PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Theories of structure formation in the universe

Neil Turok

Phil. Trans. R. Soc. Lond. A 1999 **357**, 77-92
doi: 10.1098/rsta.1999.0315

THE ROYAL SOCIETY

Theories of structure formation in the universe

By Neil Turok

Department of Applied Mathematics and Theoretical Physics, Silver St, Cambridge CB3 9EW, UK

There is currently great hope that imminent cosmological observations will establish a fundamental theory of the origin of structure in the universe. The two leading candidate theories are inflation and cosmic defects. The simplest critical-density inflation models have been known to be problematic for some time. Recent calculations are reported which appear likewise to rule out the simplest cosmic defect models. For inflation there are possible variants, namely universes which have non-zero curvature or a cosmological constant, which may yet save the day. In this paper, I review recent work with Hawking which suggests how an open inflationary universe may be obtained from a theory with a generic inflaton potential. I also discuss a potential solution to the cosmological constant problem in which the cosmological constant would be non-negligible today.

Keywords: cosmology; structure formation; topological defects; inflation; quantum cosmology; early universe

1. Introduction

We are all very excited about the planned satellite missions to map the microwave sky, and forthcoming large-scale galaxy surveys. They will produce data-sets of unprecedented precision with which we can test theories of the early universe, and especially of cosmic structure formation.

For some time there have been two contending theories. Inflation (Guth 1981; Linde 1982; Albrecht & Steinhardt 1982) is the more ambitious of the two, seeking to explain how the whole visible universe was generated along with its inhomogeneities. The theory provides a remarkable link between quantum mechanics, which causes fluctuations of the inflaton field, and the approximately scale-invariant large-scale structure we observe in today's universe (Guth & Pi 1982; Hawking 1982; Starobinski 1982; Bardeen *et al*. 1983). The fluctuations predicted by inflation are typically Gaussian distributed and 'coherent', with associated Doppler peaks in the cosmic microwave anisotropy power spectrum (see, for example, Steinhardt 1995; Albrecht 1997). Verifying these signatures will be a major step towards confirming inflation.

However, inflation rests on somewhat shaky theoretical foundations. After almost two decades of model building (for a review see Lyth & Riotto 1998) inflationary models remain contrived, with ad hoc fields and fine tuning required in the Lagrangian, and special initial conditions to start it off; inflating initial conditions are most commonly just imposed by hand, or else via a handwaving appeal to 'primordial chaos'. Despite this arbitrariness, the simplest versions of inflation are reasonably predictive. They predict a critical (matter) density universe with roughly scale-invariant fluctuations. However, over the last few years the observational evidence has grown against a critical-density universe. A non-zero cosmological constant provides one fix, but in

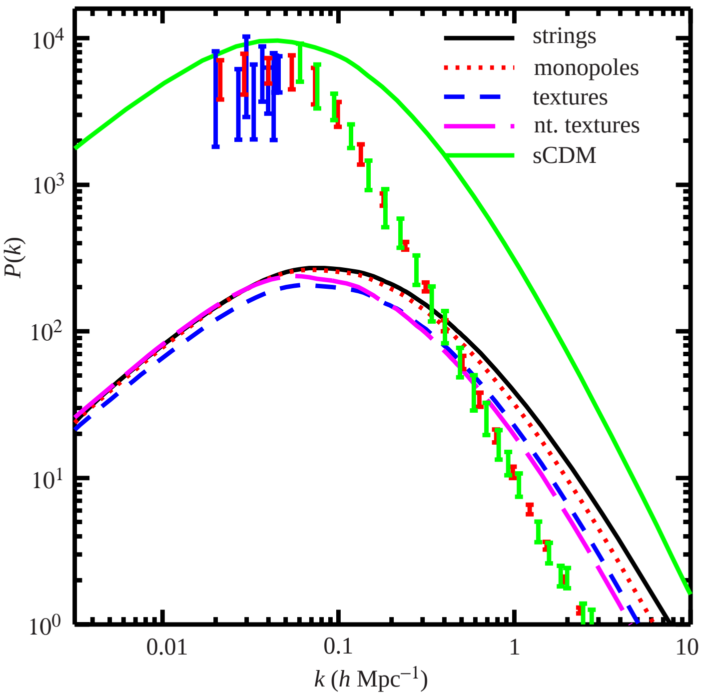

Figure 1. Predictions of the global defect theories for the matter power spectrum are compared with the observations (error bars). The upper curve shows the prediction for standard inflation plus cold dark matter, normalized to COBE.

the usual approach that requires extreme fine tuning. So inflationary theory seems to be forced down a more complex path. In this paper I will describe recent work with Hawking which makes open inflationary universes, or universes with a non-zero cosmological constant, more palatable.

However, before I begin, let me emphasize that these more complex routes depend more strongly on the inflationary model—they make us face up to the questions of the initial conditions and exactly what kind of fields are involved, questions which previously could be 'swept under the rug'. Solving them will require, more than ever before, a convincing link between inflation and fundamental theory, and a convincing theory of the initial conditions.

The main alternative theory of structure formation involves symmetry breaking and phase ordering, with associated cosmic defects such as strings and textures (Vilenkin & Shellard 1994). These theories are better motivated than inflation from high-energy theory, and more predictive. Cosmic defects were an almost inevitable consequence of unification and symmetry breaking as the universe cooled. Many of the simplest grand unified theories predict cosmic strings (Kibble *et al*. 1982; Olive & Turok 1982), and simple family symmetry theories predict cosmic texture (Joyce & Turok 1994). The amplitude of the primordial density perturbations $\delta\rho/\rho$ emerges naturally in the cosmic defect theories as the ratio $(M_{\rm GUT}/M_{\rm Pl})^2$, where $M_{\rm GUT}$ is the scale at which the observed gauge couplings unify, of order $10^{-3}M_{\rm Pl}$. The defect theories are less ambitious than inflation in that they do not address the origin of

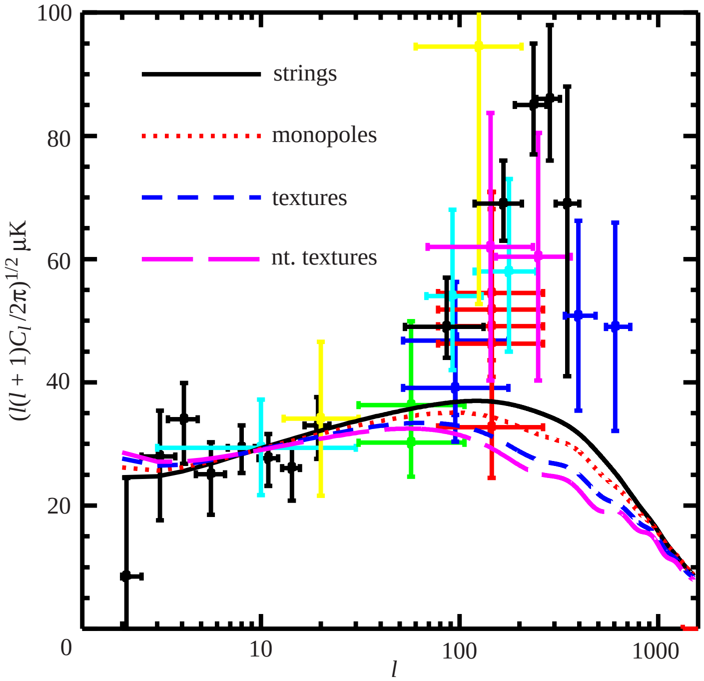

Figure 2. Predictions of the global defect theories (curves) for the cosmic microwave anisotropy are compared to the observations (error bars).

the background spacetime. One just assumes that some mechanism, such as inflation, produced a smooth flat universe in the first place.

For many years, the nonlinearity of the cosmic defect theories, and the need to evolve the defects over many expansion times, resisted the accurate determination of their predictions. However, recent breakthroughs in computational techniques (Allen *et al*. 1997; Pen *et al*. 1997) have overcome this problem (at least for the global defect theories) and made accurate predictions possible at last. The outcome is that the simplest theories fail badly with respect to the observations. There are two problems, illustrated in figures 1 and 2, respectively. The first is that if the theories are normalized to the Cosmic Microwave Background Explorer satellite (COBE), they fail to produce adequate density perturbations on galaxy clustering scales (10–100 Mpc). The conventionally defined measure of perturbations on the scale of galaxy clustering, σ_8, is too low, and the problem gets worse at larger scales. The second is that the theories show no significant Doppler peak at $l \sim 200$–300 as seems to be indicated in the cosmic microwave anisotropy data. In the global defect theories, Pen *et al*. (1997) found that these problems are not alleviated by the usual fixes for inflationary models—such as introducing a cosmological constant, making the universe open, or adding a hot dark-matter component. Avelino *et al*. (1997) and Battye *et al*. (1998) claim on the basis of more approximate calculations that the situation *is* improved for local cosmic strings in Λ or open universes. They correctly point out that lowering Ω_{matter} causes the power spectrum to shift to longer wavelengths by a factor $\Omega_{\text{matter}}^{-1}$, and this improves it's shape relative to the observations. However, there is still a problem with the amplitude of the RMS mass fluctuations, $\delta M/M$. Shifting the matter power spectrum increases $\delta M/M$ at most by $\Omega_{\mathrm{matter}}^{-2}$ (making the most optimistic assumption, that the power spectrum $P(k) \propto k$). This is offset by the loss in growth proportional to *ca.* $\Omega_{\mathrm{matter}}^{1.23}$ for a flat (Λ) universe, *ca.* $\Omega_{\mathrm{matter}}^{1.65}$ for an open one. One of the observations we need to fit is the RMS mass fluctuation on $8h^{-1}$ Mpc, estimated from the abundance of galaxy clusters to be $\sigma_8 \sim 0.57\Omega_{\mathrm{matter}}^{-0.56}$. The number comes from simulations of Gaussian models, but I would expect a similar result for cosmic strings. Comparing these two dependences, one sees that the gain in matching σ_8 is at most $\Omega_{\mathrm{matter}}^{-0.21}$ in the flat universe. E. P. Shellard (personal communication) and others argue that the change in the string density from the radiation to matter transition could be of assistance, but even including this change, Battye *et al*. (1998) obtain $\sigma_8 \sim 0.4$ in a flat universe with $\Omega_{\mathrm{matter}} = 0.3$, well below $\sigma_8 \sim 1.1$ which the cluster abundances would require.

My conclusion is that if current interpretations of the data are correct, the simplest defect theories seem in bad shape, and some fairly drastic change will be needed to restore them to good health. One possibility is a non-minimal Ricci coupling for the Goldstone bosons, as has been investigated by Pen (1998).

While the failure of the simplest defect theories is disappointing, it is important to emphasize that in many unified theories cosmic defects form at intermediate scales—anywhere from the electroweak to the Planck scale. Regardless of whether they formed cosmic structure or not, the detection of a cosmic defect would have tremendous significance for fundamental theory. It is important that we continue to search our horizon volume for them, via their signatures in the microwave sky, as gravitational lenses or as sources for very-high-energy cosmic rays.

2. Open inflation

I am now going to turn to a discussion of inflation, and our recent work (Hawking & Turok 1998*a*) linking the no-boundary proposal (Hartle & Hawking 1983) to open inflation (Gott 1982; Bucher *et al*. 1995; Yamamoto *et al*. 1995).

Inflation seeks to explain the following fundamental puzzles regarding the initial conditions of the hot Big Bang.

(i) Where did the hot matter come from?

(ii) Why was it expanding? (What went bang?!)

(iii) Why were the initial conditions so uniform and flat?

Inflation is based on the observation that scalar ('inflaton') fields have a potential $V(\phi)$ which can provide a temporary cosmological constant, driving an exponential expansion of the universe. This makes the universe very smooth and flat. As the inflaton ϕ rolls down to its minimum, the potential energy is converted into matter and radiation. The geometry is nearly flat: this requires that the universe be expanding at a rate given by the density.

The inflationary solution is beguilingly simple, but it raises many new puzzles which are still unresolved.

(1) *What is the inflaton ϕ and what is $V(\phi)$?* In contradistinction to cosmic defects, inflation does not appear to mesh nicely with unification. To obtain the correct level of density perturbations in today's universe one requires very flat potentials, characterized by a dimensionless coupling $\lambda \sim 10^{-14}$. In the simplest grand unified

theories there is no reason for scalar self-couplings to be so small—on the contrary this requires ugly fine tuning. Supersymmetric theories offer the prospect of new fields, 'moduli', with very flat potentials, but additional contrivance is needed to prevent the moduli excitations from later dominating the universe.

(2) *Why was the inflaton initially displaced from its potential energy minimum*? Initially it was hoped that inflation would result from a supercooled phase transition with the inflation field as the order parameter. But the inflaton field is so weakly coupled that it is hard to understand how thermal couplings to other fields could localize it away from its potential minimum.

(3) *What came before inflation*? Whether inflation happened at all, and how effective it was in flattening and smoothing the universe, depends on the initial conditions prior to inflation. It is often argued that the measure on the space of initial conditions does not really matter, because inflation is such a powerful (exponential) effect that *almost any* initial conditions will end up being dominated by regions of the universe which are inflating. One of the interesting things about the specific measures proposed (Hartle–Hawking or tunnelling wavefunctions) is that they involve *much greater* exponentials, which typically dwarf the exponential increase of proper volume during inflation. At the very least these measures demonstrate that the initial conditions *do* matter, and the issue cannot be ignored.

(4) *How do we avoid the initial singularity*? Inflation offers no solution to the singularity problem, certainly not if it was preceded by a hot early phase, or by primordial chaos. Open inflation and the no-boundary proposal do offer a solution, as I will describe below.

To address puzzles 3 and 4, one needs a theory of the initial conditions of the universe. Usually we regard the universe as a dynamical system, in which we need to specify the initial conditions as well as the evolutionary laws. This is true quantum mechanically just as it is classically. But it is unsatisfactory that there should be a separate input required, on top of the laws of nature, to specify the beginning of the universe. We could avoid the problem if the laws of physics defined their own initial conditions. Something like what we want occurs in thermal equilibrium—the state of the system is completely specified by the Hamiltonian, and one just 'integrates over' all possible states. We would like an analogous prescription for cosmology.

Imagine constructing a set of initial conditions for cosmology. Consider a universe which is topologically a three sphere but with an arbitrary three metric, scalar and other matter fields, and their momenta. What distribution should we assume for all these fields? The simplest choice would be a Boltzmann distribution at a specified temperature T. But then we would have T as an unwanted arbitrary parameter. Worse, due to the Jeans instability, in the presence of gravity the thermal ensemble does not exist.

The Hartle–Hawking proposal is nevertheless close in spirit to the thermal ensemble. One defines the quantum amplitude for a given three metric and scalar field configuration

$$\Psi(g^3, \phi) \propto \int^{g^3,\phi} \mathcal{D}\phi \mathcal{D}g^3 \mathrm{e}^{\mathrm{i}S}, \qquad (2.1)$$

where the lower limit of the integral is defined by continuing to Euclidean time and compact Euclidean metrics. All such metrics and the associated fields are to be 'integrated over', so that there is no additional information needed to specify the

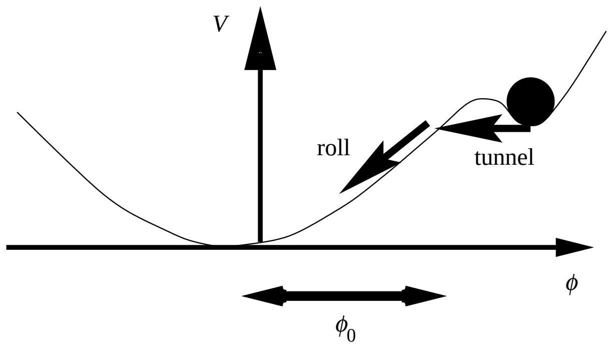

Figure 3. Scalar field potential for 'old' open inflation.

integral. The prescription is a natural generalization of the imaginary time (Matsubara) formalism in statistical physics—one can show that correlators calculated from (2.1) are given by a Euclidean path integral with no boundary, just as thermal correlators are. In thermal physics the size of the Euclidean region determines the temperature T: in the Hartle–Hawking prescription the size of the Euclidean region is dynamically determined by gravity, and this fixes the effective temperature to be the Hawking temperature of the associated background. This is nicely self-referential (for elaboration of the thermal analogy, see Turok (1998)).

The only known way of implementing the Hartle–Hawking proposal in practice is to find suitable saddle points of the path integral (i.e. classical instantons and their Lorentzian continuations) and expand around these to determine the Gaussian measure for the fluctuations. So one writes

$$\int^{g^3,\phi} \mathcal{D}\phi \mathcal{D}g^3 \mathrm{e}^{\mathrm{i}S} \approx \mathrm{e}^{-S_{\mathrm{E}}(\mathrm{inst})} \int^{g^3,\phi} \mathcal{D}\delta\phi \mathcal{D}\delta g^3 \mathrm{e}^{\mathrm{i}S_2}, \tag{2.2}$$

where S_2 is the quadratic action for small fluctuations and $S_{\mathrm{E}}(\mathrm{inst})$ is the Euclidean action of the instanton. In this approximation, we can compute all correlators of fields and fluctuations, and to second order we have complete quantum information about the universe. The quadratic fluctuation correlators are most easily calculated in the Euclidean region as well, and then analytically continued to the Lorentzian region, as in thermal physics. This is a very elegant formalism—literally everything is computed inside the microscopic Euclidean instanton (the 'pea'), and then analytically continued into the real Lorentzian universe.

The new development was the realization that a broader class of instantons than had previously been thought could contribute to the path integral. These new instantons naturally give rise to open inflation, previously thought to be a somewhat artificial special case.

The 'old' way of obtaining open inflation is illustrated in figures 3 and 4. One assumes that for some reason (primeval chaos?) a scalar field ϕ was trapped in a 'false minimum' of its potential, driving a period of exponential expansion which solves the smoothness and flatness puzzles. This period of 'old' inflation (Guth 1981) terminates with the nucleation of a bubble. Bubble nucleation is described in the semiclassical approximation using a Euclidean instanton representing the tunnelling of the scalar field through the barrier between the 'false vacuum' and the 'slow

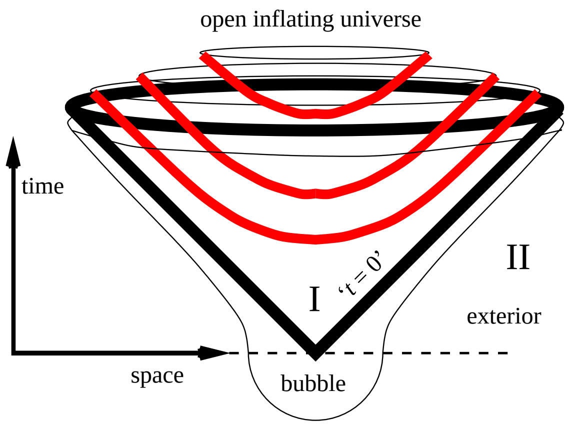

Figure 4. Spacetime picture showing an open inflationary bubble.

roll' part of the scalar field potential. The calculation was first described in the gravitational context by Coleman & De Luccia (1980).

Bucher *et al*. (1995) showed that the value of Ω today inside such a bubble is determined by the distance ϕ_0 that the inflaton field rolls during the period of inflation inside the bubble:

$$\Omega_0 = (1 + \mathcal{A}e^{-2\mathcal{N}})^{-1}, \tag{2.3}$$

where $\mathcal{A}$ represents the factor by which Ω deviates from unity during the post-inflationary epoch. This depends on the temperature to which the universe is heated immediately following the inflationary era, i.e. at the start of the 'standard hot Big Bang'. If the universe is heated to the electroweak temperature, $\mathcal{A} \sim 10^{30}$, whereas if it is heated to the Grand Unified Theory temperature, $\mathcal{A} \sim 10^{60}$. The quantity $\mathcal{N}$ is the number of inflationary e-foldings: approximating the potential as quadratic during the slow roll phase one has $\mathcal{N} \approx \frac{1}{4}(\phi_0/M_{\mathrm{Pl}})^2$, where $M_{\mathrm{Pl}} = (8\pi G)^{-1/2}$ is the reduced Planck mass.

In order to obtain an interesting value of $0.1 \lesssim \Omega_0 \lesssim 0.9$ today, one requires

$$\tfrac{1}{2}\log\mathcal{A} - 1 \lesssim N \lesssim \tfrac{1}{2}\log\mathcal{A} + 1, \tag{2.4}$$

which requires tuning of ϕ_0/M_{Pl} to a few per cent. Note that the tuning is only logarithmic in the large number $\mathcal{A}$, so it is not extreme.

The 'old' approach to open inflation served as an existence proof that inflation could produce an observationally interesting open universe. In addition, two remarkable properties should be noted. First, an infinite open universe emerges inside a bubble which at any fixed time appears finite. This is allowed by general relativity because the volume of the universe is that of the constant density hypersurfaces, and these need not correspond to the time-slicing as defined by an observer external to the bubble. Second, the singularity of the hot Big Bang (the light cone labelled $t = 0$ in figure 4) is just a coordinate singularity! This is possible only for an open universe which is potential dominated at early times. It is very interesting that the problem of the initial singularity does in this sense also point to a phase of potential domination prior to the standard hot Big Bang.

But 'old' open inflation has some unattractive features. First, the logarithmic fine-tuning mentioned above. Worse, the Coleman–De Luccia instanton only exists for

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

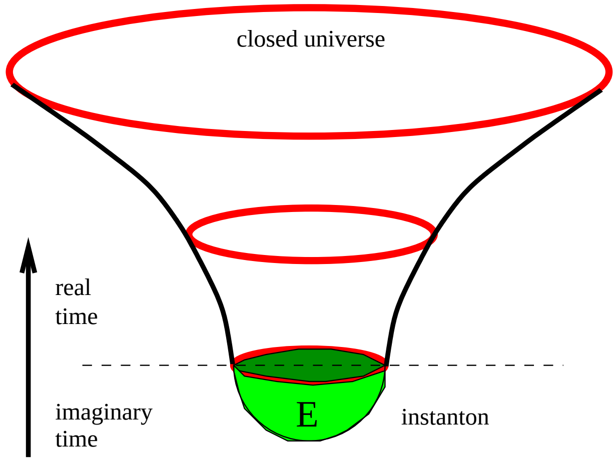

Figure 5. Obtaining de Sitter space from the Hawking–Moss instanton.

large values of the curvature (second derivative) of the scalar field potential around the tunnelling barrier. The reason for this is that the width of the bubble wall is set by the curvature scale (effective mass squared m^2) of the potential. The width m^{-1} must be smaller than the size of the de Sitter space throat H^{-1}, where H is the Hubble constant in the false vacuum, or else the bubble just does not fit inside the de Sitter space. This condition is only met if the potential has a 'sharp' false minimum, a very contrived situation. There have been attempts to find more 'generic' open inflationary models, by introducing more fields (Linde & Mehzlumian 1995) but it is hard to quantify the extent to which fine tuning has been replaced by contrivance. The physics of these models is quite complicated (see Garcia-Bellido *et al.* (1998) and Vilenkin (1998*a*) for reviews) and there is a large arbitrariness in the initial conditions.

I am going to describe a different approach which allows for open inflation in essentially any potential flat enough to allow significant inflation. Until recently it was thought that the only instantons which when analytically continued gave real Lorentzian universes were the Coleman–De Luccia instanton and a second, simpler instanton called the Hawking–Moss instanton (Hawking & Moss 1983). The latter occurs when the scalar potential has a positive extremum. In that case there is a solution in which the scalar field is constant, and the metric is that for de Sitter space with the appropriate cosmological constant. This solution is illustrated in figure 5.

This produces an inflating universe, but in the first approximation it is perfect de Sitter space, with too much symmetry—$O(4,1)$—to describe our universe. The de Sitter space is highly unstable, and ϕ would rapidly fall off the maximum of the potential. In fact, after ϕ rolled down and reheated the universe, the final state would be very inhomogeneous. If there were sufficient inflation, the inhomogeneities would be on unobservably large scales, but the universe would also be very flat and there would be no observational signature of its beginnings.

3. New instantons

Hawking and I considered instead a 'generic' potential without extrema or false vacua (see figure 6). In this case, there is no $O(5)$ invariant solution to the Euclidean field

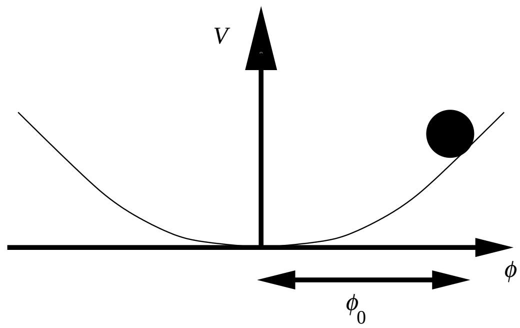

Figure 6. Potential for 'generic' open inflation.

equations simply because ϕ cannot remain constant if its potential $V(\phi)$ has a slope. This is our first observation—the maximal symmetry for an instanton in a 'generic' potential is $O(4)$. When continued to the Lorentzian region, this becomes $O(3,1)$, precisely what one needs for a homogeneous isotropic, but time-dependent, classical open universe.

There exists not just one such solution but a one-parameter family of solutions, labelled by the starting value of the scalar field ϕ_0. These solutions are singular but have finite action and can therefore contribute to the Euclidean path integral.

In the Euclidean region, the equations governing the instanton solution are

$$\phi'' + 3(b'/b)\phi' = V_{,\phi}(\phi), \qquad b'' = -\tfrac{8}{3}\pi G b(\phi'^2 + V(\phi)), \tag{3.1}$$

where the Euclidean metric is $\mathrm{d}s^2 = \mathrm{d}\sigma^2 + b(\sigma)^2\,\mathrm{d}\Omega_3^2$, with $\mathrm{d}\Omega_3^2$ the three-sphere metric. The field ϕ evolves in the 'upside down' potential $-V(\phi)$. We start from the assumption that the point $\sigma = 0$ is a regular (analytic) point of the solution—this will correspond to the vertex of the light cone containing the open universe. It follows that $\phi \sim \phi_0 + c\sigma^2 + \cdots$ and $b \sim \sigma + c'\sigma^3 + \cdots$, with c and c' constants. Assuming the slope of the potential is small, ϕ' is small and the solution is close to a four sphere for most of the range of σ, with $b \sim H^{-1}\sin(H\sigma)$. At larger σ, however, the damping term $b'/b \sim \cot(H\sigma)$ changes sign so that the motion of ϕ is antidamped. Nothing can now stop ϕ from running away as b vanishes. For the flat potentials of interest, one sees that $\phi' \sim b^{-3}$ and so $b(\sigma)$ vanishes as $(\sigma_{\max} - \sigma)^{1/3}$, and ϕ diverges logarithmically. There is a curvature singularity at $\sigma = \sigma_{\max}$.

These instantons are straightforwardly continued into the Lorentzian region, to obtain the spacetime sketched in figure 7. The infinite open universe occurs on the right of figure 7—seen from that side it would appear as in figure 4—bounded by the light cone delineated by the dashed lines. On the left is the singular boundary of the Lorentzian spacetime. We have not avoided the singularity problem of the standard Big Bang altogether, but have instead 'sidestepped it', because all the correlators of interest in the open universe are computed in the Euclidean region and these can be analytically continued into the Lorentzian region. The singularity occurs as a boundary of zero proper size to the Euclidean region. A detailed investigation (Gratton & Turok 1999) shows that the field and metric fluctuations are perfectly well defined in the presence of the singularity, and the singular boundary behaves as a perfect reflector.

Are such singular instantons allowed? The first question is whether they can contribute to the Euclidean path integral, i.e. whether they have finite Euclidean action.

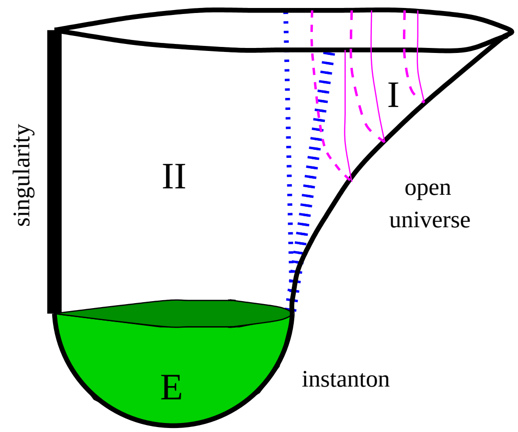

Figure 7. Complete spacetime for open instantons.

Surprisingly, they do. This is so because the Einstein action is not positive semidefinite, and the divergence in gradient energy of the scalar field is cancelled by an equal and opposite divergence in the Ricci scalar. This is essentially the same effect that allows inflation to produce arbitrary amounts of energy—the energy in matter is compensated by the negative energy of the gravitational field.

The Euclidean action turns out to be well approximated by (Turok & Hawking 1998)

$$S_{\mathrm{E}} \approx -\frac{12\pi^2 M_{\mathrm{Pl}}^4}{V(\phi_0)} - \frac{\sqrt{\frac{3}{2}} M_{\mathrm{Pl}} V_{,\phi}(\phi_0))}{V^2(\phi_0)}. \tag{3.2}$$

The second term is a contribution from the singular boundary of the instanton (Gibbons & Hawking 1977; Vilenkin 1999). It is negligible in monotonic flat potentials, but has some interesting effects near potential maxima (Bousso & Linde 1998; Turok & Hawking 1998).

There is a simple argument for the negative sign of the instanton action. The Euclidean action for an instanton of size a is

$$S_{\mathrm{E}} \sim \mathrm{Vol}(-R + \Lambda) \sim -a^2 M_{\mathrm{Pl}}^2 + a^4 \Lambda, \tag{3.3}$$

where Vol is the volume of the instanton, $R \sim a^{-2}$ the Ricci scalar and Λ the effective cosmological constant. The potential for a is of the familiar Higgs form—the negative sign of the Einstein action favours $a^2 \sim M_{\mathrm{Pl}}^2/\Lambda$, from which $S_{\mathrm{E}} \sim -M_{\mathrm{Pl}}^4/\Lambda$.

From the path integral formula (2.2), it follows that the probability of having such an instanton is proportional to $\mathrm{e}^{-2S_{\mathrm{E}}(\mathrm{inst})} \sim \mathrm{e}^{+M_{\mathrm{Pl}}^4/\Lambda}$. This means that the most likely beginning to the universe is for the effective Λ to be very small, which means there is no inflation. The initial size a of the universe is favoured to be very large. Linde (1998) argues this to be implausible—surely it is easier to create a small universe than a big one 'from nothing'. Personally, I don't view this as 'creation from nothing' (which I am not sure makes any sense!), but rather I see the Hartle–Hawking prescription as a measure for the set of initial conditions. One can show that the entropy associated with de Sitter space scales just as a negative of the Euclidean action, so there are just more states available for large initial universes.

As mentioned above, the exponent is tremendously large: for a theory with $V = \frac{1}{2}m^2\phi^2$, the probability

$$P \propto \mathrm{e}^{-2S_{\mathrm{E}}(\mathrm{inst})} \sim \mathrm{e}^{10^{12}/\mathcal{N}}, \tag{3.4}$$

where $\mathcal{N}$ is the number of e-foldings. This exponent drastically outweighs that due to the increase in proper volume in inflation, which is a puny $\mathrm{e}^{3\mathcal{N}}$.

Is this a disaster for inflation? There are three possibilities: the scalar field theory is wrong; the Hartle–Hawking ansatz is wrong; or we are asking the wrong question. Let us begin with the last possibility. Perhaps it is too much to ask that the theory only produces universes like ours. If it allows other, uninhabitable, universes we should exclude them from consideration, just as we are not surprised to find ourselves living on the surface of a planet, rather than in empty space.

We live in a galaxy, i.e. a nonlinear collapsed region, which was probably essential to the formation of heavier elements and life. Let us therefore condition the prior probability given by (3.4) by the requirement that there is a galaxy at our spatial location. Such a conditioning is of course a version of the anthropic principle, but a very minimal version. I prefer to see it as just using up some of the data we know about the universe (e.g. that we live in a galaxy) in order to explain other data. We do so in the hope that a more complete theory of the origin of life will explain why only universes containing galaxies should matter.

Hawking and I imposed the 'anthropic' requirement via the Bayes theorem

$$\mathcal{P}(\Omega_0|\mathrm{gal}) \propto \mathcal{P}(\mathrm{gal}|\Omega_0)\mathcal{P}(\Omega_0) \propto \exp\left(-\frac{\delta_{\mathrm{c}}^2}{2\sigma_{\mathrm{gal}}^2} - 2S_{\mathrm{E}}(\phi_0)\right), \tag{3.5}$$

where the first factor is the probability that a galaxy-sized region about us underwent gravitational collapse. This is Gaussian because we are assuming the usual inflationary fluctuations. The RMS perturbation σ_{gal} is the RMS amplitude for fluctuations on the galaxy scale, equal to the perturbation amplitude at Hubble radius crossing for the galaxy scale $\Delta(\phi_{\mathrm{gal}})$ multiplied by the growth factor $G(\Omega_0)$. The latter is strongly Ω_0 dependent. The quantity δ_{c} is the linear theory amplitude of a perturbation when it undergoes collapse.

The idea is very simple. The Euclidean action favours a very small number of e-foldings, which from the formula (2.3) means very small Ω_0. However, galaxy formation requires the growth of density perturbations, and this does not occur unless Ω_0 is substantial. For Gaussian perturbations (as inflationary models predict), this effect is also exponential and can compete with the Euclidean action. If one maximizes the posterior probability (3.5), the most likely value of Ω_0 is of order 0.01 for generic slow roll inflationary potentials. Such a low value is unacceptable in comparison with the data, but it is not such a bad miss. As far as I know, this is the first attempt to calculate Ω_0 from first principles, and I find it encouraging that it is not so far wrong.

In my view the most questionable component of the theory is the scalar field ϕ, with its inflaton potential. As I discussed above, the identity of ϕ is still wide open, and it is conceivable that ϕ cannot be modelled as a single simple scalar field. There may well be additional fields coupling to ϕ which affect the prior probability.

The second possibility (Linde 1988; Vilenkin 1988*b*) is that the Hartle–Hawking prescription is incorrect. Linde (1988) and Vilenkin (1988*b*) have suggested alternate prescriptions which have the effect of reversing the sign of the action for the Euclidean instanton, favouring the universe starting at a high density. I do not have space to discuss these here (for a critique see Hawking & Turok (1998*b*)) but they appear to be significantly less well defined. In particular it is not clear how they are consistent with general coordinate invariance. The appealing thing to me about the Hartle–Hawking prescription is that it is the natural generalization of the thermal ensemble to higher dimensions, based on well-established principles of relativity and quantum mechanics. I think we should be reluctant to abandon it without good reason.

The third possibility is that we must impose more stringent anthropic requirements, e.g. that the density of the galaxy we live in should not be too great or else planetary systems would be disrupted (see, for example, Tegmark & Rees 1997). This effect would also act to increase the predicted Ω_0.

Finally, one can question the whole enterprise of using singular instantons in cosmology. Vilenkin (1998*a*) has raised the objection that there are yet more general instantons than ours which are asymptotically flat and would at face value lead to an instability of Minkowski spacetime. On these grounds he suggests that *all* singular instantons should be excluded in spite of the finiteness of the action. To see Vilenkin's instanton, consider solving the field equations (3.1) starting near the singular point, with $b \sim (\sigma_{\max} - \sigma)^{1/3}$ and $\phi \sim -\log(\sigma_{\max} - \sigma)$. For simplicity consider a theory with no potential V; in general, there are similar solutions if one arranges for the field to asymptotically tend towards the value for which V vanishes. If V is negligible, we have $\phi' \propto b^{-3}$, and the equation for b has a solution $b \sim \sigma$ at large b. The absence of potential energy means that instead of closing in on itself, as in the open inflationary instantons, the solution opens out into an asymptotically flat Euclidean space. The instanton can be cut on a three surface running horizontally and intersecting the singularity at X (see figure 8). The continuation to Lorentzian spacetime describes an asymptotically flat space with an expanding hole in it. One can show that the action for such Euclidean instantons can be made arbitrarily small, so flat spacetime would be terribly unstable.

While Vilenkin's instanton is similar to ours locally (near the singularity), the global structure is very different. In his case, the singularity eventually eats up all of the Lorentzian region, and so affects all observers in that region. In our case, the singularity never enters the open universe region. The singularity is not causally disconnected from us, but all the perturbations in the universe are defined by continuation from the Euclidean region. If they are well defined there, that is all that matters for predicting what happens in the open universe.

It seems plausible that we should exclude Vilenkin's instantons on the grounds that they give the transition amplitude from flat Minkowski spacetime to an 'unreasonable' spacetime, namely flat space with a growing hole in it. The argument is not so clear in our case. First, we are not computing a transition amplitude but rather defining an 'initial condition' for the wavefunction. Second, the part of our final state in which we are interested is a 'reasonable' spacetime, namely an infinite open universe. Unfortunately we do not have a general prescription for what is or is not a 'reasonable' final state spacetime, so I do not think we can fully settle the argument at this stage.

The singularity is an indication that the theory is incomplete, but this is in any case no surprise—in any theory of quantum gravity the Einstein action will acquire significant corrections at short distances. But our ability to predict is not necessarily compromised. I would draw the analogy with the hydrogen atom, where the quantum

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

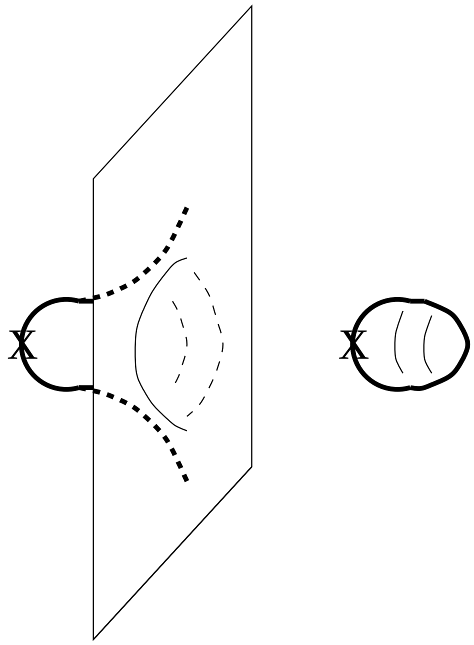

Figure 8. Vilenkin's asymptotically flat instanton (left) is compared with the open inflationary instanton (right). Both have a singularity at X, and both are cut horizontally in two by the surface on which one matches to the Lorentzian region.

mechanics is perfectly sensible, even though the Coulomb singularity points to the need for a more complete theory. In our case, calculation has shown that the perturbations are unambiguously defined in the Euclidean region (at least at one loop), so there is reason to hope that whatever underlying physics resolves the singularity will not affect our predictions.

There is an intriguing suggestion due to Garriga (1998) that the singularity may be removed in a higher-dimensional theory. Garriga shows that a solution to the five-dimensional Einstein equations with a cosmological constant (actually a five sphere) may be written in a dimensionally reduced form in such a way that the effective four-dimensional metric has precisely the same sort of singularity that our open inflationary instantons do. The singularity is then seen as an artefact of trying to look at a five-dimensional metric from a four-dimensional point of view. Garriga's model is not realistic because it has too much symmetry—the only way to continue a five sphere to a Lorentzian spacetime is to continue at an equator, obtaining five-dimensional de Sitter space. But it is likely that the idea will generalize to less symmetric instantons, which may allow a realistic expanding four-dimensional universe with a frozen extra dimension.

4. The four form and Λ

The recent observations of supernovae at high redshift, when combined with observations of the cosmic microwave anisotropy, give evidence in favour of a non-zero cosmological constant in today's universe. It is hard to understand why the cosmological constant is small: theoretical prejudice has for a long time been that there

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

must be some as yet undiscovered symmetry or dynamical mechanism that sets it zero. It is even harder to understand why the cosmological constant should have a value such that it is just beginning to dominate the density of the universe today. One possibility is an anthropic argument, and this has been pursued by a number of authors (Efstathiou 1995; Vilenkin 1995; Martel *et al*. 1997). The anthropic argument is particularly powerful here because a cosmological constant has such a dramatic effect on the expansion of the universe. Hawking and I have recently discussed how the four-form field strength in supergravity fits in with these ideas, allowing for an anthropic solution of the cosmological-constant problem in which the cosmological constant provides a non-negligible contribution to the density of today's universe.

The four-form field is a natural addition to the field content of the world, and is demanded by the simplest candidate field theory of quantum gravity, eleven-dimensional supergravity. Upon dimensional reduction to four dimensions, the four-form field strength possesses a remarkable property; namely the general solution of the field equations is parametrized by just a single spacetime-independent constant p. This constant contributes to the effective cosmological constant in the four-dimensional Einstein equations, with the contribution to Λ being proportional to p^2. If p is appropriately chosen, this contribution can cancel other contributions coming from electroweak symmetry breaking, confinement, chiral symmetry breaking and so on.

So now we have a variable cosmological constant, we have to ask what the prior probability distribution is for it. The Hartle–Hawking proposal provides us with such a distribution. There is a subtlety in calculating the appropriate Euclidean action, which led to a disagreement about the sign of the action (Hawking 1984; Duff 1989), which I think is now resolved (Turok & Hawking 1998). The point is that the arbitrary constant p mentioned above is actually the canonical momentum of a free quantum-mechanical particle. The eigenvalue of this momentum determines the value of the cosmological constant today. When one calculates the wavefunction from Feynman's path integral, one is interested in the amplitude for a particular classical universe. The effective cosmological constant in that classical universe is specified by the four-form momentum. So one needs to calculate the path integral for the four form in the momentum representation.

This gives a probability distribution analogous to that above, where the most likely universe is one with an initial net cosmological constant (i.e. including $V(\phi)$) equal to zero. Such a universe would never give inflation, but as before we can calculate the posterior probability for Λ and Ω_0 given that we live in a galaxy. This requirement is even more stringent upon Λ than it is on Ω_0—if Λ was not very tiny in Planck units, the universe would have recollapsed, or begun an epoch of inflation, even before the galaxy mass scale had crossed the Hubble radius. So, to even discuss the existence of galaxies, Λ has to be in an extremely narrow band about zero, and across this very narrow band, as anticipated by Efstathiou (1995), the prior probability distribution for Λ turns out to be very flat. Our calculations therefore support the assumptions made in their anthropic estimates of Λ. Unfortunately the inferred probability distribution for Λ is rather broad and, with only one possible measurement, the theory seems hard to test. Nevertheless if the observations of a non-zero Λ hold up, the above mechanism provides one of the few conceivable ways of understanding it.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

5. Summary

The impartial reader of this paper might conclude that we have gone backwards rather than forwards! Open inflation, and the theory of Λ I described, seem to broaden rather than narrow our theoretical options, and the use of anthropic arguments (albeit minimally as I prefer) certainly reduces our ability to predict everything in cosmology from a fundamental theory. Both developments are a result of reconsidering and generalizing fundamental theory in the light of observations. This is surely just a sign of a maturing science, in which the interplay between theory and observation is ongoing.

We should look for progress in two directions. We would like to connect inflation to a more fundamental theory of the beginnings of the universe. In tandem, we should seek specific observational signatures of these theoretical mechanisms, beyond their allowing us to simply adjust cosmological parameters. Our hope is that there will be specific observational signatures of the primordial instanton in the power spectrum of density perturbations (Gratton & Turok 1999).

References

Albrecht, A. J. 1997 In *Critical dialogues in cosmology* (ed. N. Turok). Singapore: World Scientific.

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

Allen, B., Caldwell, R. R., Dodelson, S., Knox, L., Shellard, E. P. S. & Stebbins, A. 1997 *Phys. Rev. Lett.* **79**, 2624.

Avelino, P. P., Shellard, E. P. S., Wu, J. H. P. & Allen, B. 1997 Preprint, astro-ph/9712008.

Bardeen, J. M., Steinhardt, P. J. & Turner, M. S. 1983 *Phys. Rev.* D **28**, 679.

Battye, R. A., Robinson, J. & Albrecht, A. 1998 *Phys. Rev. Lett.* **80**, 4847.

Bousso, R. & Linde, A. D. 1998 Preprint, gr-qc/9803068.

Bucher, M., Goldhaber, A. & Turok, N. 1995 *Phys. Rev.* D **52**, 3314.

Coleman, S. & De Luccia, F. 1980 *Phys. Rev.* D **21**, 3305.

Duff, M. 1989 *Phys. Lett.* B **226**, 36.

Efstathiou, G. P. 1995 *Mon. Not. R. Astr. Soc.* **274**, L73.

Garcia-Bellido, J., Garriga, J. & Montes, X. 1998 *Phys. Rev.* D **57**, 4669.

Garriga, J. 1998 Preprint, gr-qc/9804106.

Gibbons, G. W. & Hawking, S. W. 1977 *Phys. Rev.* D **15**, 2752.

Gott, J. R. 1982 *Nature* **295**, 304.

Gratton, S. & Turok, N. 1999 (In preparation.)

Guth, A. H. 1981 *Phys. Rev.* D **23**, 347.

Guth, A. H. & Pi, S.-Y. 1982 *Phys. Rev. Lett.* **49**, 1110.

Hartle, J. B. & Hawking, S. W. 1983 *Phys. Rev.* D **28**, 2960.

Hawking, S. W. 1982 *Phys. Lett.* B **115**, 295.

Hawking, S. W. 1984 *Phys. Lett.* B **175**, 395.

Hawking, S. W. & Moss, I. G. 1983 *Nucl. Phys.* B **224**, 180.

Hawking, S. W. & Turok, N. 1998*a* *Phys. Lett.* B **425**, 25.

Hawking, S. W. & Turok, N. 1998*b* Preprint, gr-qc/9802062.

Joyce, M. & Turok, N. 1994 *Nucl. Phys.* B **416**, 389.

Kibble, T. W. B., Lazarides, G. & Shafi, Q. 1982 *Phys. Lett.* B **113**, 237.

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

Linde, A. D. 1998 Preprint, gr-qc/9802038.

Linde, A. D. & Mezhlumian, A. 1995 *Phys. Rev.* D **52**, 6789.
Lyth, D. H. & Riotto, A. 1998 Preprint hep-ph/9807278.
Martel, H., Shapiro, P. R. & Weinberg, S. 1997 Preprint, astro-ph/9701099.
Olive, D. I. & Turok, N. 1982 *Phys. Lett.* B **117**, 193.
Pen, U. L. 1998 preprint, astro-ph/9804983.
Pen, U. L., Seljak, U. & Turok, N. 1997 *Phys. Rev. Lett.* **79**, 1615.
Starobinski, A. A. 1982 *Phys. Lett.* B **117**, 175.
Steinhardt, P. J. 1995 Preprint, astro-ph/9502024.
Tegmark, M. & Rees, M. 1997 Preprint, astro-ph/9709058.
Turok, N. 1998 Lecture notes on open inflation. http://www.damtp.cam.ac.uk/user/ngt1000.
Turok, N. & Hawking, S. W. 1998 *Phys. Lett.* B **432**, 271.
Vilenkin, A. 1995 Preprint, gr-qc/9512031.
Vilenkin, A. 1998*a* Preprint, gr-qc/9805252.
Vilenkin, A. 1998*b* Preprint, gr-qc/9804051.
Vilenkin, A. 1999 *Phys. Lett.* B. (In the press.)
Vilenkin, A. & Shellard, E. P. S. 1994 *Cosmic strings and other topological defects*. Cambridge University Press.
Yamamoto, K., Sasaki, M. & Tanaka, T. 1995 *Astrophys. J.* **455**, 412.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES