HDC–3–90

The Universe:

The Ultimate Free Lunch

Victor J. Stenger

Department of Physics and Astronomy, University of Hawaii,

2505 Correa Rd., Honolulu, HI 96822, USA

To appear in *European Journal of Physics*

Abstract. It is commonly believed that the origin of the universe must have involved the violation of natural laws, particularly energy conservation and the second law of thermodynamics. Here it is shown that this need not have been the case, that the universe could have begun from a state of zero energy and maximum entropy, and then naturally evolved into what we see today without violating any known principles of physics. The fundamental particles and the force laws they obey then come about through a series of random symmetry-breaking phase transitions during the period of exponential expansion in the first fraction of a second after the universe appears as a quantum fluctuation.

1. Introduction

Can the origin of the universe be understood as a natural process? Recent work in particle physics and cosmology suggests that it can. Common sense seems to dictate that the universe could not have just happened, that some transcendent force violating known principles of physics was required to bring it into existence. The commonsense arguments are basically twofold:

(1) *The No Free Lunch Argument:*

"You can't get something from nothing."

(2) *The Argument from Design:*

"How could all of this (gesturing to the world around us) have happened by chance?"

Here I will show that the physical law equivalents of these two comm onsense statements need not have been violated to bring about the universe. By means of a random quantum fluctuation, the universe could have begun in maximum disorder, as an empty spacetime of arbitrary negative curvature that then exponentially inflates, spontaneously forming structure by breaking the symmetries that existed initially.

Clearly the origin of the universe is a subject of abiding general interest. Unfortunately, most of the recent developments that have provided enormous insight into this question are buried in the esoterica of theoretical particle physics and cosmology, making them inaccessible to all but those few trained in these fields. In a recent book, I tried to communicate some of these new ideas, in particular the notion of spontaneous origins, to the general reader (Stenger 1988). Here I try to strike a balance between the two extremes by using physics concepts that should be familiar to a wide audience, yet provide sufficient rigor to at least make the basic ideas clear.

2. The First and Second Laws of Thermodynamics

The No Free Lunch argument can be translated into the language of physics as follows. When we say "You can't get something from nothing," we are giving the vernacular version of the *first law of thermodynamics:* The total heat energy added to a system equals the increase in internal energy minus any work done by the system. This is equivalent to the principle of conservation of energy. Since the universe can reasonably be regarded as an isolated system, and since it now contains energy, then energy conservation must have been violated at some time — presumably at the beginning when that energy was created out of nothing.

The argument from design can similarly be associated with an apparent violation of the *second law of thermodynamics:* The total entropy of an isolated system cannot decrease with time. Thus the universe, according to this argument, cannot have begun in chaos — maximum entropy — without an act of creation violating the second law to produce order, decreasing the entropy of the universe.

Both of these principles, the first and second laws of thermodynamics, are believed to be such fundamental statements about the universe that they are expected to have been valid at or near the beginning of time. The violation of either could reasonably be defined as a supernatural event — if by supernatural we mean the transcending of natural law.

3. The Failure of Common Sense

Much of what constitutes commonsense knowledge of the universe had been systematically incorporated into natural law by the end of the nineteenth century. Three nineteenth century concepts strongly implied that the origin of the universe violated natural law:

(1) The universe was believed to be a *firmament*. The earth and other planets moved about the sun, but the sun and other stars were assumed essentially fixed in space.

(2) The second law appeared to require that the universe started out in a state of maximum order, or minimum entropy, and was evolving toward a final end of total chaos — the *Heat Death*. Thus there had to be original order at the beginning of the universe, a *Grand Design*.

(3) It was believed that matter could not be created or destroyed. Since matter exists, it must have been created supernaturally.

The revolutions in physics and astronomy that occurred in the early twentieth century turned these conclusions on their heads. There were three relevant developments:

(1) The universe is not a firmament. Rather it is expanding as if it began in an explosion, the *Big Bang*, 10 or 15 billion years ago. Since the maximum entropy of an expanding volume increases as the volume increases, it becomes possible to begin in a state of maximum entropy — total chaos — and still obtain the formation of local order in the universe without violating the second law as the universe expands.

(2) $E = mc^2$. Matter can be created and destroyed.

(3) Quantum mechanics showed that certain events, such as atomic transitions and nuclear decays, happen spontaneously. Accidents happen (including accidental violations of the first law). Everything that occurs in the universe is not pre-determined by natural law. In fact, our so-called laws just apply to ensembles of systems, only guaranteeing the statistical behavior of the ensemble while leaving the behavior of an individual system to the vagaries of chance.

If anything characterizes these three statements, it is that each violates our common sense as derived from our everyday observations about the world. Thus it should come as no surprise that they make possible a universe whose origin also fails to follow our commonsense prejudices.

4. Order from Chaos

Let me first demonstrate that the production of order in the universe is not in violation of the second law of thermodynamics, even in where the universe starts out in total chaos. The second law says that, for any closed system, the total entropy cannot decrease:

$$\Delta S = \sum_i \Delta S_i \geq 0 \tag{1}$$

where ΔS_i is the change in entropy of subsystem i. An individual subsystem can be ordered, $\Delta S_j < 0$, as long as the decrease in entropy is compensated by increases in other subsystems.

Entropy is defined as

$$S = k \ln\{\text{no. of states}\}$$

which, for N particles of the same type, will be

$$\begin{aligned} S &= k \ln\{(\text{no. of one-particle states})^N\} \\ &= kN \ln\{\text{a not-too-big number}\} \\ &\approx kN \end{aligned} \tag{2}$$

This provides a very simple definition of entropy adequate for cosmological problems. Working in units where $k = 1$ (temperature is measured in energy units), the entropy of a system of like particles is just the number of particles in the system:

$$S = N \qquad (k = 1) \tag{3}$$

In words, the more particles, the more disorder.

The current visible universe has about 10^{22} stars. Each star contains about $N_A M_* = 10^{57}$ nucleons (or quarks) and electrons, where M_* is the mass of the star and N_A is Avagadro's number. Since the stars constitute a large fraction of the visible mass, the entropy of the visible matter $S_V \approx 10^{79}$.

As we will now see, this is negligible compared to the entropy of the 3K microwave background. From Stefan's law, the total kinetic energy in the background is that radiated by a 3K black body of radius ct in a time t:

$$E_B = \sigma T^4 \, 4 \pi c^2 t^3 \approx 10^{84} \text{ eV} \tag{4}$$

where t $\approx$ 10 billion years is the age of the universe and σ is the Stefan–Boltzmann constant. Since the average photon kinetic energy kT $= 2.6 \times 10^{-4}$ eV, the entropy of the photon background is $S_B \approx 10^{87}$.

Since $S_B >> S_V$, the entropy of the universe is, for all practical purposes, given by the entropy of the relic photons (plus neutrinos and possible dark matter particles of comparable number) left over from the Big Bang. Since that entropy is so large, and the change in entropy required to order galaxies, stars, and planets is so small by comparison, the second law provides no problem for the production of tiny pockets of order in the universe.

For example, consider ordering the earth. This is done by energy from the sun in the form of photons with an average energy kT_S, where $T_S = 5000$ K. Infrared photons with average energy kT_E, where $T_E = 300$ K, are then radiated back in space. For every photon absorbed from the sun, 5000/300 = 17 are emitted back into space. Thus the earth loses 16 units of entropy, the sun one unit, and the two bodies become correspondingly more orderly while the rest of the universe becomes more disorderly in the amount of 17 units of entropy.

The earth absorbs 2.5×10^{36} photons from the sun each second, so in the 4–5 billion years $\approx 10^{17}$ seconds of the earth's existence the entropy of the universe has increased by $(17)(2.5 \times 10^{36})(10^{17}) \approx 10^{54}$. This is a tiny fraction of the total entropy of the universe, which we saw above is about 10^{87}.

5. The Entropy of the Early Universe

We are not so much interested in the current entropy as that of the early universe. Supposing the universe begins in total chaos, how is it possible for order ever to form, consistent with the second law? To answer this, we need to determine how the entropy of the universe changes as the universe expands. When the universe has a radius R, its entropy will be that of an expanding relativistic gas of total energy E. This will depend on the temperature of the gas, however the maximum entropy will be simply (Frautschi 1982)

$$S_U^{max} = \frac{E}{\epsilon_{min}} = \frac{E}{hc/\lambda_{max}} = \frac{2\pi RE}{hc} = RE \tag{5}$$

where $\epsilon_{min} = hc/\lambda_{max}$ is the minimum energy of an individual photon, $\lambda_{max} = 2\pi R$ is the corresponding maximum photon wavelength, and in the final step we switch to natural units: $\hbar \equiv h/2\pi = c = 1$.

6. The Planck Length, Time, and Mass

The smallest distance that can be operationally defined is the *Planck length*, $R_{PL} \approx 10^{-33}$ cm. Within a sphere of this radius, the DeBroglie/Compton wavelength of a particle equals the circumference of a black hole of the same mass.

$$2\pi R_{PL} = \frac{h}{mc} \qquad (6)$$

where, for a black hole, the rest energy equals the potential energy

$$mc^2 = \frac{Gm^2}{R_{PL}} \qquad (7)$$

and so, in natural units,

$$R_{PL} = \sqrt{G} \qquad (8)$$

In addition to the Planck length, we define the Planck time $t_{PL} = R_{PL}/c \approx 10^{-43}$ s and the Planck mass, as given by (7), $m_{PL} = 1/\sqrt{G} = 10^{19}$ GeV.

7. The Entropy of a Black Hole

For reasons that will become clear, we also need to know the entropy of a black hole of mass M and radius R. Basically, the entropy will be given by the number of particles, or equally probably quantum levels, that can fit inside (Bekenstein 1973, Hawking 1976). From an argument similar to that given above for an expanding relativistic gas (Frautschi 1982),

$$S_{BH}^{max} = \frac{Mc^2}{\epsilon_{min}} = \frac{Mc^2}{hc/\lambda_{max}} = \frac{Mc^2 2\pi R}{hc} = RM \qquad (9)$$

in natural units. Now, $Mc^2 = GM^2/R$ as in (7), so $M = R/G = R/R_{PL}^2$ and

$$S_{BH}^{max} = \frac{R^2}{R_{PL}^2} \qquad (10)$$

Since we cannot see inside a black hole, we have no way of obtaining information about what is inside. The only properties of a black hole that can be described are those measurable from the outside, like mass, size, angular momentum, and charge. Operationally, a black hole thus can have no internal structure so its inside must be total chaos. And if a black hole has no internal structure, it must have maximum entropy. It follows that the maximum entropy of a body of radius R is that of a black hole of the same radius, as given in equation (10).

8. Complete Chaos at the Planck Time

If the inside of a black hole is total chaos, then any system with some degree of order must have an entropy less than that of a black hole of the same size. From (10), the maximum entropy of a black hole of radius $R = R_{PL}$ is unity. From (5), the maximum entropy for an expanding sphere of relativistic gas of that radius will exceed unity unless its total energy $E \leq m_{PL}$. However, from the uncertainty principle, the uncertainty in the momentum of a particle confined to a region of Planck dimensions will be $\Delta p \geq 1/R_{PL} = m_{PL}$, so its energy cannot be less than something of the order of m_{PL}. Thus, at $R = R_{PL}$, the entropy of an expanding relativistic gas must equal that of a black hole of the same radius. Since the entropy of the universe is given by the number of relativistic particles regardless of epoch, the maximum entropy of the universe at any radius R can be written:

$$S_U^{max} = \frac{R}{R_{PL}} \tag{11}$$

The simple relations (10) and (11) make a profound point. Suppose we extrapolate back to the time when the universe was a sphere of radius R_{PL}. The entropy of the universe at that time was a great as it could possibly be, equal to that of a Planck-sized black hole. The universe then was in complete chaos, with no structure, design, or order. So instead of starting at a high level of order, as once was thought, we are forced to conclude that not only is a high level of order unnecessary at the origin of the universe, the opposite — maximum disorder — is *required* if the universe was ever as small as a Planck length. If the universe was created with any Grand Design, it would have had to come into being with a size greater than the Planck length. Thus, the creation model of the origin of the universe is reminiscent of the anti-evolutionist view that the earth was created with all its structure and living species in place. Unsurprisingly, it is similarly ad hoc.

At $R = R_{PL}$, the universe is operationally indistinguishable from a black hole. As Hawking has shown, quantum mechanics implies that a black hole is unstable, with a mean lifetime determined just by its mass M: $\tau = (M/m_{PL})t_{PL}$ (Hawking 1974). So we can view the universe beginning as a chaotic black hole of Planck dimensions that then explodes into an expanding relativistic gas at the Planck time. How can the order we clearly observe today then have resulted, consistent with the second law of thermodynamics? From (10) and (11), the entropy of the universe is less than maximum once $R > R_{PL}$. As soon as the universe becomes an expanding gas, order can form without violating the second law. Although, from (11), the entropy of the expanding relativistic gas from the initial explosion increases with R, the maximum allowable entropy of the universe, from (10), is increasing faster as R^2, allowing room for structure to form (Frautschi 1982). Since the Planck time, the universe has expanded 61 orders of magnitude in radius, and so now has an entropy at least 61 orders of magnitude below its maximum possible value.

The structure of the universe, at the most fundamental level, is described by the properties of the elementary constituents of matter and the force laws that govern how those constituents interact. We will now see how this structure can come about as a series of symmetry-breaking spontaneous phase transitions during the first fraction of a second of the expansion. In the process, we will see that the first law of thermodynamics is also not violated.

9. The Universe of Matter and Radiation

Suppose the universe to be an expanding sphere of uniform mass density ρ. In general relativity, space itself is expanding and the radius of the sphere R is a scale factor that multiplies all distances. However we can avoid the formalism of general relativity and get many of the same results from a suitably modified Newtonian approach that I will use here. Also, since all reference frames are equivalent, we can work in the egocentric one in which we earthlings are the center of the universe.

Let us neglect all except radial motions. Consider a body at a distance r. The Newtonian equation of motion for the body can be written:

$$\frac{\ddot{r}}{r} = -\frac{4}{3}\pi G\rho \tag{12}$$

Since ρ is the fourth component of a four-vector, (12) is not relativistically invariant. In general relativity, gravity couples to momentum as well as mass-energy through the stress-energy tensor and (12) is modified by replacing ρ with the invariant density, $\rho + 3P$, where P is the pressure in units where the speed of light $c = 1$. Then

$$\frac{\ddot{r}}{r} = -\frac{4}{3}\pi G(\rho + 3P) \tag{13}$$

When non-relativistic matter dominates, $P = 0$ and we get Newton's result (12). When the matter becomes extreme relativistic (speed $v \rightarrow c$), people unfortunately rename it "radiation," like photons, for which the equation of state is $P = \rho/3$. In that case,

$$\frac{\ddot{r}}{r} = -\frac{8}{3}\pi G\rho \tag{14}$$

That is, extreme relativistic matter (like light) interacts via gravity with twice the strength as non-relativistic matter.

We are primarily interested in the early universe, where all matter is extreme relativistic and we can expect (14) to apply. In that case, the total energy of a body of relativistic mass m at a distance r will be

$$E = mc^2 - \frac{8}{3}\pi mG\rho r^2 \tag{15}$$

Recall that mc^2 contains both rest and kinetic energy in general, but the rest energy is negligible when the particles are extreme relativistic.

The body will have zero total energy when the universe has the critical density

$$\rho_c = \frac{3H^2}{8\pi G} \tag{16}$$

where $H = v/r$ is the Hubble constant and $v = c$ in this case. Since we considered any arbitrary body interacting with the rest of the universe, an exact balance between the total kinetic and gravitational potential energies of the universe will occur when $\rho = \rho_c$.

10. The Empty Universe

One of the curiousities of Einstein's General Theory of Relativity is the *cosmological constant.* In Newtonian gravity, the stars cannot remain eternally at fixed points in space (for a finite universe, and Olber's paradox rules out an infinite firmament). Eventually everything must collapse to a point. Einstein found that his equations allowed for an additional repulsive term, which he used to balance the attraction of gravity and provide for the firmament of stars. When Hubble later discovered that the universe was not a firmament, the cosmological constant was put aside. Einstein never liked it anyway, since it seemed to violate Mach's Principle, and called it his "biggest blunder."

Interest in the cosmological constant has been revived in today's investigations of the early universe. Its role can best be seen from a very simple argument that does not require any of the technical detail of relativity. In general relativity, gravity results from the curvature of spacetime. For cosmological purposes, this curvature can be represented by the quantity $-\ddot{R}/R$, where R is the scale factor that multiplies distances as the universe expands. In the usual case, the presence of matter results in positive curvature, as near a star or as with the homogenous fluid in equation (13), which can be applied for $r = R$. However, a perfectly good solution of Einstein's equations is an empty universe with constant spacetime curvature,

$$\frac{\ddot{R}}{R} = \frac{\Lambda}{3} \tag{17}$$

where Λ is the cosmological constant that Einstein originally added to (13) to provide the repulsion needed to cancel the attraction of normal gravity and give a steady state universe. In the empty universe, the cosmological constant acts alone.

In that case, we can solve (17) to give the time dependence of R:

$$R(t) = R_o e^{Ht} \tag{18}$$

where $H = \dot{R}/R = (\Lambda/3)^{1/2}$ is the Hubble constant (though not the current one). So an empty (de Sitter) universe with negative constant curvature (positive Λ) will exponentially inflate with a Hubble constant determined by the cosmological constant.

As the volume of space expands, the total energy increases. This would seem to violate energy conservation; however it does not. All that is required is for no heat to be added or subtracted for an isolated system, that is, the process of expansion must be adiabatic. From the first law of thermodynamics, the increase in the internal energy for an adiabatic expansion is equal to the work done on the system:

$$dE = -PdV \tag{19}$$

Thus $P = -dE/dV = -\rho$, where ρ is the energy density (equal to the mass density in units where $c = 1$). Going back to (13) and substituting $P = -\rho$, we see that we get (17) with $\Lambda = 8\pi G\rho$. So a vacuum of negative spacetime curvature has a positive energy density but *negative* pressure. This is called the *false vacuum.* Normally we think of an expanding gas with positive pressure doing work on the outside world, lowering its internal energy if no heat is added. Here we have curved empty spacetime with constant negative pressure doing work on itself, adiabatically raising its own total internal energy as it expands. As we see, this does not violate the first law of thermodynamics. We can get something from nothing, if by nothing we mean no matter or radiation.

11. Starting the Universe from Nothing

So can we understand how the universe could have started from nothing? First we must free ourselves from the instinct of looking for a causal explanation for everything. If we extrapolate back in time to the Planck time, $t_{PL} = 10^{-43}$ second, the universe was within the Planck radius, $R_{PL} = 10^{-33}$ centimeter. We have seen that this was *required* to be a situation of maximum chaos. As such, it could not have been the result of any causal process; or if it was, all the memory of that causal process would have been wiped out and the situation would be indistinguishable from one that is purely spontaneous. So the universe had to have begun as a random fluctuation at the Planck time. The alternative theory is that it was created with a Grand Design after the Planck time, but this is ad hoc and can be ruled out by the law of parsimony: the supernatural creation of the universe is a hypothesis not required by the data.

Now in a time interval $\Delta t = t_{PL}$, the uncertainty principle, $\Delta E \Delta t \geq \hbar/2$, implies that the initial energy of the universe E_o will have a (positive or negative) random value chosen from a normal distribution with a mean of zero and standard deviation $\Delta E = m_{PL}/2$. Thus we expect that, by accident, E_o will be of the order of m_{PL}. The initial energy density ρ_o then is of the order $\pm m_{PL}/R^3_{PL} = \pm m^4_{PL}$. Since the universe is empty, this energy density is contained in the curvature of spacetime, or, equivalently, we can say that the universe began with a random cosmological constant Λ of the order of $\pm 8\pi G m^4_{PL}$. Suppose Λ was accidentally positive. Then the universe exponentially inflated according to $R = e^{Ht}$, where the Hubble factor $H = (\Lambda/3)^{1/3}$.

Now let me indulge in a little further magic. The Planck radius determines our basic units of distance, time, and mass. Thus we can choose $R_{PL} = t_{PL} = \sqrt{G} = 1$, so Newton's constant $G = 1$ and the Planck mass $m_{PL} = 1$. I will call these *Planck units* ($\hbar = c = k = G = 1$). The advantage of working in Planck units is twofold. First it is simpler, with fewer symbols to carry around. But more important, it manifestly specifies that the value of G, like $\hbar$, c and k, is arbitrary. We do not need to find some grand principle to calculate the value of G. It has the only value it can have. As we will next see, the only quantity that it not arbitrary at the start of the universe is Λ, and that is random but required to be of the order of unity in Planck units.

Later I will discuss the mechanism by which inflation stops, but for now let me assume that it does so at some time t'. At that time the volume of the universe will have expanded by a factor $e^{3Ht'}$ and so the energy contained in the volume will have increased by the same factor.

This energy was available to produce the matter of the universe. If, as is now commonly believed, the current mass density of the universe is equal to the critical density given by (16), which is 10^{-30} gcm^{-3} using the current value of the Hubble constant (around 10^{-61} in Planck units), and if the radius of the universe is 10 billion light-years, then the total current mass is 10^{55} grams, corresponding to an energy of 10^{88} eV = 10^{60} in Planck units. If the universe started out with unit energy and unit radius, then it must have gone through 20 orders of magnitude of radial expansion before stopping at a radius of $10^{20}R_{PL} = 10^{-13}$ cm. Hey! That's the radius of the proton! This is an interesting coincidence, but I am not sure how profound it is. It says that, if you ask what radius a sphere must have to contain all the mass of the universe at the Planck density, you get the proton radius. In any case, $e^{Ht'} = 10^{20} = e^{46}$, and since $H \approx (8\pi/3)^{1/3} = 3$ in Planck units, $t' \approx 15$ Planck times or around 10^{-42} second.

12. The Inflaton Field

We have seen how a spontaneously-appearing empty universe can inflate just as the result of the vacuum energy density it initially accidentally contains by virtue of the uncertainty principle. And, if it is allowed to inflate to the size of a proton, sufficient energy will have been

been pumped into the system by the work done on itself by the negative pressure associated with this vacuum energy density, to create all the matter in the universe. But, why doesn't the universe just continue to inflate, remaining empty of matter forever? Some mechanism must exist to stop inflation and produce the particles and forces of nature. Let us proceed to discuss that mechanism.

Since the energy density of the vacuum is associated with the scalar (that is, one-component) cosmological constant Λ, we can think of it as being carried by a scalar field ϕ. This is called the *inflaton* field. As a first approximation, we assume the field is real and classical, with quantum effects introducing statistical fluctuations. A classical scalar field can be treated as a unit mass particle with a Lagrangian density:

$$\ell = \frac{1}{2} (\dot{\phi})^2 - u(\phi) \tag{20}$$

where $u(\phi)$ is the potential energy density that describes the self-interaction of ϕ. Multiplying by the volume to get the Lagrangian, and using Lagrange's equations (making sure to take into account the expansion of the universe: volume $\sim R^3$), we get the equation of motion:

$$\ddot{\phi} + 3H\dot{\phi} = - \frac{\partial}{\partial\phi} u(\phi) \tag{21}$$

where $H = \dot{R}/R$ is the Hubble constant, as before. Thus the expansion of the universe adds a friction term $3H\dot{\phi}$ to the equation of motion of the inflaton field.

So, depending on the functional form of $u(\phi)$, ϕ will evolve with time. And, if and when the energy density $\rho = \frac{1}{2}(\dot{\phi})^2 + u(\phi)$ goes to zero, the inflation described by (18), where $H = (8\pi G\rho/3)^{1/2}$, will stop.

To get some idea of the form of $u(\phi)$, let us consider the example of a ferromagnet. In the Landau theory of phase transitions, the free energy density of a system like a ferromagnet can be written:

$$f = \frac{1}{2} \alpha M^2 + \frac{1}{4} \beta M^4 \tag{22}$$

where M is the magnetization, or more generally, the *order parameter.* There is nothing very profound about (22); this is what you would get for the first two terms in a Taylor expansion around $M = 0$. The coefficient α is assumed to have a temperature dependence, $\alpha = a(T-T_c)$, so that the free energy is minimum at $M = 0$ when $T > T_c$. When the temperature drops below T_c, α becomes negative and the phase transition to ferromagnetism occurs with minimum free energy now at a non zero value of $M = \pm(-\alpha/\beta)^{1/2}$.

Suppose that the free energy of the inflaton field has some similar properties to the magnetic field of a ferromagnet. That is, the self interaction of the field contains short range correlations that tend to minimize the free energy at a field value other than zero. Assuming the same form for $u(\phi)$, which can be interpreted as the free energy density with order parameter ϕ, we have

$$u(\phi) = u_o + \frac{1}{2} \alpha \phi^2 + \frac{1}{4} \beta \phi^4 \tag{23}$$

where I have added a constant term at $\phi = 0$ to correspond to the initial state of the universe with the energy density $\rho_o = u_o$, which we have seen is of the order of unity in Planck units. If, like the ferromagnet, we are below the transition temperature where α is negative, we will have the potential energy density shown in Fig. 1, where the minima are at $\phi = \pm\ \sigma$, with $\sigma = (-\alpha/\beta)^{1/2}$.

The form (23) will hold for small deviations of ϕ from zero for any analytic even function of ϕ. The function must be even if ϕ is scalar, but otherwise the shape is arbitrary and σ is another accidental parameter, like ρ_o. The value it happens to have, gave us the universe we have. Another value would have given another universe, and perhaps did.

The spontaneously empty universe starts out at the top of the potential hill with no order, $\phi = 0$, but with random energy of the order of m_{PL} contained in its spacetime curvature so it begins to exponentially inflate. However the situation is not a stable one. Quantum fluctuations send the universe rolling down one side of the hill or the other toward one of the valleys at $\phi = \pm\sigma$. As ϕ evolves, the energy density ρ decreases so the inflation is no longer a simple exponential, but the universe still grows enormously with the friction term in (21) acting to slow the fall to equilibrium.

Near the bottom of the hill, ϕ will oscillate and the universe will continue to inflate until ϕ is brought to rest by friction. This will eventually happen from the friction due to expansion alone. However this just produces a universe filled with a uniform scalar field. Some additional mechanism must have existed — one that resulted in the production of the matter of the universe.

12. Spontaneous Symmetry Breaking

The ferromagnet is an example of *spontaneous symmetry breaking*, in which the system relaxes to a state that does not preserve the symmetry of the underlying dynamics. Broken symmetry is order. Thus order appears, as measured by the order parameter M, where none previously existed. That order is accidental, since all orientations of magnetization are initially equally likely. When the ferromagnet cools below the critical temperature, an arbitrary direction in space is randomly selected for the direction of magnetization.

Now obviously the early universe was not a ferromagnet and this is just an analogy. What is the actual nature of the inflaton field ϕ and what does it have to do with particle production and the spontaneous generation of the "laws of nature?" To explain, I must finally connect the discussion to recent developments in particle physics — a subject I have been able to studiously avoid so far.

Currently all experimental data are consistent with the *Standard Model* of elementary particles and forces. In this model, interactions between quarks and leptons are mediated by vector (spin 1) particles called *gauge bosons:* the photon, W, Z, and gluon. The gauge bosons are intrinsically massless, but gain mass through their interaction with a background scalar field, called the Higgs field, which spontaneously breaks the underlying symmetry (Higgs 1964, 1966). This results in short range forces, since the range of an interaction is inverse to the mass of the exchanged boson. The Higgs mechanism has been applied with greatest success in the Glashow–Weinberg–Salam unification of the infinite range electromagnetic interaction with the weak interaction, whose range is a fraction of a proton diameter (Glashow 1961, Salam 1968, Weinberg 1967).

In 1979, Linde proposed that the phase transition associated with the spontaneous symmetry breaking process could have cosmological implications — leading to a time dependence of particle masses, coupling constants, and the cosmological term (Linde 1979). In 1980, Kazanas argued that the vacuum energy of the early universe could lead to an exponential expansion in the way I have outlined here, resulting in a symmetry breaking phase transition. He suggested that this could explain the fact that the universe is far more homogeneous than can be explained by the normal Big Bang expansion (Kazanas 1980). Shortly thereafter, Guth recognized the fuller implications of these developing ideas and showed how exponential inflation explains why the universe is now so flat and solves a number of other problems associated with the standard Big Bang (Guth 1981). Some problems with Guth's original scenario were resolved by the New Inflationary Model independently proposed in 1982 by Linde in the Soviet Union (Linde 1982) and Albrecht and Steinhardt in the U.S. (Albrecht and Steinhardt 1982). Alternate ways in which inflation could have been triggered randomly from nothingness were proposed by Vilenkin (1983) and Linde (1983). A fuller discussion of these topics can be found in any number of reviews (e.g., Guth and Steinhardt 1984, Unruh and Semenoff 1986, Kolb et al. 1986).

In the original inflationary models, the scalar field ϕ was taken to be the Higgs field specifically associated with Grand Unification Theories (GUT) that attempted to bring the strong nuclear interaction into the unification scheme (Georgi and Glashow 1974, Georgi 1981). However the simplest GUT, minimal SU(5), was falsified by the failure of its prediction of the proton lifetime and strong–electroweak unification is no longer fashionable as unification schemes, such as superstrings, attempt to include gravity in the picture. At this writing, the exact structure of the inflaton field is still unknown, however it can be shown that the form

(23) results for Higgs–type fields from general considerations not limited to any particular unification scheme (Linde 1979, Sher 1988).

13. The Production of Order

The association of the inflaton field ϕ with a Higgs–type field, not necessarily that of GUT, is a reasonable one and allows us to at least qualitatively understand how both matter and the "laws of nature" came about spontaneously. There are basically two types of laws of physics: conservation principles and force laws. Conservation of energy, momentum, angular momentum, charge, and other quantities derive from fundamental symmetries of spacetime and the inner dimensions that account for certain degrees of freedom such as spin, charge, and baryon number. These would exist as basic symmetries when spacetime and the inner dimensions first spontaneously appear.

Rather than representing order, symmetry principles actually correspond to a state of maximum disorder; they describe situations where no particular axis is preferred and thus a system has no structure. Order is not symmetry — order is broken symmetry. It occurs as the result of a phase transition from more symmetric but less orderly states, as with the freezing of a cloud of water vapour into a six–pointed snowflake. And force laws result from broken symmetry. Thus Newton's law of gravity describes how space translation symmetry, or momentum conservation, is broken as a body falls to earth.

Complex and unique structures exist in the universe because of the variety of fundamental particles from which they are assembled, and the variety of the forces between these particles. Starting out alike in the early symmetric and chaotic universe, diverse properties result as the universe cools and passes through a series of symmetry–breaking phase transitions.

At the Planck time there is maximum disorder, so the order parameter ϕ is initially zero, give or take quantum fluctuations. By analogy with the ferromagnet, and from our association of ϕ with the Higgs field, $u(\phi)$ can be assumed to be of the form shown in Fig. 1. Particle production then occurs when the ϕ field oscillates around the potential minimum before coming to rest, in a way analogous to the production of spin waves in a magnet by the oscillation of the magnetization about equilibrium. These are called Goldstone oscillations. And just as the oscillation of an electromagnetic field can produce photons, Goldstone oscillations will produce massless vector bosons. These vector bosons then gain mass through the Higgs mechanism which results from the non–zero equilibrium value of the background scalar field $\phi = \pm\sigma$ that is left when the oscillations cease. Once the vector bosons exist, they can interact producing pairs of fermions — quarks and leptons or some precursors — which then interact with one another via vector boson exchange.

At this point the universe is no longer in the state of chaos that initially existed. Order exists, as manifested by the non–zero value of ϕ. This order, like that of the ferromagnet, was accidental, resulting from the spontaneous breaking of the original symmetry of the system.

Thus we imagine the complex array of particles and forces of nature appearing — by chance — through a series of phase transitions in the early universe, as the universe is heated by particle production, cooled by expansion below the critical temperature for the next phase transition, and then reheated again for the next stage. Finally the sequence stops with the quarks, leptons and forces we know today, the curvature becoming negligible and the linear expansion of the Big Bang takes over. Because the symmetry between matter and antimatter that initially existed was broken by one of the early phase transitions, the annihilation of particles and antiparticles into the photons now part of the microwave background was not perfect, leaving a small reside of one part in a billion quarks and electrons that then stuck together in the clumps that we call galaxies, stars, planets, rocks, trees, and people.

The author is grateful for the helpful comments of X. Tata and S. Pakvasa.

References

Albrecht A and Steinhardt P J 1982 *Phys. Rev. Lett.* **48** 1220.

Bekenstein J D 1973 *Phys. Rev.* D7 2333.

Institute Slansky R ed. (World Scientific).

Frautschi Steven 1982 *Science* **217** 593 (1982).

Georgi H and Glashow S L 1972 *Phys. Rev. Lett.* **32** 438.

Georgi H 1981 *Sci. Am.* **244** 48.

Glashow SL 1961 *Nucl. Phys.* **22** 579.

Guth A H 1981 *Phys. Rev.* D23 347–356.

Guth A H and Steinhardt PJ 1984 *Sci. Am.* **250** 116.

Hawking S W 1974 *Nature* **248** 30.

Hawking S W 1976 *Phys. Rev.* **D13** 191.

Higgs P W 1964 *Phys. Rev. Lett.* 13 508

—— 1966 *Phys. Rev.* **145** 1156.

Kazanas Demosthenes 1980 *Astrophys. J.* **241** L59–L63.

Kolb E W, Turner M S, Lindley D, Olive K, and Seckel D 1986 (eds.) *Inner Space/Outer Space* (Chicago: University of Chicago Press).

Linde A D 1979 *Rep. Prog. Phys.* **42** 289–437.

Linde A D 1982 *Phys. Lett.* **108B** 389.

Linde AD 1983 *Phys. Lett.* 129B 177–181.

Linde A D 1987 *Physics Today* September 61–68.

Salam A 1968 *Elementary Particle Theory* ed. N. Svartholm (Stockholm: Almqvist and Wiksells) p. 367.

Sher M 1988 Washington University St. Louis report

WU–TH–88–8 to be published in *Phys. Rep.*

Stenger V J 1988 *NOT BY DESIGN: The Origin of the Universe* (Buffalo NY: Prometheus Books).

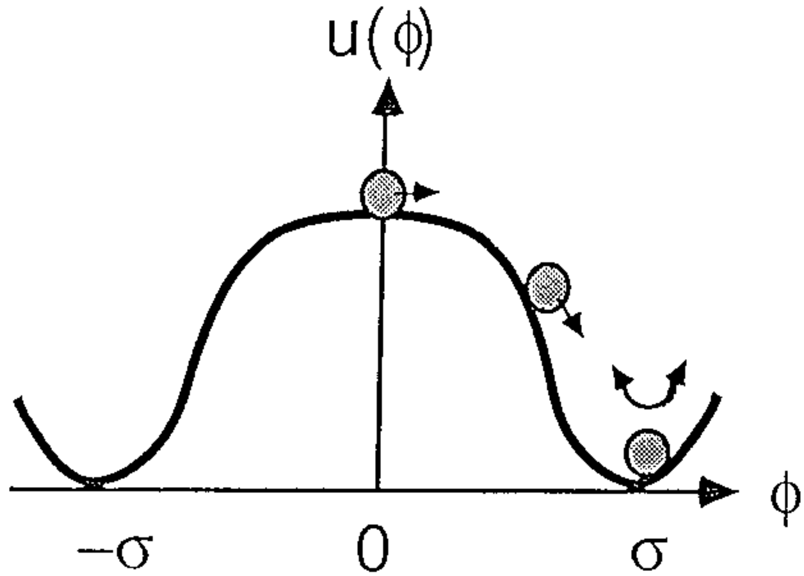
Unruh W G and Semenoff G W eds. 1986 *The Early Universe* (Dordrecht Holland: D Reidel Publishing Co.).

Vilenkin A 1983 *Phys. Rev.* D27 2848.

Weinberg S 1967 *Phys. Rev. Lett.* **19** 1264.

Figure Caption

1. The universe starts out empty with zero order, $\phi = 0$, and an energy density of the order of m_{PL}^4. This corresponds to a negative spacetime curvature, or negative pressure, and the universe exponentially inflates. Since the free energy is not minimum, the universe is in unstable equilibrium. A quantum fluctuation then sends it rolling down one side of the potential energy density hill or the other toward a minimum at $\phi = \pm\sigma$. Before stopping at the bottom, it oscillates about the minimum, dissipating its energy in the form of particle production. Since ϕ is an order parameter, $\phi = \pm\sigma$ corresponds to a broken symmetry. By the Higgs mechanism, the background field can give mass to vector bosons resulting in short range forces.



u(ϕ)
ϕ
−σ
0
σ