxies recede from us with speeds proportional to their distance. We seem to inhabit a homogeneous universe where the distances between any two widely separated pairs of galaxies stretch as some uniform function of time. This doesn't imply that we are in some central "plague spot," because an observer sitting in any other galaxy would see the same uniform expansion around him.

Hubble's work suggested that galaxies must have been crowded together at some time in the past, 10 or 20 billion years ago. But he had no direct evidence for a "beginning." The clinching evidence that there was a so-called Big Bang came in 1965 when Arno Penzias and Robert Wilson of Bell Telephone Labs detected the cosmic microwave background radiation. This was an accidental discovery. Their prime motive was a practical one—to communicate with artificial satellites. At first they didn't realize what they had found, but the excess background noise in their instruments could only mean that even intergalactic space wasn't completely cold. It's about 3 degrees above absolute zero. This may not sound like much, but it implies that there are about 400 quanta of radiation (photons) for every cubic centimeter. Indeed, there are about a billion photons for every atom in the universe.

There's no way of accounting for this radiation, its spectrum (roughly that of a blackbody), and its isotropy except on the hypothesis that it is indeed a relic of a phase when the entire universe was hot, dense, and opaque. Everything must have once been a very compressed and hot gas, hotter than the centers of stars. And although the intense radiation in this primordial fireball was cooled and diluted by the expansion, the wavelengths being stretched and redshifted, the radiation would still be around. After all, it fills the entire universe and has nowhere else to go.

This microwave background radiation is a relic of an era long before any stars or galaxies existed. We've come to believe that another such relic is the element helium, which would have been made from protons



A good analogy to the expanding universe is M.C. Escher's inifinite lattice, which would expand if all the rods lengthened at the same rate, and which has no center.

and neutrons during the first few minutes when the fireball was at a temperature of a billion degrees. Helium would have been made in just the proportion that astronomers now find by spectroscopic studies of stars and nebulae. And it's extraordinary that we can extrapolate back to such an early epoch on the basis of a simple theoretical model, assuming the laws of nuclear physics were the same as they are now, and account for the extremely high and uniform cosmic helium abundance.

More detailed work, much of it done here at Caltech, has firmed up the consensus that everything did indeed emerge from the hot Big Bang. Discrepancies could have emerged in the last 20 years, but none have done so. Still, this isn't yet a firm dogma. Conceivably, satisfactory proof is as illusory as it was for a Ptolemaic astronomer who had just fitted a new epicycle. Cosmologists are sometimes chided for being "often in error but never in doubt."

But, at the moment, the hot Big Bang model certainly seems far more plausible than any other equally specific alternative. Most of us therefore adopt a cosmogonic framework that looks like this: Stars and galaxies all emerged from a universal thermal soup. It was initially smooth and almost featureless—but not quite. There were (although we don't know why) small fluctuations from place to place in the expansion rate. Embryonic galaxies were slightly over-dense regions whose expansion lagged behind the mean expansion. And these embryos eventually evolved into disjoint clouds whose internal expansion halted. These protogalactic clouds then collapsed to make individual galaxies when the universe was perhaps 10 percent of its present age. Subsequently the galaxies would have grouped into clusters, a process that can be quite well simulated by n-body dynamical computer calculations.

The galaxies that Hubble observed were all within a few hundred million light years of us—relatively close compared to the distance we can now probe. But because of the large-scale homogeneity of the universe, Hubble got a fair sample of it. His classification of galaxies has survived and stood the test of time. But Hubble was acutely aware of observational limitations, and his great book, *The Realm of the Nebulae,* concludes with these words:

> With increasing distance our knowledge fades and fades rapidly. Eventually we reach the dim boundary, the utmost limits of our telescope. There we measure shadows, and we search among ghostly errors of measurement for landmarks that are scarcely more substantial. The search will continue. Not until the empirical resources are exhausted need we pass on to the dreamy realm of speculation.

This search *has* continued as more powerful telescopes and detectors have been employed. Because light travels at a finite speed, we see distant parts of the universe as they were a long time ago. So we can sample the past even if we can't repeat it. To see any cosmic evolutionary trend one must look back in time by a good fraction of the 10-billion-plus years for which the universe has been expanding. The first person to do this was Sir Martin Ryle at Cambridge in the late 1950s. He found clear evidence that conditions were different when galaxies were young. His telescopes picked up radio waves from some active galaxies (the ones that we now think harbor massive black holes) even when these were too far away to be observed by the optical techniques of the time. He couldn't determine the distance by radio measurements alone, but he assumed that, at least statistically, the ones appearing faint were more distant than those appearing intense. He counted the numbers with various apparent intensities and found that there were too many faint galaxies (in other words, more distant ones) compared to brighter and closer ones. This was discomforting to those who believed in a steady-state universe, with whom Ryle was having a running battle at the time. But it was compatible with an evolving universe, if galaxies were more prone to violent outbursts in the remote past.

Optical astronomers joined this enterprise after the discovery of quasars in 1963. Quasars are hyperluminous nuclei of galaxies, and optical astronomers have now seen some so far away that the light set out when the universe had less than a fifth of its present scale. And it's also clear from quasars, as it was first from Ryle's data, that the cosmic scene was much more violent when galaxies were young. Most of the runaway catastrophes, the formation of great black holes, happened early in galactic history, when less gas was locked up in stars and more was still available to fuel the central monster.

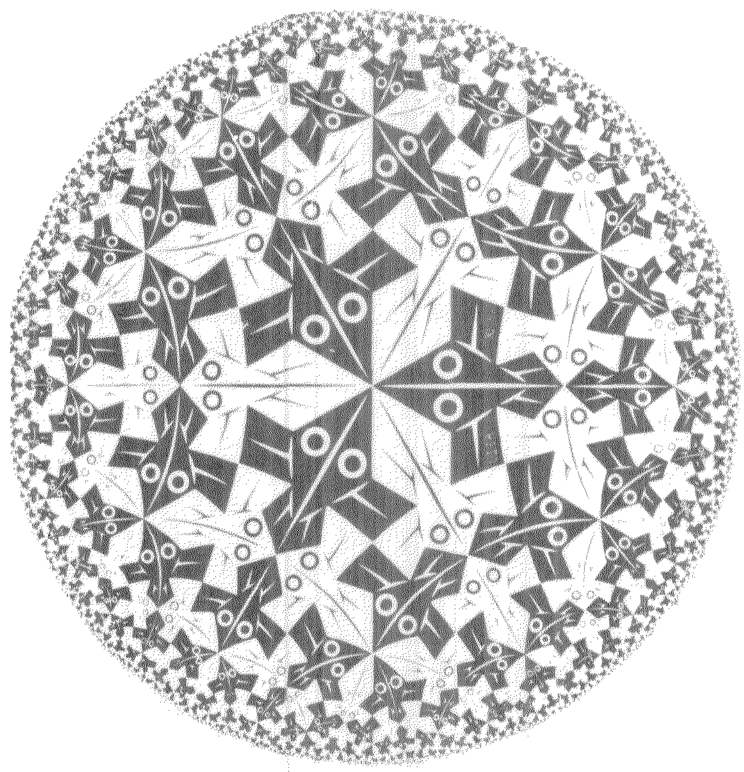
Ordinary galaxies, those without these hyperluminous quasar nuclei, would be almost invisibly faint at these great distances. But the latest sensitive detectors, such as charge coupled devices (CCDs), have recently revealed huge numbers of objects, closely packed over the sky, which are probably young galaxies at the stage of a protogalactic cloud contracting to form a disk. We must await the next generation of telescopes, of which the 10-meter Keck Telescope will be the first, to image these objects brightly enough to reveal their shape and form with any clarity. We shall then be able to obtain "snapshots" of groups of galaxies at different distances (and therefore different evolutionary stages) and trace directly how galaxies emerged from amorphous beginnings at high red shifts.

One stumbling block in understanding galaxies is the rather embarrassing fact that 90 percent of their mass is unaccounted for. When we study the orbital speeds of gas in the outer parts of galaxies, we find that the gas a long way out is still moving surprisingly fast and indeed would be escaping from the system were its centrifugal force not counterbalanced by the gravitational pull of more stuff than we see. We get this evidence also from the motions of galaxies in groups and clusters.

There's no reason, really, that we should be amazed by evidence of this "dark matter." There's no reason why everything in the universe should shine brightly, but it's still a mystery what this dark matter actually is. It could be a huge population of faint stars, too small to have ignited their nuclear fuel. Or it could be the remnants of massive stars that might have been bright in the early phases of galactic evolution but now have all died. A third idea, much discussed in recent years, is that the primordial fireball might have had extra ingredients apart from the ordinary atoms and radiation we observe. Elementary particles of some novel type could collectively exert large-scale gravitational forces.

There are various observational ways of deciding among such varied options. We might look for very faint infrared stars with high motions or for gravitational lensing due to compact stars or black holes. If there are some mysterious particles filling our galaxy, we might even (though their interaction with matter would be very small), be able to detect them by laboratory experiments. It would be especially interesting if we could learn by astronomical methods more about neutrinos, ghostly and elusive particles that hardly interact at all with ordinary matter; or, better still, if we could discover some new fundamental particle—the photino, for instance, which has been predicted by some theorists.

If such particles turned out to account for dark matter, we would then have to view the



Another Escher drawing illustrates what we actually see as we look out from our "center" at the light from distant galaxies crowded together in earlier epochs.

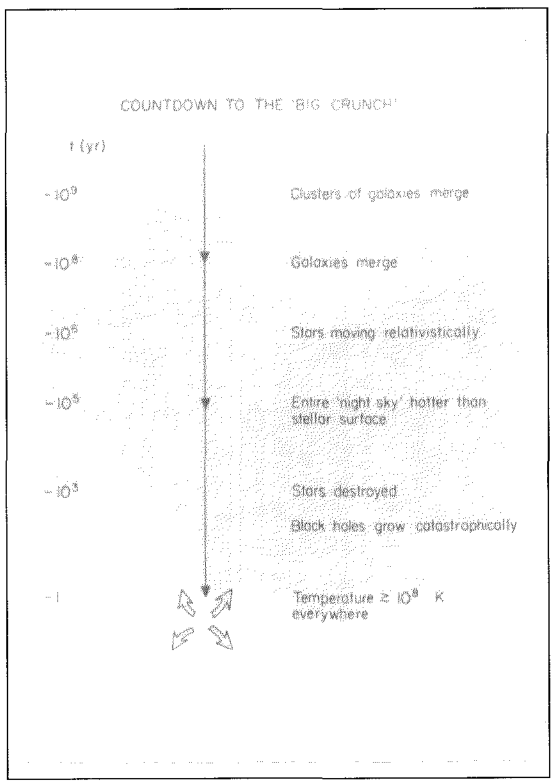
galaxies, the stars, and ourselves in a downgraded perspective. Copernicus, more than four centuries ago, dethroned the Earth from any central position. Early in this century, Shapley and Hubble demoted us from any privileged location in space. But now even *particle chauvinism* would have to be abandoned. The protons, neutrons, and electrons of which we and the entire astronomical world are made could be a kind of afterthought in a cosmos where neutrinos or photinos control the overall dynamics. Great galaxies could be just a sort of puddle or sediment in a cloud of invisible matter, 10 times more massive and extensive.

The amount of dark matter in the universe, important for galactic structure, is even more crucial for the very long term future of the universe. Will it go on expanding forever so that the galaxies fade and disperse into an ultimate heat death? Or will it collapse so that our descendants all share the fate of someone who falls into a black hole, the firmament falling on their heads to recreate a fireball like that from which we believe the universe emerged?

To answer this question we need to know the amount of gravitating matter tending to brake or slow down the expansion. We're now expanding; we don't know whether we're decelerating a lot or only a little, but it's easy to calculate how much gravitating matter is needed to bring the expansion to a halt. This critical density works out at about three atoms per cubic meter. It doesn't sound like very much, but even if we include the galaxies we see, plus all the dynamically inferred dark matter in galaxies and clusters, the mean density still falls short of this critical value by a factor of about five. There could still, however, be some more elusive material between clusters of galaxies. Absent evidence is not evidence of absence, and our knowledge of dark matter is still very biased and incomplete. That being so, it is at least amusing to consider both of the eschatologies suggested by our simple theories.

What would happen if our universe recollapsed? The red shifts of distant galaxies would be replaced by blue shifts, and galaxies would crowd together again. Space is already becoming more and more punctured as isolated regions—dead stars, and galactic nuclei—collapse to form black holes, but this would then just be a precursor of a universal squeeze to the Big Crunch that would eventually engulf everything. Galaxies would merge; stars would move faster, just as the atoms in a gas move faster as you compress the gas. Stars would eventually be destroyed, not by hitting each other, but because the night sky had become hotter than their centers. The final outcome would be a fireball like that which initiated the universe's expansion—though somewhat more lumpy and unsynchronized, and with extra entropy from starlight. When might this happen? The earliest would be 50 billion years from now—at least 10 times the future lifetime of our sun.

What about the other case? What happens if there isn't enough gravitating stuff ever to halt the universe's expansion? Gravitational binding energy is being released as stars, galaxies, and clusters progressively contract. This inexorable trend is delayed by rotation, nuclear energy, and the sheer scale of astronomical systems, which makes things happen slowly and staves off gravity's final victory. But if the universe expands indefinitely, even the slowest processes can run their full course, and the universe then has enough time to run down to a final heat death. If protons don't live forever, ordinary stars may eventually decay. If protons do last forever, then the final heat death will be spun out over a much longer period, as neutron stars tunnel quantum-mechanically into black holes. The length of time it would take for this to happen is up to $10^{10^{76}}$ (seconds or



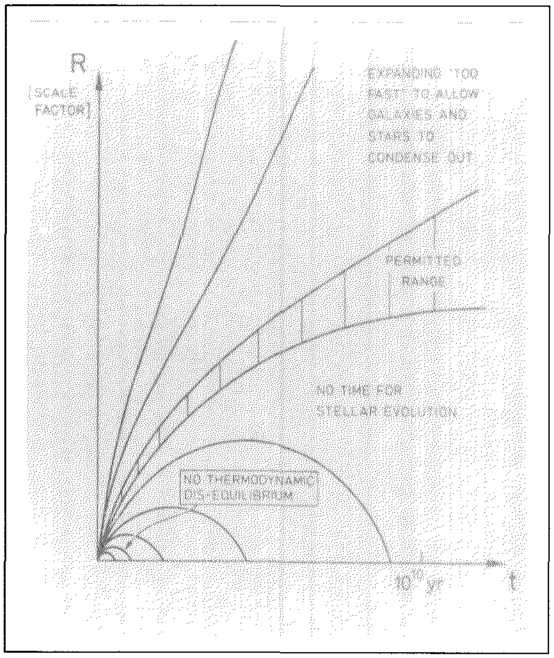
years; it doesn't matter), or 1 followed by about as many zeros as the number of atoms in the observed universe. Even if the universe were made of ink, you couldn't write this number down.

In an article written some years ago in *Reviews of Modern Physics*, Freeman Dyson discussed the future of the universe in great detail. He doesn't say much about the Big Crunch and the collapsing universe (I think that idea gives him claustrophobia), but he does address in detail some of the other points that I've summarized here, and he goes on to contemplate the outcome for intelligent life in this universe. Can it survive and develop intellectually on finite energy reserves forever, thinking infinite thoughts and storing or communicating an ever-increasing body of information? He shows, comfortingly, that in principle this can be done: As the background temperature falls, you must keep cooler, think progressively more slowly, and hibernate for very long periods. Will our descendants need to follow Dyson's conservationist maxims to survive an infinite future, or will they fry in the Big Crunch a few tens of billions of years hence? We need a more complete inventory of what's in the universe by observing it in all wavebands, searching for black holes, and understanding all sorts of exotic particles before we can pronounce a long-term forecast for the next hundred billion years.

THE FAR FUTURE OF AN EVER-EXPANDING UNIVERSE

10^{14} yr	Ordinary stellar activity completed
10^{17} yr	Significant dynamical relaxation in galaxies
10^{20} yr	Gravitational radiation effects in galaxies
[10^{31}–10^{36} yr	Proton decay]
$10^{64}(m/m_{\odot})^3$ yr	Quantum evaporation of black holes
10^{1500} yr	White dwarfs → iron*
$10^{10^{26}}$–$10^{10^{76}}$	Neutron stars undergo quantum* tunneling to black holes, which then "quickly" evaporate

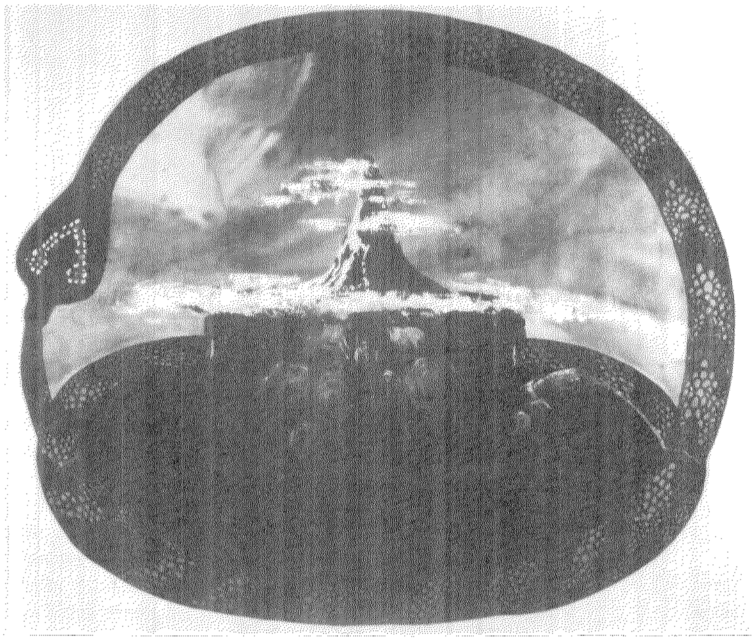
*If proton decay does not occur



The dynamics of the early universe must have been finely tuned to allow stars and galaxies to form in the "permitted range" here. If it had re-collapsed sooner, there would have been no time for stellar evolution. It it had expanded much faster, the kinetic energy would have overwhelmed gravity, and the clouds from which galaxies are born could not have coalesced.

These two alternative long-range futures, which seem very different, present a puzzle, because the initial conditions that could have led to anything like our present universe are very restrictive compared to the possibilities that might have been set up. We know that our universe is still expanding after 10 billion years. Had it recollapsed sooner, there would have been no time for stellar evolution, possibly not even time for it to have gone through anything other than the fireball state, precluding any thermodynamic disequilibrium. On the other hand, the expansion could not have been much faster than the critical rate; otherwise the kinetic energy would have overwhelmed gravity, and the clouds that developed into galaxies wouldn't have been able to pull themselves together. That's equivalent to saying that the density of the universe can't be far below the critical density. So the dynamics of the early universe must have been finely tuned in order to end up in the shaded region on the graph above. In Newtonian terms, the fractional difference between the initial potential and kinetic energies of any spherical region must have been very small.

So why was the universe set up to expand in this rather special way? There are other issues that baffle us similarly. In particular, why does the universe contain small-scale initial fluctuations that are necessary as "seeds" for galaxy formation, while still remaining so homogeneous overall? We can't answer these questions, even though we can trace in broad outline the course of cosmic evolution back



Ancient Indian cosmologists envisaged the Earth as supported by four elephants standing on a giant turtle. But what held up the turtle remained a mystery, as do the initial conditions of the universe to scientists today.

to when galaxies first formed and even all the way back to a universal fireball a few seconds old.

The ancient Indian cosmologists envisaged the earth being supported by four elephants standing on a giant turtle, but they weren't sure what held up the turtle. Conceptually we still end up in similarly bad shape with an appeal to initial conditions, saying things are as they are because they were as they were. Key features of the universe must have been imprinted before the first second had elapsed. So what happened during the first second?

The further back we extrapolate, the less confidence we have in the adequacy or applicability of known physics. For instance, the material will be squeezed above nuclear densities for the first microsecond. But if you think of time on a logarithmic scale, it seems a severe omission to ignore these early eras. And theorists differ on how far back they are prepared to extrapolate with a straight face. Some have higher credulity thresholds than others. In particular, those whose intellectual habitat is the gee-whiz fringe of particle physics are interested in the possibility that the early universe might once have been at colossally high temperatures. The goal of such physicists is to develop a so-called grand unified theory of all the forces governing the microphysical world. But they are faced with a stumbling block: the critical energy at which the so-called symmetry breaking is supposed to have occurred is about 10^{15} giga electron volts (GeV), which is a million million times higher than experiments on Earth can reach. It's hard, therefore, to test these theories, because only tiny effects are predicted in our low-energy world. For instance, protons may decay very slowly. But if we are emboldened to extrapolate the Big Bang theory back not just to one second but to 10^{-36} seconds, then, but only then, all thermal energies would exceed 10^{15} GeV. So perhaps the early universe was the only accelerator where the requisite energies for unifying the forces could ever be attained.

But the snag is that this accelerator shut down 10 billion years ago. So we can't learn anything about its activities unless the 10^{-36} second era left behind some fossils, just as helium is the fossil left from the first few minutes. Physicists would enthusiastically seize at even the most trifling vestige surviving from this ultra-early phase. An especially exciting possibility raised by these theories is that the particular mix of matter and radiation in our universe, a billion light quanta for every atom, may result from a small fractional favoritism of matter over antimatter established at that time.

Unified theories bring a new set of questions, such as the origin of matter, into the scope of serious discussion. The realization that protons aren't strictly conserved suggests furthermore that our universe may possess no conserved quantities other than those that are actually zero—such as total electric charge. This, combined with the concept of a so-called inflationary phase, in which our universe could have emerged from a single quantum fluctuation, allows us to envisage a sort of *ex nihilo* creation of the entire universe.

These concepts are still very tentative of course. Their present status resembles that of the theory of element synthesis in the Big Bang when Gamow and Lemaitre first discussed it 40 years ago. And just as the ideas of those pioneers were put on a surer footing by later developments, so we can hope that the concept of the ultra-early universe will also firm up. Indeed, we may not have to wait as long: In earlier decades only a few physicists took cosmology seriously, but now these ideas engage the interest of many leading mainstream theorists. And these developments certainly offer cosmologists a psychological boost, creating a symbiotic rather than a parasitical relationship with their physicist colleagues. It also makes me feel, in comparison with some of my colleagues, like a cautious empiricist, very reluctant to stray far from the data. That's an unusual feeling for an astrophysicist to have.

With phenomena such as ordinary stars we feel fairly confident that we know the relevant physics. When conditions get more extreme (in galactic nuclei, for instance) we're less confident, although it's astounding how far we can go without running against a contradiction. One theme that has emerged is the interdependence of different phenomena. The everyday world is determined by atomic structure, the stars are probably determined by atomic nuclei, and galaxies may be held together by some kind of subnuclear particles that are relics of a high-energy phase.

But in the early Big Bang or in gravitational collapse inside black holes we're confronted by conditions so extreme that we know for sure that we *don't* know enough physics. Above all, physics is conceptually unsatisfactory in that we lack an adequate theory of quantum gravity. Two great foundations of physics are the quantum uncertainty principle and Einstein's general relativity. The theoretical superstructures erected on these foundations are disjoint. There's normally no overlap in their domain of relevance because quantum effects are important on a microphysics scale, gravity only on the astronomical scale. But when the universe was squeezed to colossal densities (at 10^{-43} seconds, the Planck time), gravity could be important on the scale of a single particle, a single thermal quantum. Even the boldest physicists can extrapolate back no further.

Despite these difficulties some theorists believe that it's no longer premature to explore physical laws prevailing at the Planck time. They've come up with many fascinating ideas. There's no consensus about which concept might really fly, but it's certainly no longer just cranks who try to consider all physical forces in one go. We may have to jettison commonsense notions of space and time, the dimensionality of our world, and many other things.

What about gravity? Two features of this peculiar force that holds together individual stars and entire galaxies are quite crucial for cosmogonic processes. The first feature is that gravity drives things *further from equilibrium*, not toward equilibrium. When gravitating systems *lose* energy they get *hotter*; for example, an artificial satellite speeds up as it spirals downward due to atmospheric drag. Another example is the sun. If its radiative losses were not compensated by nuclear fusion, the sun would contract and deflate but would end up with a hotter interior than before. It needs more pressure inside it to balance the stronger gravity when it's more compressed. So, from the initial Big Bang to our present solar system, this antithermodynamic behavior of gravity has been amplifying density contrast and creating temperature gradients—prerequisites for the emergence of any complexity.

The second key feature of gravity is its *weakness*. The gravitational force within an atom is almost 40 powers of 10 weaker than the electrical forces that bind it. As I explained in discussing stars, gravity holds sway on sufficiently large scales, but those scales are vast because gravity is weak. If gravity were somewhat stronger, say 30 rather than 40 powers of 10 weaker than electromagnetism, then a small-scale speeded up universe could exist, in which stars, gravitationally bound fusion reactors, had 10^{-15} of the sun's mass and lived for less than a year. This might not allow enough time for complex systems to evolve. There would be fewer powers of 10 between astrophysical time scales and basic microscopic time scales for physical or chemical reactions. Moreover, complex structures could not get very large without themselves being crushed by gravity. Our universe is large and diffuse and evolves slowly *because* gravity is so weak. Its extravagant scale, billions of light years, is necessary to provide enough time for the cooking of elements inside stars and for interesting complexity to evolve around even just one star in just one galaxy. So a force like gravity is essential if structures are to emerge from amorphous starting points; but, paradoxically, the weaker it is, the greater and more complex are its consequences.

The evidence for apparent fine-tuning in the initial expansion rate (in Dyson's words, "The universe seems to have known we were coming") has led some physicists to highlight other coincidences in the physical laws. All key features of the everyday world and the astronomical scene are determined by a few basic physical laws and constants—the masses of elementary particles and the strength of the basic forces between them. And in many cases a rather delicate balance seems to prevail. For example, if the nuclear forces were slightly stronger relative to electromagnetism, the diproton would be stable, ordinary hydrogen wouldn't exist, and stars would evolve quite differently. If nuclear forces were slightly weaker, no chemical elements other than hydrogen would exist and chemistry would be a trivially simple subject.

The details of stellar nucleosynthesis—the nuclear transmutations inside stars that forge the elements we are made of—are sensitive to other apparent accidents. For instance, Fred Hoyle showed that carbon and oxygen can both be readily synthesized only because there's a sort of specially tuned resonance in the carbon nucleus.

What shall we make of all this? It shouldn't occasion any surprise that we've evolved to fit our local environment around a star. But what surprises some of us is that the physical laws should permit any complexity to evolve anywhere. Some physicists don't take this very seriously, but others envisage a kind of natural selection among an ensemble of universes governed by different laws. Most universes would be still-born in the sense that no complexity could develop within them. But some, including ours, could perhaps exist with any requisite tuning of the parameters. In other words, given that we know that our cosmic environment permits observers to exist, maybe we shouldn't take the Copernican principle too far. We wouldn't feel justified in assigning ourselves a central position in the cosmos, but it may be equally unrealistic to deny (or to be surprised) that our situation can be privileged in any sense.

The eventual status of this so-called "anthropic principle" will depend, I think, on what the laws of nature are really like. If some fundamental theory yields unique values for all the ratios, then it may be inconceivable to envisage a universe with different constants. We then have to accept it as coincidental, or even providential, that these constants happen to lie in the restricted range that allow complexity and consciousness to evolve in the low-energy world we inhabit. The intricacy implicit in these unique laws may astonish us, but our reaction would be no less subjective than a mathematician's surprise at the rich intellectual structures that can stem from simple axioms.

But if, contrariwise, the basic laws turned out to involve some random elements, then the ensemble idea could be put on a serious footing. Some cosmologists suggest that different parts of an infinite universe could have cooled down after the Big Bang with different constants. There could be different domains in which the physics could be different, and complex evolution could occur only in oases where the laws of nature were of propitious dimensions.

Our oasis must be at least 10 billion light years across, because the physical laws seem the same everywhere we can observe. But the desert, or still-born, regions may in principle be observed within the distant future (maybe 10^{12} years hence) when our horizon is expanded sufficiently for light from more remote domains to reach us. This time delay is, to be sure, an impediment to practical empirical tests, but conceptually the situation is no different from the conjectures of early "cosmographers" about continents beyond the limits of the then-known world.

Einstein said, "The most incomprehensible thing about the universe is that it is comprehensible." The physical laws that our brains are somehow attuned to understand apply not just in the lab but in the remotest quasar and even in the early instants of the Big Bang. Were this not so, were there not a firm link with local physics, cosmology could never rise beyond ad hoc explanations on the level of the *Just So Stories*. Some optimists indeed believe that a comprehensive and comprehensible theory for all the fundamental forces may emerge from a symbiosis between cosmology and particle physics.

This would mean in a sense, as some physicists have emphasized, the end of fundamental physics. But it would emphatically *not* mean the end of challenging science. I first heard the following metaphor for what the physicist does from Dick Feynman.

Suppose you were unfamiliar with the game of chess. Then, just by watching games being played you could infer what the rules were. The physicist likewise finds patterns in the natural world and learns what dynamics and what transformations govern its basic elements. But in chess, learning how the pieces move is no more than a trivial preliminary to the absorbing progression from novice to grand master. The whole point and interest of that game lie in exploring the complexity implicit in a few deceptively simple rules. Likewise, all that's happened in the universe over the last 10 billion years—the emergence of galaxies, the formation of their constituent stars, and the intricate evolution that, on a planet of at least one star, has led to creatures able to wonder about it all—may be implicit in a few fundamental equations of physics. Exploring and trying to understand all this offers an unending quest and a challenge that has barely begun. □