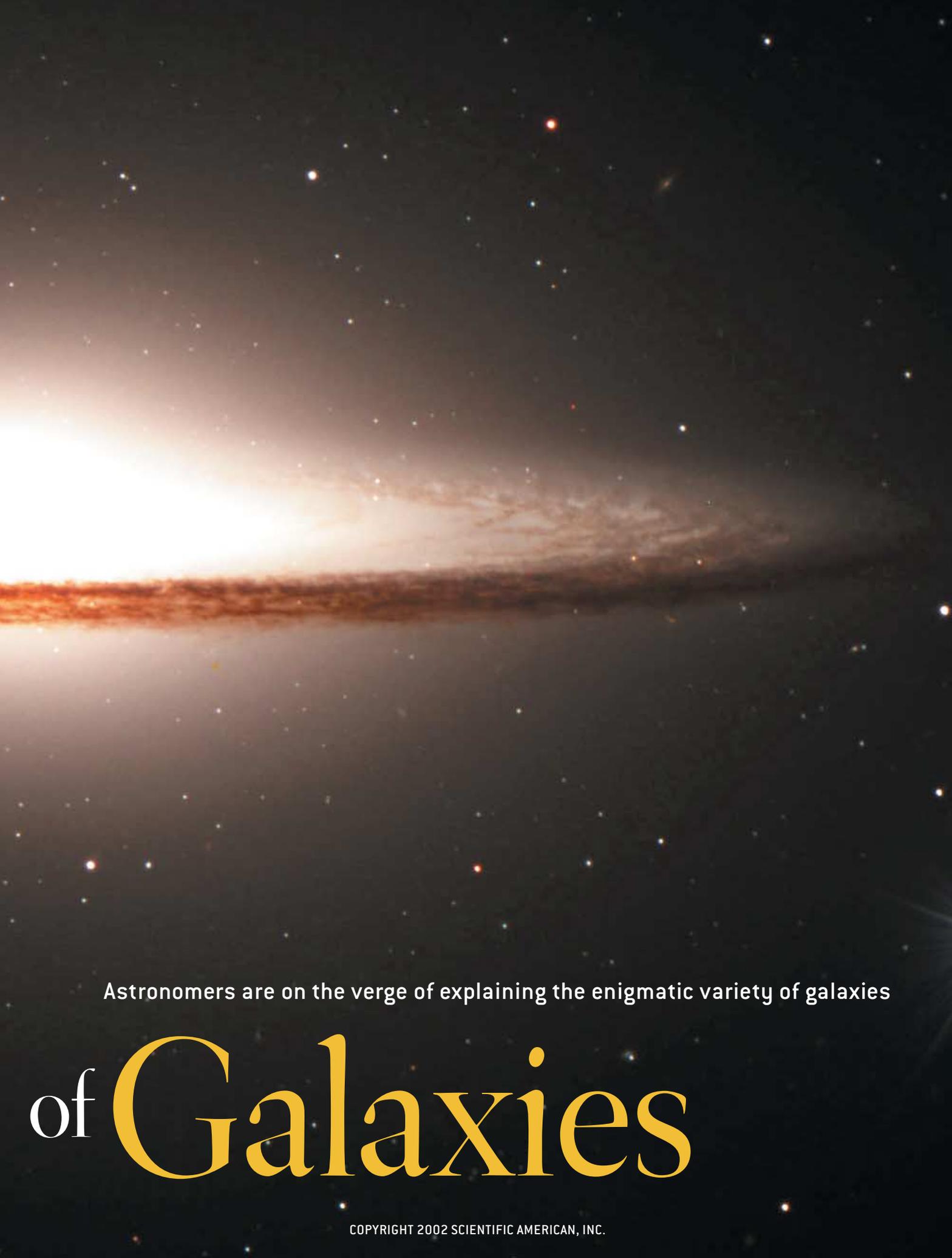


SOMBRERO GALAXY is an all-in-one package: it exemplifies nearly every galactic phenomenon that astronomers have struggled for a century to explain. It has a bright ellipsoidal bulge of stars, a supermassive black hole buried deep within that bulge, a disk with spiral arms [*seen close to edge-on*], and star clusters scattered about the outskirts. Stretching beyond this image is thought to be a vast halo of inherently invisible dark matter.

By Guinevere Kauffmann and Frank van den Bosch

The Life Cycle

COPYRIGHT 2002 SCIENTIFIC AMERICAN, INC.



Astronomers are on the verge of explaining the enigmatic variety of galaxies

of Galaxies

In many science-fiction stories,

a mighty empire dooms itself through its hubris: it presumes to conquer and rule an entire galaxy. That seems a lofty ambition indeed. To bring our Milky Way galaxy to heel, an empire would have to vanquish 100 billion stars. But cosmologists—those astronomers who study the universe as a whole—are unimpressed. The Milky Way is one of 50 billion or more galaxies within the observable reaches of space. To conquer it would be to conquer an insignificant speck.

A century ago nobody knew all those galaxies even existed. Most astronomers thought that the galaxy and the universe were synonymous. Space contained perhaps a billion stars, interspersed with fuzzy splotches that looked like stars in the process of forming or dying. Then, in the early decades of the 20th century, came the golden age of astronomy, when American astronomer Edwin Hubble and others determined that those fuzzy splotches were often entire galaxies in their own right.

Why do stars reside in gigantic agglomerations separated by vast voids, and how do galaxies take on their bewildering variety of shapes, sizes and masses? These questions have consumed astronomers for decades. It is not possible for us to observe a galaxy forming; the process is far too slow. Instead researchers have to piece the puzzle together by observing many different galaxies, each caught at a different phase in its evolutionary history. Such measurements did not become routine until about

a decade ago, when astronomy entered a new golden age.

Spectacular advances in telescope and detector technology are now giving astronomers a view of how galaxies have changed over cosmic timescales. The Hubble Space Telescope has taken very deep snapshots of the sky, revealing galaxies down to unprecedentedly faint levels. Ground-based instruments such as the giant Keck telescopes have amassed statistics on distant (and therefore ancient) galaxies. It is as if evolutionary biologists had been handed a time machine, allowing them to travel back into prehistory and take pictures of the animals and plants inhabiting the earth at a series of different epochs. The challenge for astronomers, as it would be for the biologists, is to grasp how the species observed at the earliest times evolved into what we know today.

The task is of truly astronomical proportions. It involves physics on wildly disparate scales, from the cosmological evolution of the entire universe to the formation of a single star. That makes it difficult to build realistic models of galaxy formation, yet it brings the whole subject full circle. The discovery of all those billions of galaxies made stellar astronomy and cosmology seem mutually irrelevant. In the grand scheme of things, stars were just too small to matter; conversely, debates over the origin of the universe struck most stellar astronomers as hopelessly abstract. Now we know that a coherent picture of the universe must take in both the large and the small.

Overview/*Galaxy Evolution*

- One of the liveliest subfields of astrophysics right now is the study of how galaxies take shape. Telescopes are probing the very earliest galaxies, and computer simulations can track events in unprecedented detail.
- Researchers may soon do for galaxies what they did for stars in the early 20th century: provide a unified explanation, based on a few general processes, for a huge diversity of celestial bodies. For galaxies, those processes include gravitational instability, radiative cooling, relaxation (whereby galaxies reach internal equilibrium) and interactions among galaxies.
- Several vexing questions remain, however. A possible answer to these questions is that stars, seemingly insignificant to such large bodies as galaxies, actually have a profound and pervasive effect on their structure.

Galactic Species

TO UNDERSTAND HOW galaxies form, astronomers look for patterns and trends in their properties. According to the classification scheme developed by Hubble, galaxies may be broadly divided into three major types: elliptical, spiral and irregular [*see illustration on opposite page*]. The most massive ones are the ellipticals. These are smooth, featureless, almost spherical systems with little or no gas or dust. In them, stars buzz around the center like bees around a hive. Most of the stars are very old.

Spiral galaxies, such as our own Milky Way, are highly flattened and organized structures in which stars and gas move on circular or near-circular orbits around the center. In fact, they are also known as disk galaxies. The pinwheel-like spiral arms are filaments of hot young stars, gas and dust. At their centers, spiral galaxies contain bulges—spheroidal clumps of stars that are reminiscent of miniature elliptical galaxies. Roughly a third of spiral galaxies have a rectangular structure toward the cen-

TYPES OF GALAXIES

ASTRONOMERS SORT GALAXIES using the “tuning fork” classification scheme developed by American astronomer Edwin Hubble in the 1920s. According to this system, galaxies come in three basic types: elliptical (*represented by the handle of the fork at right*), spiral (*shown as prongs*) and irregular (*shown below at left*). The smallest galaxies, known as dwarfs, have their own uncertain taxonomy.

Within each of the types are subtypes that depend on the details of the galaxy’s shape. Going from the top of the tuning fork to the bottom, the galactic disk becomes more prominent in optical images and the central bulge less so. The different Hubble types may represent various stages of development. Galaxies start off as spirals without bulges, undergo a collision during which they appear irregular, and end up as ellipticals or as spirals with bulges.

—G.K. and F.v.d.B.

IRREGULARS

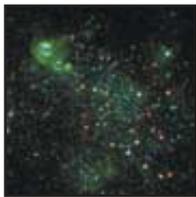


M82
Irregular

DWARF TYPES



M32
Elliptical



VII Zw 403
Blue Compact



Small Magellanic Cloud
Irregular



Leo I
Spheroidal

ELLIPTICALS



M89
E0



M49
E4



M110
E5



M84
S0

BARRED SPIRALS



NGC 660
SBa



NGC 7479
SBb



M58
SBc

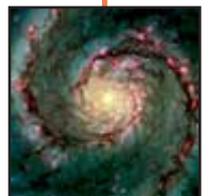
NORMAL SPIRALS



NGC 7217
Sa



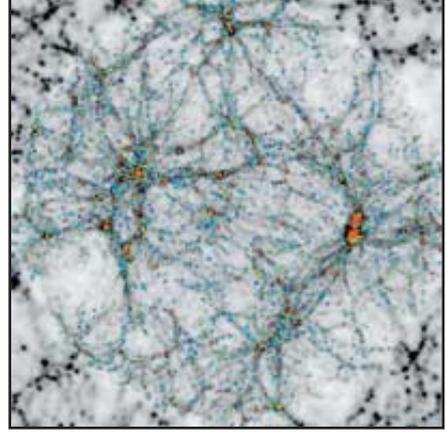
NGC 4622
Sb



M51
Sc

N. A. SHARP/NOAO/AURA/NSF (M82); B. KEEL/HALL TELESCOPE/LOWELL OBSERVATORY (M32); R. SCHULTE-LADBECK/U. HOPP/M. CRONE/ASTROPHYSICAL JOURNAL (blue compact dwarf); NOAO/AURA/NSF (Small Magellanic Cloud); DAVID MALIN, © ANGLO-AMERICAN OBSERVATORY (Leo I); NOAO/AURA/NSF (M89, M49, M110, M84); R. BRANCH/R. MILNER/A. BLOCK/NOAO/AURA/NSF (NGC 660); A. BLOCK/NOAO/AURA/NSF (NGC 7479); F. CIESLAK/A. BLOCK/NOAO/AURA/NSF (M58); B. KEEL/R. BUTA/G. PURCELL/CERRO TOLOLO INTER-AMERICAN OBSERVATORY, CHILE (NGC 7217); G. BYRD/R. BUTA/T. FREEMAN/NASA (NGC 4622); NASA/STSCI/AURA (M51)

Supercomputer simulations of the spatial distribution of galaxies are in excellent agreement with observations.



ter. Such “bars” are thought to arise from instabilities in the disk.

Irregular galaxies are those that do not fit into the spiral or elliptical classifications. Some appear to be spirals or ellipticals that have been violently distorted by a recent encounter with a neighbor. Others are isolated systems that have an amorphous structure and exhibit no signs of any recent disturbance.

Each of these three classes covers galaxies with a wide range of luminosities. On average, however, ellipticals are brighter than spirals, and fainter galaxies are more likely than their luminous counterparts to be irregular. For the faintest galaxies, the classification scheme breaks down altogether. These dwarf galaxies are heterogeneous in nature, and attempts to pigeon-hole them have proved controversial. Loosely speaking, they fall into two categories: gas-rich systems where stars are actively forming and gas-poor systems where no stars are forming.

An important clue to the origin of the galaxy types comes from the striking correlation between type and local galaxy density. Most galaxies are scattered through space far from their nearest neighbor, and of these only 10 to 20 percent are ellipticals; spirals dominate. The remaining galaxies, however, are packed into clusters, and for them the situation is reversed. Ellipticals are the majority, and the spirals that do exist are anemic systems depleted of gas and young stars. This so-called morphology-density relation has long puzzled astronomers.

Light and Dark

A SMALL PERCENTAGE of spirals and ellipticals are peculiar in that they contain an exceedingly luminous, pointlike core—an active galactic nucleus (AGN). The most extreme and rarest examples are the quasars, which are so bright that they completely outshine their host galaxies. Astronomers generally believe that AGNs are powered by black holes weighing millions to billions of solar masses. Theory predicts that gas falling into these monsters will radiate about 10 percent of its intrinsic energy, sufficient to generate a beacon that can be detected on the other side of the universe.

Once considered anomalies, AGNs have recently been shown to be integral to the process of galaxy formation. The peak of AGN activity occurred when the universe was approximately a fourth of its present age—the same time that most of the stars in ellipticals were being formed. Furthermore, supermassive black holes are now believed to reside in virtually every elliptical galaxy, as well as every spiral galaxy that has a bulge, regardless of whether those galaxies contain an AGN [see “The Hole Shebang,” by George Musser; *News and Analysis*, *SCIENTIFIC AMERICAN*, October 2000]. The implication is

that every galaxy may go through one or more episodes of AGN activity. As long as matter falls into the black hole, the nucleus is active. When no new material is supplied to the center, it lies dormant.

Most of the information we have about all these phenomena comes from photons: optical photons from stars, radio photons from neutral hydrogen gas, x-ray photons from ionized gas. But the vast majority of the matter in the universe may not emit photons of any wavelength. This is the infamous dark matter, whose existence is inferred solely from its gravitational effects. The visible parts of galaxies are believed to be enveloped in giant “halos” of dark matter. These halos, unlike those found above the heads of saints, have a spherical or ellipsoidal shape. On larger scales, analogous halos are thought to keep clusters of galaxies bound together.

Unfortunately, no one has ever detected dark matter directly, and its nature is still one of the biggest mysteries in science. Currently most astronomers favor the idea that dark matter consists mostly of hitherto unidentified particles that barely interact with ordinary particles or with one another. Astronomers typically refer to this class of particles as cold dark matter (CDM) and any cosmological model that postulates their existence as a CDM model.

Over the past two decades, astronomers have painstakingly developed a model of galaxy formation based on CDM. The basic framework is the standard big bang theory for the expansion of the universe. Cosmologists continue to debate how the expansion got going and what transpired early on, but these uncertainties do not matter greatly for galaxy formation. We pick up the story about 100,000 years after the big bang, when the universe consisted of baryons (that is, ordinary matter, predominantly hydrogen and helium nuclei), electrons (bound to the nuclei), neutrinos, photons and CDM. Observations indicate that the matter and radiation were distributed smoothly:

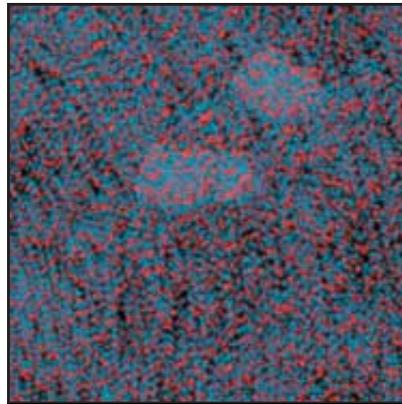
THE AUTHORS

GUINEVERE KAUFFMANN and *FRANK VAN DEN BOSCH* are researchers at the Max Planck Institute for Astrophysics in Garching, Germany. They are among the world’s experts on the theoretical modeling of galaxy formation. Kauffmann has recently turned her attention to analyzing data from the Sloan Digital Sky Survey, which she believes holds the answers to some of the mysteries highlighted in this article. In her spare time, she enjoys exploring Bavaria with her son, Jonathan. Van den Bosch is particularly intrigued by the formation of disk galaxies and of massive black holes in galactic centers. In his free time, he can often be found in a Munich beer garden.

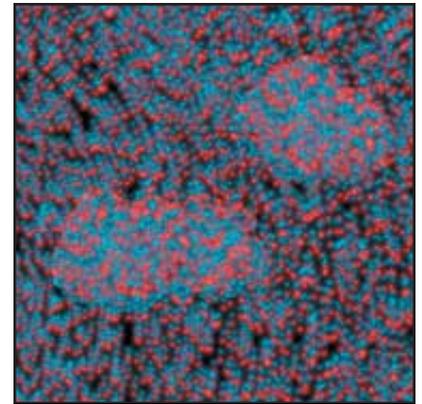
H. MATHIS, V. SPRINGEL, G. KAUFFMANN AND S.D.M. WHITE: Max Planck Institute for Astrophysics, Garching, Germany.
AND G. LEMSON, A. ELДАР AND A. DEKEL: Hebrew University, Israel (simulation of galaxy formation in a region 900 million light-years across)

COOKING UP A GALAXY

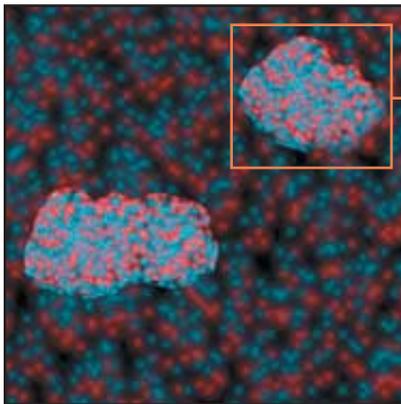
THREE BASIC PROCESSES dictated how the primordial soup congealed into galaxies: the overall expansion of the universe in the big bang, the force of gravity, and the motion of particles and larger constituents. The shifting balance among these processes can explain why galaxies became discrete, coherent bodies rather than a uniform gas or a horde of black holes. In this theory, small bodies coalesce first and then glom together to form larger objects. A crucial ingredient is dark matter, which reaches a different equilibrium than ordinary matter. —G.K. and F.v.d.B.



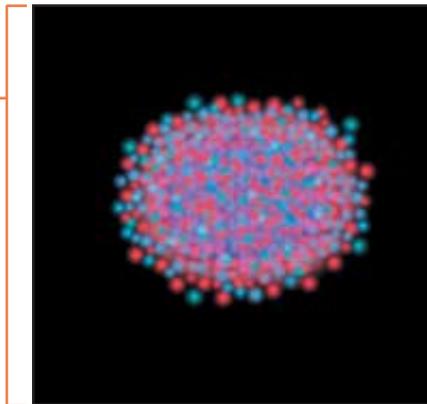
1 In the beginning, a primordial fluid—a mixture of ordinary matter (*blue*) and dark matter (*red*)—fills the universe. Its density varies subtly from place to place.



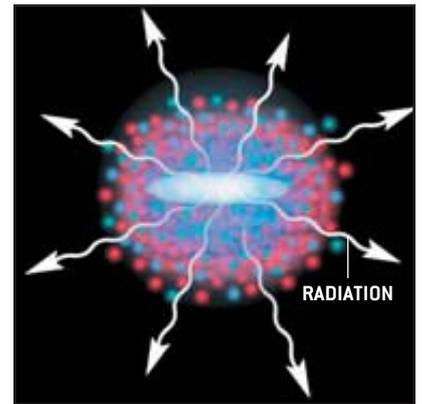
2 At first, cosmic expansion overpowers gravity. The fluid thins out. But patches of higher density thin out more slowly than other regions do.



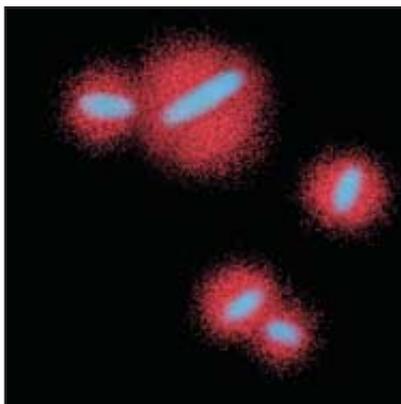
3 Eventually these patches become so dense, relative to their surroundings, that gravity takes over from expansion. The patches start to collapse.



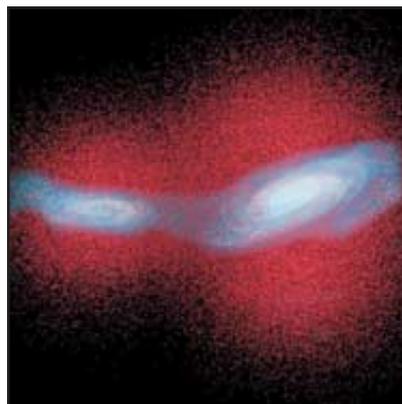
4 As each patch collapses, it attains equilibrium. The density, both of ordinary and of dark matter, peaks at the center and decreases toward the edge.



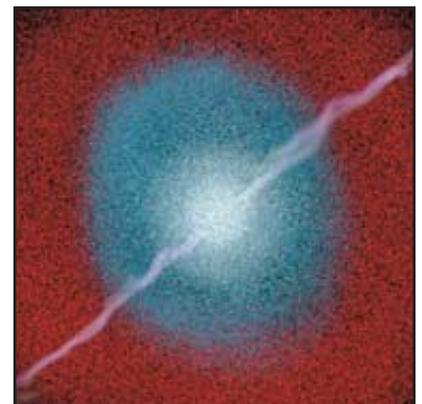
5 Dark matter, being unable to radiate, retains this shape. But ordinary matter emits radiation, collapses into a rotating disk and begins to condense into stars.



6 Protogalaxies interact, exerting torques on one another and merging to form larger and larger bodies. (This step overlaps with steps 4 and 5.)



7 When two disks of similar size merge, the stellar orbits become scrambled. An elliptical galaxy results. Later a disk may develop around the elliptical.



8 The merger triggers new star formation and feeds material into the central black hole, generating an active galactic nucleus, which can spew plasma jets.

the density at different positions varied by only about one part in 100,000. The challenge is to trace how these simple ingredients could give rise to the dazzling variety of galaxies.

If one compares the conditions back then with the distribution of matter today, two important differences stand out. First, the present-day universe spans an enormous range of densities. The central regions of galaxies are more than 100 billion times as dense as the universe on average. The earth is another 10 billion billion times as dense as that. Second, whereas the baryons and CDM were initially well mixed, the baryons today form dense knots (the galaxies) inside gargantuan halos of dark matter. Somehow the baryons have decoupled from the CDM.

The first of these differences can be explained by the process of gravitational instability. If a region is even slightly more dense than average, the excess mass will exert a slightly stronger-than-average gravitational force, pulling extra matter toward itself. This creates an even stronger gravitational field, pulling in even more mass. This runaway process amplifies the initial density differences.

Sit Back and Relax

ALL THE WHILE, the gravity of the region must compete with the expansion of the universe, which pulls matter apart. Initially cosmic expansion wins and the density of the region decreases. The key is that it decreases more slowly than the density of its surroundings. At a certain point, the overdensity of the region compared with its surroundings becomes so pronounced that its gravitational attraction overcomes the cosmic expansion. The region starts to collapse.

Up to this point, the region is not a coherent object but merely a random enhancement of density in the haze of matter that

fills the universe. But once the region collapses, it starts to take on an internal life of its own. The system—which we shall call a protogalaxy from here on—seeks to establish some form of equilibrium. Astronomers refer to this process as relaxation. The baryons behave like the particles of any gas. Heated by shock waves that are triggered by the collapse, they exchange energy through direct collisions with one another, thus achieving hydrostatic equilibrium—a state of balance between pressure and gravity. The earth’s atmosphere is also in hydrostatic equilibrium (or nearly so), which is why the pressure decreases exponentially with altitude.

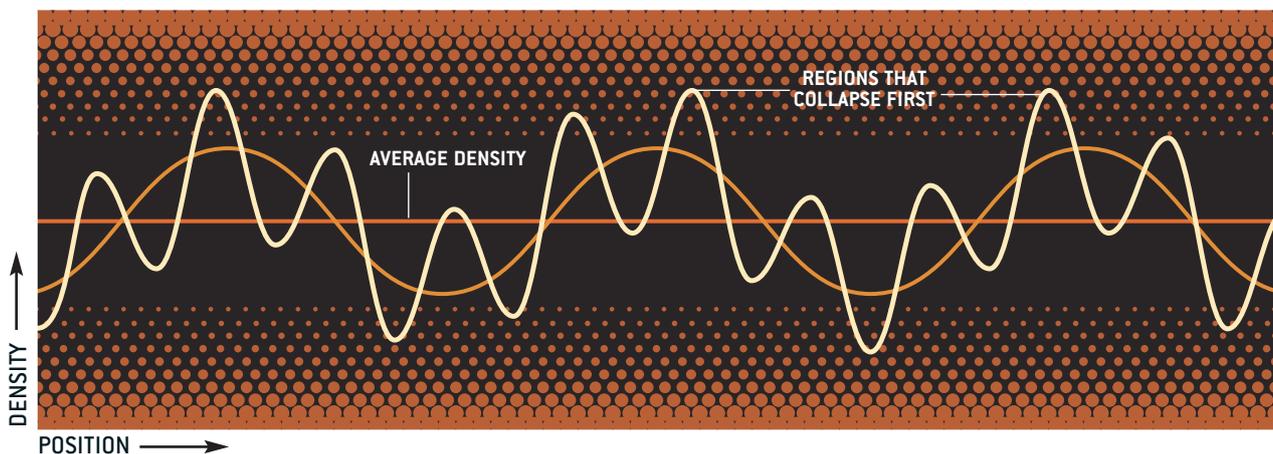
For the dark matter, however, relaxation is distinctively different. CDM particles are, by definition, weakly interactive; they are not able to redistribute energy among themselves by direct collisions. A system of such particles cannot reach hydrostatic equilibrium. Instead it undergoes what is called, perhaps oxymoronically, violent relaxation. Each particle exchanges energy not with another individual particle but with the collective mass of particles, by way of the gravitational field.

Bodies traveling in a gravitational field are always undergoing an exchange of gravitational and kinetic energy. If you throw a ball into the air, it rises to a higher altitude but decelerates: it gains gravitational energy at the expense of kinetic energy. On the way down, the ball gains kinetic energy at the expense of gravitational energy. CDM particles in a protogalaxy behave much the same way. They move around and change speed as their balance of gravitational and kinetic energy shifts. But unlike balls near the earth’s surface, CDM particles move in a gravitational field that is not constant. After all, the gravitational field is produced by all the particles together, which are undergoing collapse.

GALACTIC DENSITY VARIATIONS

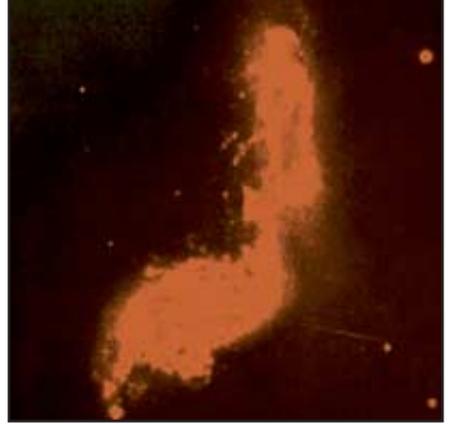
DENSITY VARIATIONS in the pregalactic universe followed a pattern that facilitated the formation of protogalaxies. The variations were composed of waves of various wavelengths in a pattern that music connoisseurs will recognize as “pink noise.” [Indeed, they originated as sound waves in the

primordial plasma.] A small wave was superimposed on a slightly larger wave, which was superimposed on an even larger wave, and so on. Therefore, the highest density occurred over the smallest regions. These regions collapsed first and became the building blocks for larger structures. —G.K. and F.v.d.B.



Astronomers

may be directly observing,
for the first time, the formation of
elliptical galaxies.



Changes in the gravitational field cause some particles to gain energy and others to lose energy. Just as for the baryons, this redistribution of the energies of the particles allows the system to relax, forming a CDM halo that is said to be in virial equilibrium. The process is complicated and has never been worked out in great theoretical detail. Instead researchers track it using numerical simulations, which show that all CDM halos in virial equilibrium have similar density profiles.

The end point of the collapse and relaxation of a protogalaxy is a dark matter halo, inside of which the baryonic gas is in hydrostatic equilibrium at a temperature of typically a few million degrees. Whereas each CDM particle conserves its energy from then on, the baryonic gas is able to emit radiation. It cools, contracts and accumulates at the center of the dark matter halo. Cooling, therefore, is the process responsible for decoupling the baryons from the CDM.

So far we have focused on a single protogalaxy and ignored its surroundings. In reality, other protogalaxies will form nearby. Gravity will pull them together until they merge to form a grander structure. This structure will itself merge, and so on. Hierarchical buildup is a characteristic feature of CDM models. The reason is simple. Because small-scale fluctuations in density are superimposed on larger-scale fluctuations, the density reaches its highest value over the smallest regions. An analogy is the summit of a mountain. The exact position of the peak corresponds to a tiny structure: for example, a pebble on top of a rock on top of a hill on top of the summit. If a cloud bank descends on the mountain, the pebble vanishes first, followed by the rock, the hill and eventually the whole mountain.

Similarly, the densest regions of the early universe are the smallest protogalaxies. They are the first regions to collapse, followed by progressively larger structures. What distinguishes CDM from other possible types of dark matter is that it has density fluctuations on all scales. Neutrinos, for example, lack fluctuations on small scales. A neutrino-dominated universe would be like a mountain with an utterly smooth summit.

The hierarchical formation of dark matter halos cannot be described using simple mathematical relationships. It is best studied using numerical simulations. To emulate a representative part of the universe with enough resolution to see the formation of individual halos, researchers must use the latest supercomputers. The statistical properties and spatial distribution of the halos emerging from these simulations are in excellent agreement with those of observed galaxies, providing strong support for the hierarchical picture and hence for the existence of CDM.

Take a Spin

THE HIERARCHICAL PICTURE naturally explains the shapes of galaxies. In spiral galaxies, stars and gas move on circular orbits. The structure of these galaxies is therefore governed by angular momentum. Where does this angular momentum come from? According to the standard picture, when protogalaxies filled the universe, they exerted tidal forces on one another, causing them to spin. After the protogalaxies collapsed, each was left with a net amount of angular momentum.

When the gas in the protogalaxies then started to cool, it contracted and started to fall toward the center. Just as ice skaters spin faster when they pull in their arms, the gas rotated faster and faster as it contracted. The gas thus flattened out, in the same way that the earth is slightly flatter than a perfect sphere because of its rotation. Eventually the gas was spinning so fast that the centrifugal force (directed outward) became equal to the gravitational pull (directed inward). By the time the gas attained centrifugal equilibrium, it had flattened into a thin disk. The disk was sufficiently dense that the gas started to clump into the clouds, out of which stars then formed. A spiral galaxy was born.

Because most dark matter halos end up with some angular momentum, one has to wonder why all galaxies aren't spirals. How did ellipticals come into being? Astronomers have long held two competing views. One is that most of the stars in present-day ellipticals and bulges formed during a monolithic collapse at early epochs. The other is that ellipticals are relative latecomers, having been produced as a result of the merging of spiral galaxies.

The second view has come to enjoy increasing popularity. Detailed computer simulations of the merger of two spirals show that the strongly fluctuating gravitational field destroys the two disks. The stars within the galaxies are too spread out to bang into one another, so the merging process is quite similar to the violent relaxation suffered by dark matter. If the galaxies are of comparable mass, the result is a smooth clump of stars with properties that strongly resemble an elliptical. Much of the gas in the two original disk galaxies loses its angular momentum and plummets toward the center. There the gas reaches high densities and starts to form stars at a frenzied rate. At later times, new gas may fall in, cool off and build up a new disk around the elliptical. The result will be a spiral galaxy with a bulge in the middle.

The high efficiency of star formation during mergers explains why ellipticals typically lack gas: they have used it up. The merger model also accounts for the morphology-density relation: a galaxy in a high-density environment will undergo more mergers and is thus more likely to become an elliptical.

Observational evidence confirms that mergers and interac-

tions have been common in the universe, particularly early on. In Hubble Space Telescope images, many ancient galaxies have disturbed morphologies, a telltale sign of interaction. Moreover, the number of starburst galaxies—in which stars form at a frenetic pace—increases dramatically at earlier times. Astronomers may be directly observing, for the first time, the formation of elliptical galaxies.

If elliptical galaxies and spiral bulges are linked to galaxy mergers, then it follows that supermassive black holes may be created in these events, too. Hole masses are strongly correlated with the mass of the surrounding elliptical galaxy or bulge; they are not correlated with the mass of the spiral disk. Merger models have been extended to incorporate supermassive holes and therefore AGNs. The abundant gas that is funneled toward the center during a merger could revive a dormant black

hole. In other words, quasars were more common in the past because mergers were much more common then.

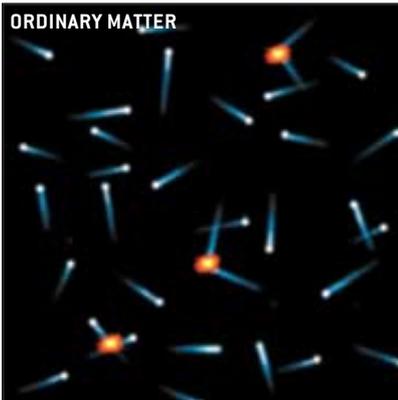
As for dwarf galaxies, in the hierarchical picture they are the leftovers—small clumps that have yet to merge. Recent observations show that star formation in dwarfs is particularly erratic, coming in short bursts separated by long quiescent periods [see “Dwarf Galaxies and Starbursts,” by Sara C. Beck; *SCIENTIFIC AMERICAN*, June 2000]. In heftier galaxies such as the Milky Way, star formation occurs at a more constant rate. These results are intriguing because astronomers have often hypothesized that the mass of a galaxy determines its fertility. In lightweight galaxies, supernova explosions can easily disrupt or even rid the system of its gas, thus choking off star formation. Even the smallest perturbation can have a dramatic effect. It is this sensitivity to initial conditions and random events

HOW RELAXING

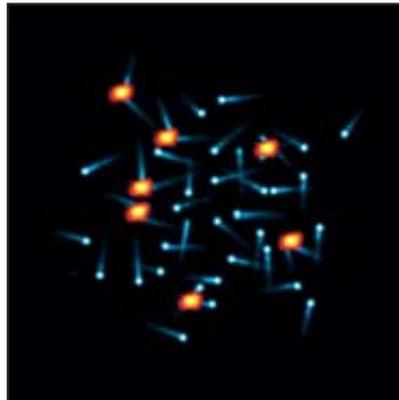
AN INTERNAL STATE OF EQUILIBRIUM is what makes a galaxy a distinct object rather than merely an arbitrary patch of space. This equilibrium determines the overall properties of the galaxy, such as

its shape and density profile. (An analogous equilibrium determines the size and temperature of stars.) The ordinary matter and dark matter attain equilibrium by different means.

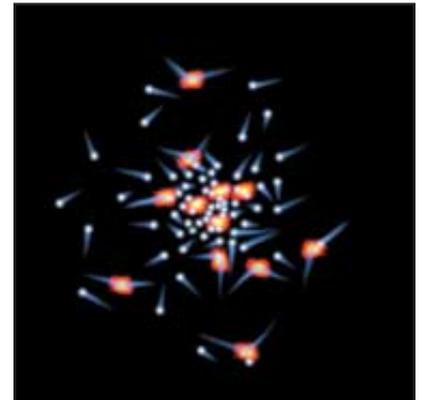
ORDINARY MATTER



1 The ordinary matter—predominantly hydrogen gas—starts off moving every which way. Its density varies randomly.

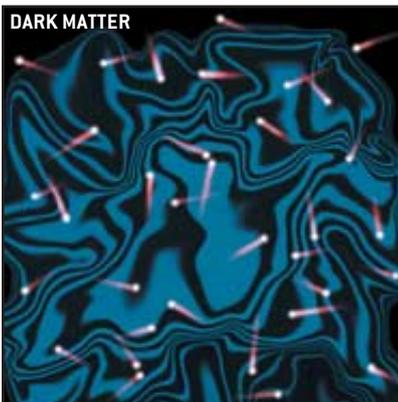


2 The gas particles bang into one another, redistributing energy and generating a pressure that resists gravity.

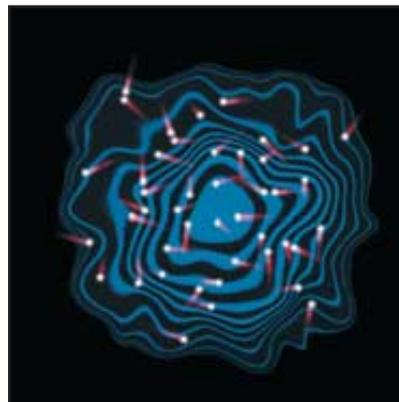


3 Eventually the gas settles down into hydrostatic equilibrium, with the density highest near the center of gravity.

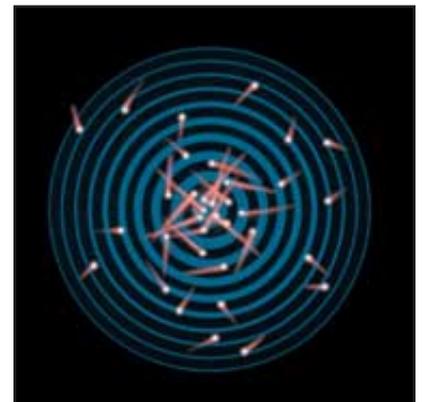
DARK MATTER



1 Initially the dark matter has the same arrangement as ordinary matter. The difference is that the particles do not collide.



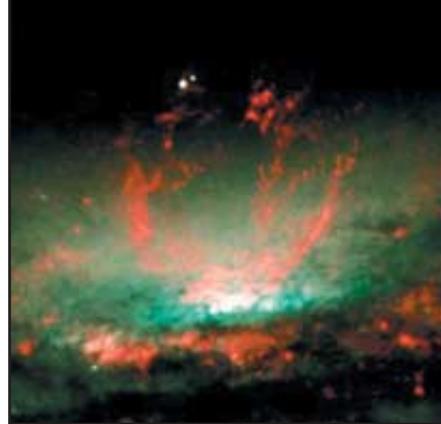
2 As the particles move around, the gravitational field changes, which causes particles to gain or lose energy.



3 Gradually the system settles down into virial equilibrium, in which the gravitational field no longer fluctuates.

Supernova explosions

could expel mass from low-mass galaxies so efficiently that hardly any stars would form.



that may account for the heterogeneity of the galactic dwarfs.

Although the standard picture of galaxy formation is remarkably successful, researchers are still far from working out all the processes involved. Moreover, they have yet to resolve some troubling inconsistencies. The simple picture of gas cooling inside dark matter halos faces an important problem known as the cooling catastrophe. Calculations of the cooling rates imply that the gas should have cooled briskly and pooled in the centers of halos, leaving intergalactic space virtually empty. Yet the space between galaxies is far from empty. Some extra input of energy must have prevented the gas from cooling down.

Some Feedback, Please

ANOTHER PROBLEM CONCERNS angular momentum. The amount of angular momentum imparted to protogalaxies in the models is comparable to the angular momentum that we actually see in spiral galaxies. So long as the gas retains its angular momentum, the CDM picture reproduces the observed sizes of spirals. Unfortunately, in the simulations the angular momentum leaks away. Much of it is transferred to the dark matter during galaxy mergers. As a result, the disks emerging from these simulations are a factor of 10 too small. Apparently the models are still missing an essential ingredient.

A third inconsistency has to do with the number of dwarf galaxies. Hierarchical theories predict a proliferation of low-mass dark matter halos and, by extension, dwarf galaxies. These are simply not seen. In the neighborhood of the Milky Way, the number of low-mass dwarfs is a factor of 10 to 100 lower than theories predict. Either these dark matter halos do not exist or they are present but have eluded detection because stars do not form within them.

Several solutions have been suggested for these problems. The proposals fall into two classes: either a fundamental change to the model, perhaps to the nature of dark matter [see “What’s the Matter?” by George Musser; News and Analysis, *SCIENTIFIC AMERICAN*, May 2000], or a revision of our picture of how the cooling gas is transformed into stars. Because most astronomers are reluctant to abandon the CDM model, which works so well on scales larger than galaxies, they have concentrated on improving the treatment of star formation. Current models gloss over the process, which occurs on scales that are much smaller than a typical galaxy. Incorporating it in full is far beyond the capabilities of today’s supercomputers.

Yet star formation can have profound effects on the structure of a galaxy [see “The Gas between the Stars,” by Ronald J. Reynolds; *SCIENTIFIC AMERICAN*, January]. Some astronomers

think that the action of stars might actually solve all three problems at once. The energy released by stars can heat the gas, obviating the cooling catastrophe. Heating also slows the descent of gas toward the center of the galaxy and thereby reduces its tendency to transfer angular momentum to the dark matter—alleviating the angular momentum problem. And supernova explosions could expel mass from the galaxies back into the intergalactic medium [see “Colossal Galactic Explosions,” by Sylvain Veilleux, Gerard Cecil and Jonathan Bland-Hawthorn; *SCIENTIFIC AMERICAN*, February 1996]. For the lowest-mass halos, whose escape velocity is small, the process could be so efficient that hardly any stars form, which would explain why we observe fewer dwarf galaxies than predicted.

Because our understanding of these processes is poor, the models still have a lot of wiggle room. It remains to be seen whether the problems really can be fixed or whether they indicate a need for a completely new framework. Our theory of galaxy formation will surely continue to evolve. The observational surveys under way, such as the Sloan Digital Sky Survey, will enormously improve the data on both nearby and distant galaxies. Further advances in cosmology will help constrain the initial conditions for galaxy formation. Already, precise observations of the cosmic microwave background radiation have pinned down the values of the large-scale cosmological parameters, freeing galactic modelers to focus on the small-scale intricacy. Soon we may unite the large, the small and the medium into a seamless picture of cosmic evolution. 

MORE TO EXPLORE

Cosmological Physics. John A. Peacock. Cambridge University Press, 1999.

The Formation of Ellipticals, Black Holes and Active Galactic Nuclei: A Theoretical Perspective. Guinevere Kauffmann, Stéphane Charlot and Martin G. Haehnelt in *Philosophical Transactions of the Royal Society of London*, Series A, Vol. 358, No. 1772, pages 2121–2132; July 15, 2000.

The Big Bang. Joseph Silk. W. H. Freeman and Company, 2001.

The Morphological Evolution of Galaxies. Roberto G. Abraham and Sidney van den Bergh in *Science*, Vol. 293, No. 5533, pages 1273–1278; August 17, 2001. Available at astro-ph/0109358

The Angular Momentum Content of Dwarf Galaxies: New Challenges for the Theory of Galaxy Formation. Frank C. van den Bosch, Andreas Burkert and Rob A. Swaters in *Monthly Notices of the Royal Astronomical Society*, Vol. 326, No. 3, pages 1205–1215; September 21, 2001. Available at astro-ph/0105082

New Perspectives in Astrophysical Cosmology. Martin Rees. Second edition. Cambridge University Press, 2002.

Galaxy Formation and Evolution: Recent Progress. Richard S. Ellis. Lecture given at the 11th Canary Islands Winter School of Astrophysics, “Galaxies at High Redshift” [in press]. Available at arxiv.org/abs/astro-ph/0102056