Was Cosmic Inflation the

Two thousand years after Lucretius proclaimed
that nothing can be created from nothing,
inflationary cosmology asserts that he was wrong.

WHEN AN OBSCURE RUSSIAN METEOROLOGIST named Alexander Friedmann proposed, in 1922, that the Universe might be expanding, Albert Einstein was sure that he was wrong. Five years earlier Einstein had published a static model of the Universe, and he was still convinced that it was correct. In a rare but dramatic blunder, Einstein bolstered his unfounded beliefs with an erroneous calculation, and fired off a note to the *Zeitschrift für Physik* claiming that Friedmann's theory violated the conservation of energy. Eight months later, however, after a visit from a colleague of Friedmann's, Einstein admitted his mistake and published a retraction. The equations of general relativity do, he conceded, allow for the possibility of an expanding universe.

Today the Big Bang theory, which began with Friedmann's calculations in 1922, has become the accepted view of cosmology. The expansion of the Universe was first observed in the early 1920s by Vesto Melvin Slipher, and in 1929 was codified by Edwin Hubble into what we now know as "Hubble's Law": on average, each distant galaxy is receding from us with a velocity that is proportional to its distance. In 1965 Arno Penzias and Robert Wilson detected a background of microwave radiation arriving at Earth from all directions—the afterglow of the primordial hot, dense fireball. Today we know, based on

‘Bang’ of the Big Bang?

data from the *Cosmic Background Explorer* (COBE) satellite (see *Beam Line*, Vol. 23, No. 3, Fall/Winter 1993), that the spectrum of this background radiation agrees with exquisite precision—to 1/30 of 1 percent—with the thermal spectrum expected for the glow of hot matter in the early Universe. In addition, calculations of nucleosynthesis in the early universe show that the Big Bang theory can correctly account for the cosmic abundance of the light nuclear isotopes: hydrogen, deuterium, helium-3, helium-4, and lithium-7. (Heavier elements, we believe, were synthesized much later, in the interior of stars, and were then explosively ejected into interstellar space.)

Despite the striking successes of the Big Bang theory, there is good reason to believe that the theory in its traditional form is incomplete. Although it is called the “Big Bang theory,” it is not really the theory of a bang at all. It is only the theory of the *aftermath* of a bang. It elegantly describes how the early Universe expanded and cooled, and how matter clumped to form galaxies and stars. But the theory says nothing about the underlying physics of the primordial explosion. It gives not even a clue about what banged, what caused it to bang, or what happened before it banged. The inflationary universe theory, on the other hand, is a description of the bang itself, and provides plausible answers to these questions and more.

A VERY SPECIAL BANG

Could the Big Bang have been caused by a colossal stick of TNT, or perhaps a thermonuclear explosion? Or maybe a gigantic ball of matter collided with a gigantic ball of antimatter, releasing an untold amount of energy in a powerful cosmic blast.

In fact, none of these scenarios can plausibly account for the Big Bang that started our Universe, which had two very special features distinguishing it from any typical explosion.

First, the Big Bang was far more homogeneous, on large scales, than can be explained by an ordinary explosion. In discussing homogeneity, however, I must first clarify that the Universe is in many ways conspicuously inhomogeneous. Palo Alto is very different from San Francisco, and the stars, galaxies, and clusters of galaxies are scattered through space in a lumpy, complex pattern. Cosmologically speaking, however, all this structure is small-scale. We can focus on the large scales, for example, by dividing space into cubes of 300 million light-years or more on a side. We would find that each such cube closely resembles the others in all its average properties, such as mass density, galaxy density, and light output. This large-scale uniformity can be seen in galaxy surveys, but the most dramatic evidence comes from the cosmic background radiation. Data from the COBE satellite, confirmed by subsequent ground-based observations, show that this radiation has the same temperature in all directions (after correcting for the motion of the Earth) to an accuracy of one part in 100,000.

> *Inflation is a wildfire that will inevitably take over the forest, as long as there is some chance that it will start.*

To see how difficult it is to explain this uniformity as the result of an ordinary explosion, we need to know a little about the history of the cosmic background radiation. The early Universe was so hot that electrons would have been ripped away from atoms, resulting in a plasma that filled space. Such a plasma is very opaque, so the photons that now make up the cosmic background radiation were constantly absorbed and re-emitted. After about 300,000 years, however, the Universe cooled enough for the plasma to form a gas of neutral atoms, which is very transparent. The photons of the cosmic background radiation have traveled on straight lines ever since, so they provide today an image of the Universe at an age of 300,000 years, just as the photons reaching your eye at this moment provide an image of the page in front of you. Thus, the observations of the cosmic background radiation show that the Universe was uniform in temperature, to one part in 100,000, at an age of several hundred thousand years.

Under many circumstances such uniformity would be easy to understand, since anything will come to a uniform temperature if left undisturbed for a long enough time. In the standard Big Bang theory, however, the Universe evolves so quickly that there is no time for the uniformity to be established. One can pretend, for the sake of discussion, that the Universe is populated by little purple creatures, each equipped with a furnace and a refrigerator, and each dedicated to the cause of creating a uniform temperature. Those little creatures, however, would have to communicate at roughly 100 times the speed of light if they are to achieve their goal of creating a uniform temperature across the visible Universe by 300,000 years after the Big Bang. Since neither sticks of dynamite nor balls of matter and antimatter can transmit their energy faster than light, they cannot account for the uniformity. The classical form of the Big Bang theory requires us to postulate, without explanation, that the primordial fireball filled space from the beginning. The temperature was the same everywhere *by assumption*, not as a consequence of any physical process. This shortcoming is known as the "horizon problem," since cosmologists use the word "horizon" to indicate the largest distance that information or energy could have traversed since the instant of the Big Bang, given the restriction of the speed of light.

The second special feature of the Big Bang, which is very difficult to imagine arising from a standard explosion, is a remarkable coincidence called the "flatness problem." This problem concerns the pinpoint precision with which the mass density of the early Universe must be specified for the Big Bang theory to agree with reality.

First, we need to review a little vocabulary. If the mass density of the Universe exceeds a value called the *critical density*, then gravity will be strong enough to reverse the expansion eventually, causing the Universe to recollapse into what is sometimes called the *big crunch*. If the mass density is less than the critical value, the Universe will go on expanding forever. The ratio of the actual mass density to the critical value is known to cosmologists by the Greek letter omega (Ω). General relativity implies that the geometry of the Universe is Euclidean only if omega is one, so an $\Omega = 1$ universe is called “flat” (see box on the right).

Omega is very difficult to determine, but it is safe to say that its present value lies somewhere in the range of 0.1 to 2. That seems like a broad range, but consideration of the time development of the Universe leads to a spectacularly different point of view. $\Omega = 1$ is an unstable equilibrium point of cosmological evolution, which means that it resembles the situation of a pencil balancing on its sharpened tip. The phrase equilibrium point implies that if omega is ever exactly equal to one, it will remain exactly equal to one forever—just as a pencil balanced precisely on end will, according to the laws of classical physics, remain forever vertical. The word unstable means that any deviation from the equilibrium point, in either direction, will rapidly grow. If the value of omega in the early Universe was just a little above one, it would have rapidly risen toward infinity; if it was just a smidgen below one, it would have rapidly fallen toward zero. For omega to be anywhere near one today, it must have been extraordinarily close to one at early times. For example, consider one second after the Big Bang, the time at which the processes related to Big Bang nucleosynthesis were just beginning. For omega to be anywhere in the allowed range today, at that time omega must have equaled one to an accuracy of 15 decimal places!

A simple explosion gives no explanation for this razor-sharp fine-tuning, and indeed no explanation can be found in the traditional version of the Big Bang theory. The initial values of the mass density and expansion rate are not predicted by the theory, but must be postulated. Unless we postulate that the mass density at one second just happened to have a value between 0.999999999999999 and 1.000000000000001 times the critical density, however, the theory will not describe a universe that resembles the one in which we live.

THE INFLATIONARY UNIVERSE

Although the properties of the Big Bang are very special, we now know that the laws of physics provide a mechanism that produces exactly this sort of a bang. The mechanism is known as cosmic inflation.

The crucial property of physical law that makes inflation possible is the existence of states of matter that have a high energy density that cannot be rapidly lowered. Such a state is called a “false vacuum,” where the word “vacuum” indicates a state of lowest possible energy density, and the word “false” is used to mean

Critical Mass Density and Flatness

THE CRITICAL MASS density ρ_c is related to the Hubble constant H by

$$\rho_c = \frac{3H^2}{8\pi G}$$

where G is Newton’s gravitational constant. The quantity Ω is defined by $\Omega \equiv \rho/\rho_c$, where ρ is the actual mass density. It is often assumed that the cosmological constant Λ introduced by Einstein is zero, in which case the Universe will recollapse if and only if $\Omega > 1$. If Λ is nonzero, the condition for recollapse is more complicated, but the equation above is still taken as the definition of ρ_c.

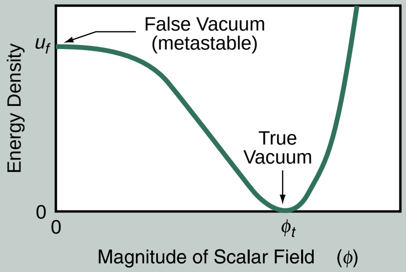
The spatial geometry of the Universe is determined by the quantity

$$\Omega + \frac{\Lambda}{3H^2}.$$

If this quantity exceeds one, the Universe curves back on itself to form a closed space of finite volume, but without boundary. In such a space the sum of the angles in a triangle would exceed 180 degrees, and a starship traveling on a straight line would eventually return to its point of origin. If the quantity above is less than one, the Universe is an open space in which triangles contain less than 180 degrees. If the quantity is exactly one, the space is Euclidean, which is also called flat.

Physics of the False Vacuum

THE FALSE VACUUM arises naturally in any theory that contains scalar fields, that is, fields that resemble electric or magnetic fields except that they have no direction. The Higgs fields of the Standard Model of particle physics or the more speculative grand unified theories are examples of scalar fields. It is typical of Higgs fields that the energy density is minimal not when the field vanishes, but instead at some nonzero value of the field. For example, the energy density diagram might look like



The energy density is zero if $\phi = \phi_t$, so this condition corresponds to the ordinary vacuum of empty space. In this context it is usually called the “true” vacuum. The state in which the scalar field is near $\phi = 0$, at the top of the plateau, is called the “false” vacuum. If the plateau of the energy density diagram is flat enough, it can take a very long time, by early Universe standards, for the scalar field to “roll” down the hill of the energy density diagram so that the energy can be lowered. For short times the false vacuum acts like a vacuum in the sense that the energy density cannot be lowered.

temporary. For a period that can be long by the standards of the early Universe, the false vacuum acts as if the energy density cannot be lowered, since the lowering of the energy is a slow process. The underlying physics of the false vacuum state is described in the box on the left.

The peculiar properties of the false vacuum stem from its pressure, which is large and negative (see box on the right). Mechanically such a negative pressure corresponds to a suction, which does not sound like something that would drive the Universe into a period of rapid expansion. The mechanical effects of pressure, however, depend on pressure differences, so they are unimportant if the pressure is reasonably uniform. According to general relativity, however, there is a gravitational effect that is very important under these circumstances. Pressures, like energy densities, create gravitational fields, and in particular a positive pressure creates an attractive gravitational field. The negative pressure of the false vacuum, therefore, creates a repulsive gravitational field, which is the driving force behind inflation.

There are many versions of inflationary theories, but generically they assume that some small patch of the early Universe somehow came to be in a false vacuum state. Various possibilities have been discussed, including supercooling during a phase transition in the early Universe, or a purely random fluctuation of the fields. A chance fluctuation seems reasonable even if the probability is low, since the inflating region will enlarge by many orders of magnitude, while the non-inflating regions will remain microscopic. Inflation is a wildfire that will inevitably take over the forest, as long as there is some chance that it will start.

Once a patch of the early Universe is in the false vacuum state, the repulsive gravitational effect drives the patch into an inflationary period of exponential expansion. To produce a universe with the special features of the Big Bang discussed above, the expansion factor must be at least about 10^{25}. There is no upper limit to the amount of expansion. Eventually the false vacuum decays, and the energy that had been locked in it is released. This energy produces a hot, uniform, soup of particles, which is exactly the assumed starting point of the traditional Big Bang theory. At this point the inflationary theory joins onto the older theory, maintaining all the successes for which the Big Bang theory is believed.

In the inflationary theory the Universe begins incredibly small, perhaps as small as 10^{-24} cm, a hundred billion times smaller than a proton. The expansion takes place while the false vacuum maintains a nearly constant energy density, which means that the total energy increases by the cube of the linear expansion factor, or at least a factor of 10^{75}. Although this sounds like a blatant violation of energy conservation, it is in fact consistent with physics as we know it.

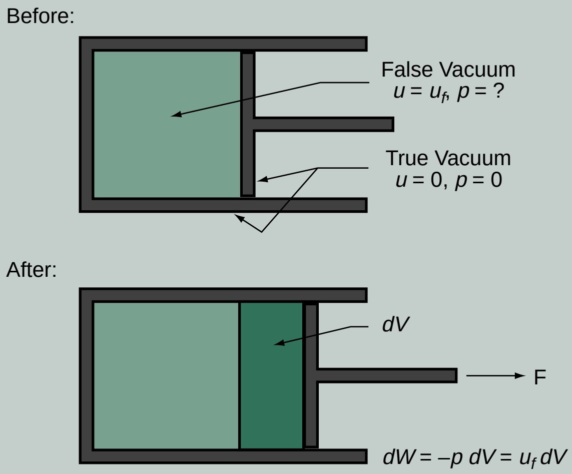
The resolution to the energy paradox lies in the subtle behavior of gravity. Although it has not been widely appreciated, Newtonian physics unambiguously implies that the energy of a gravitational field is always negative, a fact which holds also in general relativity. The Newtonian argument closely parallels the derivation of the energy density of an electrostatic field, except that the answer has the opposite sign because the force law has the opposite sign: two positive masses attract, while two positive charges repel. The possibility that the negative energy of gravity could balance the positive energy for the matter of the Universe was suggested as early as 1932 by Richard Tolman, although a viable mechanism for the energy transfer was not known.

During inflation, while the energy of matter increases by a factor of 10^{75} or more, the energy of the gravitational field becomes more and more negative to compensate. The total energy—matter plus gravitational—remains constant and very small, and could even be exactly zero. Conservation of energy places no limit on how much the Universe can inflate, as there is no limit to the amount of negative energy that can be stored in the gravitational field.

This borrowing of energy from the gravitational field gives the inflationary paradigm an entirely different perspective from the classical Big Bang theory, in which all the particles in the Universe (or at least their precursors) were assumed to be in place from the start. Inflation provides a mechanism by which the entire Universe can develop from just a few ounces of primordial matter. Inflation is radically at odds with the old dictum of Democritus and Lucretius, "Nothing can be created from nothing." If inflation is right, everything can be created from nothing, or at least from very little. If inflation is right, the Universe can properly be called the ultimate free lunch.

Pressure of the False Vacuum

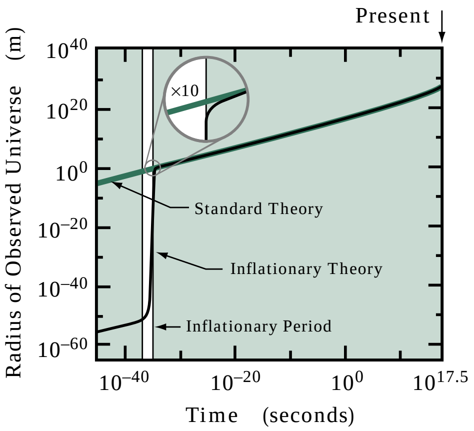
THE PRESSURE OF THE FALSE VACUUM can be determined by a simple energy-conservation argument. Imagine a chamber filled with false vacuum, as shown in the diagram below.



For simplicity, assume that the chamber is small enough so that gravitational effects can be ignored. Since the energy density of the false vacuum is fixed at some value u_f, the energy inside the chamber is $U = u_f V$, where V is the volume. Now suppose the piston is quickly pulled outward, increasing the volume by dV. If any familiar substance were inside the chamber, the energy density would decrease. The false vacuum, however, cannot rapidly lower its energy density, so the energy density remains constant and the total energy increases. Since energy is conserved, the extra energy must be supplied by the agent that pulled on the piston. A force is required, therefore, to pull the piston outward, implying that the false vacuum creates a suction, or negative pressure p. Since the change in energy is $dU = u_f dV$, which must equal the work done, $dW = -p dV$, the pressure of the false vacuum is glven by

$$p = -u_f .$$

The pressure is negative, and extremely large. General relativity predicts that the gravitational field which slows the expansion of the universe is proportional to $u_f + 3p$, so the negative pressure of the false vacuum overcomes the positive energy density to produce a net repulsive gravitational field.



The solution to the horizon problem. The green line shows the radius of the region that evolves to become the presently observable Universe, as described by the traditional Big Bang theory. The black line shows the corresponding curve for the inflationary theory. Due to the spectacular growth spurt during inflation, the inflationary curve shows a much smaller Universe than in the standard theory for the period before inflation. The uniformity is established at this early time, and the region is then stretched by inflation to become large enough to encompass the observed Universe. Note that the numbers describing inflation are illustrative, as the range of possibilities is very large.

INFLATION AND THE VERY SPECIAL BANG

Once inflation has been described, it is not hard to see how it produces just the special kind of bang that was discussed earlier.

Consider first the horizon problem, the difficulty of understanding the large-scale homogeneity of the Universe in the context of the traditional Big Bang theory. Suppose we trace back through time the observed region of the Universe, which has a radius today of about 10 billion light-years. As we trace its history back to the end of the inflationary period, our description is identical to what it would be in the traditional Big Bang theory, since the two theories agree exactly for all times after the end of inflation. In the inflationary theory, however, the region undergoes a tremendous spurt of expansion during the inflationary era. It follows that the region was incredibly small before the spurt of expansion began—10^{25} or more times smaller in radius than in the traditional theory. (Note that I am not saying that Universe as a whole was very small. The inflationary model makes no statement about the size of the Universe as a whole, which might in fact be infinite.)

Because the region was so small, there was plenty of time for it to come to a uniform temperature, by the same mundane processes by which a cup of hot coffee cools to room temperature as it sits on a table. So in the inflationary model, the uniform temperature was established before inflation took place, in an extremely small region. The process of inflation then stretched this region to become large enough to encompass the entire observed Universe. The uniformity is preserved by this expansion, because the laws of physics are (we assume) the same everywhere.

The inflationary model also provides a simple resolution for the flatness problem, the fine-tuning required of the mass density of the early Universe. Recall that the ratio of the actual mass density to the critical density is called omega, and that the problem arose because the condition $\Omega = 1$ is unstable: omega is always driven away from one as the Universe evolves, making it difficult to understand how its value today can be in the vicinity of one.

During the inflationary era, however, the peculiar nature of the false vacuum state results in some important sign changes in the equations that describe the evolution of the Universe. During this period, as we have discussed, the force of gravity acts to accelerate the expansion of the Universe rather than to retard it. It turns out that the equation governing the evolution of omega also has a crucial change of sign: during the inflationary period the Universe is driven very quickly and very powerfully *towards* a critical mass density. This effect can be understood if one accepts from general relativity the relationship between a critical mass density and the geometric flatness of space. The huge expansion factor of inflation drives the Universe toward flatness for the same reason that the Earth appears flat, even though it is really round. A small piece of any curved space, if magnified sufficiently, will appear flat.

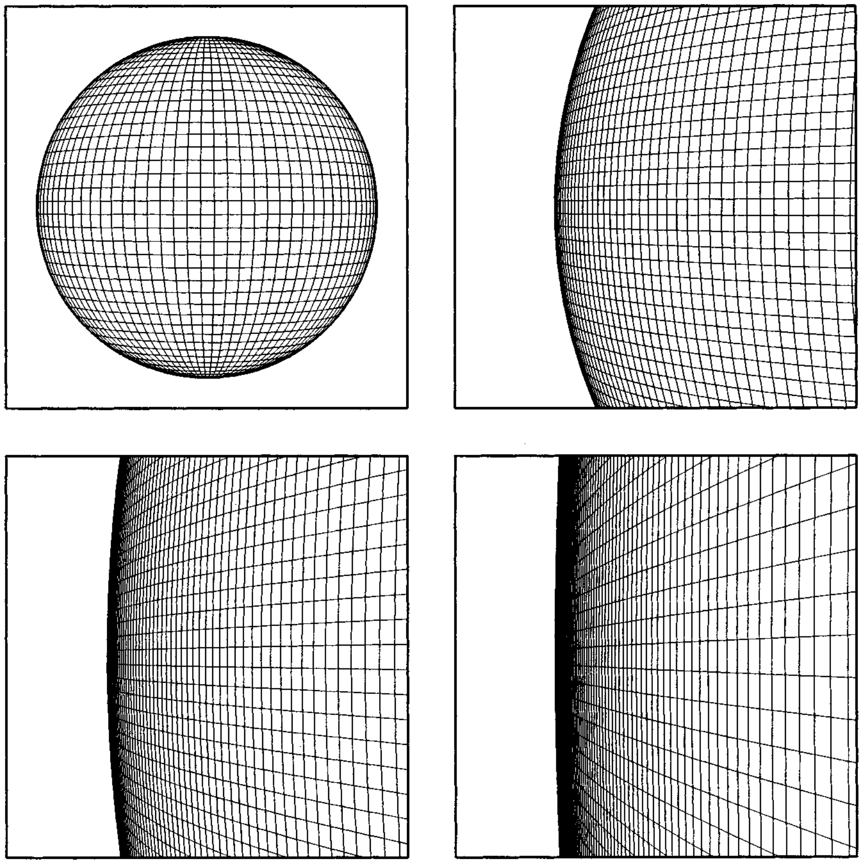
Thus, a short period of inflation can drive the value of omega very accurately to one, no matter where it starts out. There is no longer any need to assume that the initial value of omega was incredibly close to one.

Furthermore, there is a prediction that arises from this behavior. The mechanism that drives omega to one almost always overshoots, which means that even today the mass density should be equal to the critical value to a high degree of accuracy. (If Einstein's cosmological constant Λ is nonzero, this prediction is modified to become $\Omega + \Lambda/3H^2 = 1$, where H is Hubble's constant.) Thus, the determination of the mass density of the Universe could be a very important test of the inflationary model. Unfortunately, it is very difficult to reliably estimate the mass density of the Universe, since most of the matter in the Universe is "dark," detected only through its gravitational pull on visible matter. Current estimates of omega range from 0.2 to 1.1. Nonetheless, it is likely that this issue can be settled in the near future. The high precision measurements of the microwave background radiation that will be made by the Microwave Anisotropy Probe, scheduled for launch in about 2001, are expected to pin down the value of omega to about 5 percent accuracy.

THE CURRENT PICTURE

While it may be too early to say that inflation is proved, I claim that the case for inflation is compelling. It is hard to even conceive of an alternative theory that could explain the basic features of the observed Universe. Not only does inflation produce just the kind of special bang that matches the observed Universe, but quantum fluctuations during inflation could have produced nonuniformities which served as the seeds of cosmic structure. These nonuniformities can be observed directly in the cosmic background radiation, with an amplitude of about one part in 100,000. So far the measurements of the spectrum have been beautifully consistent with the predictions of inflation, although it must be admitted that nonuniformities created by cosmic strings are also consistent with the observations. Cosmic strings, however, cannot explain the large-scale homogeneity or the flatness of the Universe.

While the case for inflation is strong, it should be stressed that inflation is really a paradigm and not a theory. The statement that the Universe arose from inflation, if it is true, is not the end of the study of cosmic origins—it is in fact closer to the beginning. The details of inflation depend upon the details of the underlying particle physics, so cosmology and particle physics become intimately linked together. While I cannot see any viable alternative to the general idea of inflation, there is still much work to be done before a detailed picture is established. And I suspect that there is room for many new important ideas.



The expanding sphere illustrates the solution to the flatness problem in inflationary cosmology. As the sphere becomes larger, its surface becomes flatter and flatter. Similarly the inflation of space causes it to become geometrically flat, and general relativity implies that the mass density of a flat universe must equal the critical value.