The Uncaused Beginning of the Universe

Author(s): Quentin Smith
Source: *Philosophy of Science*, Vol. 55, No. 1 (Mar., 1988), pp. 39-57
Published by: University of Chicago Press on behalf of the Philosophy of Science Association
Stable URL: http://www.jstor.org/stable/187819
Accessed: 24-02-2016 04:26 UTC

THE UNCAUSED BEGINNING OF THE UNIVERSE*

QUENTIN SMITH

Marion, Indiana

There is sufficient evidence at present to justify the belief that the universe began to exist without being caused to do so. This evidence includes the Hawking-Penrose singularity theorems that are based on Einstein's General Theory of Relativity, and the recently introduced Quantum Cosmological Models of the early universe. The singularity theorems lead to an explication of the beginning of the universe that involves the notion of a Big Bang singularity, and the Quantum Cosmological Models represent the beginning largely in terms of the notion of a vacuum fluctuation. Theories that represent the universe as infinitely old or as caused to begin are shown to be at odds with or at least unsupported by these and other current cosmological notions.

My purpose in this paper is to argue that there is sufficient evidence at present to warrant the conclusion that the universe probably began to exist over ten billion years ago, and that it began to exist without being caused to do so. I believe accordingly that the positions held by many if not most contemporary philosophers concerning this issue are unjustified, for their beliefs typically fall into one of three mutually exclusive categories, (1) the universe is probably infinitely old, (2) the universe began to exist and its beginning was caused by God, and (3) insufficient evidence is available to enable us to decide about whether the universe began to exist or is infinitely old.

1. The Prediction of a Space-time Singularity in our Past. Most philosophers today are aware that the Big Bang cosmological theory has superseded the Steady State theory, but a great number of these philosophers erroneously believe either that there is probably an infinite number of cycles of expansion and contraction of the universe, or that there is insufficient evidence to decide between the infinitely oscillating model and the theory that there was an earliest or single expansion, or that there is a first expansion that needs to be explained by introducing divine causality. That these beliefs are unfounded becomes apparent once the evidence for the prediction of a singularity in our past by the Big Bang cosmological model is adequately clarified.

The most important but by no means the only observational evidence

*Received September 1985.

Philosophy of Science, 55 (1988) pp. 39–57.

for the Big Bang theory is the redshift of the light from distant galactic clusters, first discovered by Slipher and Hubble, which indicates the universe to be expanding uniformly in all directions.[1] This suggests that there is some time in the past when all the galactic clusters, or all the materials in these clusters, were arbitrarily close together, and that this time represents the beginning of the universe.

This is more exactly understood in terms of the models of the universe provided by the so-called Friedmann solutions of the field equations of Einstein's General Theory of Relativity (GTR). The field equations show that the metric of space-time is dependent upon the matter present in that space-time.[2] The field equations can be solved for the universe as a whole if figures reflecting the observed values of the universe are introduced. Since the universe is isotropic (the same in all directions) and homogeneous (the matter is evenly distributed), it is described by the Robertson-Walker metric,[3] which applied to the field equations enable them to be reduced to (with the cosmological constant $\wedge$ omitted):

$$-3d^2a/dt^2 = 4\pi G\,(p + 3P/c^2)a$$

$$3(da)^2/dt = 8\pi Gpa^2 - 3kc^2$$

[1]See G. Stromberg's summary of V. M. Slipher's measurements in Stromberg 1925; also see Hubble 1929. Other observational evidence that supports Big Bang cosmology includes the background microwave radiation of 2.7 K, which is a remnant of the intense heat generated at an early stage of the expanding universe. This radiation was first discovered (and initially measured to be 3.5 K) by A. Penzias and R. Wilson in Penzias and Wilson 1965.

A third major set of data supporting the Big Bang cosmology is the abundance of helium 4, deuterium, helium 3, and lithium 7, the formation of which is predicted to occur in the first minutes of the Big Bang.

[2]According to John Wheeler (1973, p. 220), the simplest expression of the Einstein equations is

> (curvature of space time) = 8π (density of the mass-energy present in that space-time).

More completely, it can be said that the field equations relate the metric tensor $g_{\mu\nu}$ and its derivatives, which describe the geometry of space-time, to the energy-momentum tensor $T_{\mu\nu}$, which is determined by the distribution of the mass and energy in that space-time. These equations enable paths in space-time (specifically, geodesic paths) to be calculated. The formula summarizing the ten field equations is

$$R_{\mu\nu} - (1/2)Rg_{\mu\nu} + \lambda g_{\mu\nu} = -(8\pi G/c^2)T_{\mu\nu}$$

The terms on the left-hand side are composed of $g_{\mu\nu}$ and its derivatives, and also of the constant λ. G is the constant of gravitation and c the velocity of light.

[3]The Robertson-Walker metric is determined by a, the radius of the universe at a certain time, and by the curvature of space-time. The metric of a homogeneous and isotropic universe is

$$ds^2 = dt^2 - (1/c^2)a^2 d\sigma^2$$

where ds is the space-time interval between two events, $d\sigma$ is the line element of a space of constant curvature, and c the velocity of light.

a is the scale factor representing the radius of the universe at a given time. da/dt is the rate of change of a with time; it is the rate at which the universe expands or contracts. d^2a/dt^2 is the rate of change of da/dt, that is, the acceleration of the expansion or the deceleration of the contraction. G is the gravitational constant and c the speed of light. P is the pressure of matter and p its density. k is a constant which takes one of three values: 0 for a flat Euclidean space (in which case the universe is open, that is, expands forever), -1 for a hyperbolic space (in which case the universe is also open), or $+1$ for a spherical space (in which case the universe is closed, that is, will contract).

What is important to note about these Friedmann equations is that if p, the density of matter in the universe, is positive, then the right side of the first equation is positive, and this entails that d^2a/dt^2, the acceleration of the expansion or the deceleration of the contraction, cannot be zero. d^2a/dt^2 must be negative, which means that the acceleration of the expansion is decreasing or that the acceleration of the contraction is increasing. In a word, if there is matter present in the universe, then the universe must be either expanding or contracting with a varying acceleration.

It is the case of expansion that especially interests us, for the universe is now expanding. If the acceleration of the expansion is decreasing, this implies that the further we go into the past the greater the increase of the acceleration and the smaller the scale factor a of the radius of the universe, until a time t_0 is reached when $a = 0$. As d^2a/dt^2 increases and a decreases, the density of matter p increases, until at t_0 the value of p is infinite. At this time the entire universe is squeezed into at least one point of infinite density, infinite temperature, and infinite curvature. We have reached a space-time singularity.

I shall argue in the next sections that these considerations support the idea that there is an uncaused beginning of the universe. In the remainder of this section I shall discuss the issue of whether or not the singularity is real.

At first it was thought that the singularity predicted by the Friedmann equations was fictitious, since its prediction depended upon the assumption that the universe is exactly homogeneous and isotropic, whereas in reality it is only approximately so. Consider an inexactly symmetrical contracting universe: as the radius of the universe approaches zero the convergence of particles, due to small perturbations, would not focus upon a single point; rather the particles would rebound off one another and result in a "bounce" of the universe and a new phase of expansion. In the words of E. M. Lifshitz and I. M. Khalatnikov, who developed one of the more recent arguments for this scenario, the fluctuations "exclude the possibility of the existence of a singularity in the future of the contracting universe and imply that the contraction of the universe (if this must in general occur) must finally be turned into an expansion" (Lifschitz and Khalatnikov 1963, p. 207). This is the basis of the idea of an oscillating universe, according to which the universe runs through successive cycles of expansion and contraction. The present phase of the expansion, accordingly, can be understood as a result of a prior phase of contraction.

Before I explain how it can be proven that the above argument is mistaken and that a singularity must occur even if the universe is inexactly symmetrical, I shall show first that the assumption that the singularity is fictitious and that the universe oscillates does not render probable the idea that the universe is *infinitely* old.

Models of an oscillating universe usually predict that with each new cycle there is an increase in the size of the radius of the universe, amount of radiation present, and entropy.[4] Radiation from previous cycles accumulates in each new cycle, and the accompanying increase in pressure causes the new cycle to be longer than the last one; the universe expands to a greater radius and takes a longer time to complete the cycle. This disallows an infinite regress into the past, for a regress will eventually arrive at a cycle that is infinitely short and a radius that is infinitely small; this cycle, or the beginning of some cycle with values approaching the values of this cycle, will count as the beginning of the oscillating universe.

The inference to a finite past can also be made from a measure of the amount of radiation present in the universe; if there were an infinite number of previous cycles, an infinite amount of radiation would be present in the current cycle, but the amount measured is finite. Joseph Silk calculates that the amount of radiation observed in the present expansion allows there to be "about 100 previous expansion and collapse cycles of the universe" (Silk 1980, p. 311).

The conclusion that the past is finite also follows from facts about entropy; if an infinite number of previous cycles have elapsed, each with increasing entropy, then the present cycle would be in a state of maximum entropy—but in fact it is in a state of relatively low entropy.

John Wheeler sweeps away these objections to an infinitely oscillating universe by supposing that at the end of each contracting phase all the constants and laws of that cycle disappear and the universe is "reprocessed probabilistically" (Misner, Thorne and Wheeler 1973, p. 1214) so as to acquire new constants and laws in the next cycle. No information about a previous cycle is passed on to the next cycle. Accordingly, no inference to a finite past can be made on the basis of present observations

[4]The most widely discussed models have been developed in Tolman (1934, pp. 440 ff) and in Landsberg and Park (1975).

and the laws and constants that hold in the current cycle.

Now there is no reason to think that such a universe is logically impossible, but that is not germane to our present concern, which is to establish probabilist grounds for a belief in the finitude or infinitude of the universe's past. It is logically possible that at the point of onset of each new cycle all laws and constants are transformed, but since these occurrences cannot be predicted according to any known physical law, there is no reason to think that these transformations occur.

Indeed, there is a theoretical reason to prefer the finite oscillatory models to Wheeler's model (supposing that we must choose among oscillating models). The finite models, through being constructed in accordance with the known physical laws and constants, obey a principle related to the principle of induction; the related principle is that physical laws and constants originally inductively established for one domain of physical events should be applied to other domains of physical events if there is no observational evidence that events in these other domains differ in the relevant respects from those in the original domain. In the present context, the domains are cycles; since there is no observational evidence that events in past cycles differ relevantly from those in our cycle, we are not justified in supposing that the laws and constants inductively established in our cycle do not apply to the events in previous cycles.

The issue of whether oscillating universes are finite or infinite in respect of the past lost much of its urgency in the middle and late 1960s, with the development of the Hawking-Penrose singularity theorems (Penrose 1965; Hawking 1965, 1966, 1970), which entailed that an inexactly homogeneous and isotropic universe must have a singularity. A space-time contains a singularity if (1) the space-time satisfies the equations of GTR, (2) time-travel into one's own past is impossible and the principle of causality is not violated (there are no closed timelike curves), (3) the mass density and pressure of matter never becomes negative,[5] (4) the universe is closed and/or there is enough matter present to create a trapped surface, and (5) the space-time manifold is not too highly symmetric.[6]

It is reasonable to assume that all of these conditions, except perhaps (4), apply to our universe. Condition (4) might seem to be open to question if the universe is not closed and the condition of a trapped surface

[5]That is, the stress-energy tensor satisfies

$$(T_{\alpha\beta} - \frac{1}{2} g_{\alpha\beta} T) u^{\alpha} u^{\beta} \geq 0$$

[6]That is, the space-time is such that

$${}^{t}(a^{R}b)cd(e^{t}f)^{t^{c}t^{d}} \neq 0$$

holds at some point along each timelike or null geodesic. t^{a} is the tangent vector.

must obtain. A trapped surface is one from which light and matter cannot escape due to the intensity of the gravitional forces, such that under the influences of these forces the space-time paths of radiation and matter within the trapped surface converge towards a singularity. If the singularity is in the past, the geodesics of the rays and particles stem from the singularity; if it is in the future, the geodesics aim towards it. In the case of our universe, there is a singularity in its past (be the universe open or closed) if there is enough matter present to create a trapped surface. And there *is* enough matter:

> Recent observations of the microwave background indicate that the universe contains enough matter to cause a time-reversed closed trapped surface. This implies the existence of a singularity in the past.[7]

2. The Beginning of the Universe Defined. In order to show how the foregoing considerations render probably true the idea that the universe spontaneously began, a precise definition of the beginning of the universe must be developed. That is the task of the present section.

There are at least three possible definitions that might seem to be consistent with the ideas of the last section. Either the universe began (i) at the singularity, (ii) after the singularity, or (iii) neither at nor after the singularity.

(i) If the universe were closed and perfectly homogeneous and isotropic, then the definition of the beginning of the universe "at the singularity" would be relatively simple. At the first time, t_0, when $a = 0$, there exists a single point in which the entire universe is compressed, and the existence of this point counts as the beginning of the universe. This point exists for one instant before exploding in the Big Bang. However, since the universe is imperfectly homogeneous and isotropic, a more complicated definition is necessary. The imperfect symmetry implies that the universe began nonsimultaneously at a series of points.[8] Moreover, it is necessary that these points be infinite in number if the universe is open and space is infinite, for in one point there can be compressed only a finite volume of space. Given these factors, the appropriate definition of the beginning of the universe is that it began at the earliest singularity, this singularity consisting of the existence at t_0 of the first point(s) to

[7]Hawking and Ellis (1973, p. 3). The proof that the trapped surface created by this matter implies a singularity in our past (rather than in our future) is given on pages 356–359.

[8]The less dense parts of the universe exploded from points first, followed by the more dense parts. See Barrow and Silk (1983, p. 42).

It should also be noted that if the universe is not sufficiently isotropic and homogeneous, some past-directed timelike geodesics will not end in singularities. Observational evidence, however, suggests that the universe is sufficiently symmetric so that all do end in singularities. See Hawking and Ellis (1973, pp. 358–359).

explode in a Big Bang. The universe began at this time in the sense that this is the earliest time at which some part of the universe exists.

There is a serious problem with this definition. The universe is standardly defined as the set of events, each event being a point in a 4-dimensional space-time continuum, such that each event is characterized by four coordinates (x_1,x_2,x_3,t), the first three being spatial and the fourth temporal. But the singularity at t_0 is not in a 3-d space; it is in a space either of 0 dimensions (if it is just one point), 1 dimension (if it is a series of points constituting a line or line segment) or 2 dimensions (if it is a series of points comprising a surface-like space). Accordingly, the singularity at t_0 is not a part of the universe and *a fortiori* not the earliest part of the universe. Rather it is a *source* of the universe. The universe itself began at some time after t_0, which leads us to the second definition of the beginning of the universe.[9]

(ii) On this definition the beginning of the universe is the explosion of 4-dimensional space-time out of the earliest singularity, the singularity at t_0. In other words, the beginning of the universe is the Big Bang. The Big Bang is the first state of the universe.

The Big Bang occurs at $t > t_0$. However there is not some instant at which the Big Bang occurs, for (assuming that time is dense or continuous) there is no earlest instant after the first instant t_0; for every instant $t_a > t_0$ there is another instant $t_b < t_a$. Accordingly, if the phrase "the Big Bang" is to be used unequivocally, it must be used to designate a state occupying an interval that is the first interval of some length to elapse after t_0. Although on *a priori* grounds there is no nonarbitrary basis for selecting this length, there are empirical reasons for identifying the first post-t_0 interval of length 10^{-43} second as the time of the Big Bang. The earliest state of the universe that cosmologists have determined to be unpreceded by a state of a different kind is the state constitutive of the Planck era, which occupies the first post-t_0 interval of length 10^{-43} second. A cosmological state of some kind K is the only state at which all and only the types of particles and forces present in that state exist. It is speculated by many cosmologists that during the Planck era and only during the Planck era there existed only one type of force, the superforce, and one type of particle, the superparticle; the superforce is the force from which the gravitational, strong, weak and electromagnetic forces subsequently separated due to symmetry breaking, and the superparticle likewise became differentiated into the various types of bosons and fermions. Following the Planck era there is the GUT era from 10^{-43} to 10^{-35} second

[9]Although the space of the singularity is standardly defined as less than 3-d, Roger Penrose has proposed a definition of the cosmological singularity as a 3-d spacelike surface, in which case it could count as a part of the universe and thus as its beginning. See Penrose (1974).

(the inflationary expansion occurs at 10^{-35} second at the end of the GUT era), the electroweak era from 10^{-35} second to 10^{-10} second, the free quark era from 10^{-10} second to 10^{-4} second, and so on up to the present.

In order to eliminate possible confusion about the identity of this Big Bang, it must be noted that it is the explosion of 4-dimensional space-time out of the point(s) that exist(s) at t_0; there are other explosions from points that exist at instants later than t_0. The Big Bang that explodes from the singularity at t_0 is the first Big Bang, and can be designated as "the Big Bang_1". It is the Big Bang_1 that is the beginning of the universe.

The Big Bang_1 is "the first state of the universe" in two senses. In the first sense, the Big Bang_1 begins and ends before every other state of a different kind begins; in this sense, the first state of the universe is a nonoverlapped state. In the second sense, the Big Bang_1 begins before every other Big Bang begins; in this sense the first state of the universe is a partially overlapped state, for it is likely that other Big Bangs begin before the Big Bang_1 ends.

(iii) A third possible definition is that the universe began with the Big Bang_1 at the earliest interval of 10^{-43} second, but that it did not begin at or after the earliest singularity. This does not entail that the universe began before this singularity, for it is possible for the universe to begin neither before, at nor after this singularity; this possibility is actual if there is no time at which the singularity exists. The concept of a singularity, on this view, is a limiting concept that refers to nothing existent. The prediction by the Friedmann equations of a time t_0 when $a = 0$ is interpreted as a prediction of a limit to time and to the radius of the universe. "t_0" does not refer to a time but expresses a concept of an ideal limit that past times can approach with arbitrary closeness but can never reach; every actual time t is such that $t > t_0$. The same holds for the concept of $a = 0$, and the concepts of infinite density, temperature and curvature.

An alternate explication of the definition of the universe as beginning "neither before, at nor after the singularity" is implied by some remarks of Richard Swinburne. According to this explication, there is no time at which the singularity exists, but there is time, empty time, prior to the Big Bang_1. If t_0 is the time at which the first singularity would have existed had it existed, then the Friedmann equations can be taken as predicting that "the Universe must have come into being after t_0" (Swinburne 1981, p. 254). Paul Fitzgerald interprets Swinburne to mean something that Fitzgerald takes to be absurd, that "the universe popped into being, preceded by a finite empty lapse of time!" (Fitzgerald 1976, p. 635). But this is a misinterpretation of Swinburne, for Swinburne argues that it is logically necessary for past time to be infinite (Swinburne 1981, pp. 172–173), and therefore that if the universe began at $t > t_0$ there must be an infinite amount of empty time prior to the Big Bang_1 or to t_0.

I have elsewhere shown that Swinburne's and others' putative proofs of the necessary infinitude of time are fallacious, so Swinburne's characterization of the beginning of the universe need not be accepted (Smith 1985c). But that is not to say that it is logically impossible for the Big Bang_1 to have occurred after an infinite (or finite) period of empty time, as J. G. Whitrow (1980, p. 32) and others have argued it to be; this *is* logically possible.[10] What is pertinent to my present investigation is that this is *improbable,* given that the empirically established cosmological theories that predict a beginning of the universe do so by predicting a beginning of time (this will be proven in the next section).

If we reject the "empty time" explication of the third definition of the beginning of the universe, that leaves us with two seemingly viable definitions, the second and the third as originally explicated. I shall assume the second definition to be the correct one, as it treats the singularity as real and thus complies with the Hawking-Penrose singularity theorems. The few objections that have been made to the reality of the singularity since the development of the singularity theorems have been based for the most part on philosophical grounds, and do not seem very convincing.[11] In any case, I will show that the conclusion that the universe spontaneously began follows no less if the third definition is used.

3. Arguments that the First Singularity and the Big Bang_1 are Uncaused. The idea that the Friedmann equations and the Hawking-Penrose singularity theorems predict an uncaused beginning of the universe is resisted by many philosophers. W. H. Newton-Smith writes:

> . . . supposing that the Big Bang emerged from a singularity of infinite density, it is hard to see what would constitute a reason for denying that that singularity itself emerge from some prior cosmological goings-on. And as we have reasons for supposing that macroscopic events have causal origins, we have reason to suppose that some prior state of the universe led to the production of this particular singularity. (1980, p. 111)

This argument fails on several accounts. Note first that

> (1) We have reason for supposing that macroscopic events have causal origins

entails

[10]I have shown that it is logically possible for there to be empty time before the Big Bang_1 in Smith (1985a).

[11]A frequent objection is that singularities involve infinite values and that infinities cannot be real. See for example Craig (1979, pp. 116–117). I have rebutted Craig's and others' arguments against infinite realities in Smith (1987).

(2) We have reason for supposing that the cosmological singularity has a causal origin

only given the additional premise

(3) The cosmological singularity is a macroscopic event

which is false, for the singularity, far from being a macroscopic event, is infinitely smaller than the smallest *microscopic* event that physicists have yet detected. Moreover, the singularity is not even an *event*, that is, a point in 4-dimensional space-time; it is not a part of but a boundary or edge of the 4-d-space-time continuum.

Furthermore, it belongs analytically to the concept of the cosmological singularity that it is not the effect of prior physical events. The definition of a singularity that is employed in the singularity theorems entails that it is *impossible* to extend the space-time manifold beyond the singularity. The definition in question is based on the concept of inextendible curves, a concept that has been most completely and precisely explicated by B. G. Schmidt (1971). In a space-time manifold there are timelike geodesics (paths of freely falling particles), spacelike geodesics (paths of tachyons), null geodesics (paths of photons), and timelike curves with bounded acceleration (paths along which it is possible for observers to move). If one of these curves terminates after a finite proper length (or finite affine parameter in the case of null geodesics), and it is impossible to extend the space-time manifold beyond that point (for example, because of infinite curvature), then that point, along with all adjacent terminating points, is a singularity. Accordingly, if there is some point *p* beyond which it is possible to extend the space-time manifold, beyond which geodesics or timelike curves can be extended, then *p* by definition is not a singularity.

This effectively rules out the idea that the singularity is an effect of some prior natural process. A more difficult question is whether or not the singularity or the Big Bang probably is an effect of a supernatural cause, God. I will consider first the question of whether the Big Bang_1 is probably supernaturally caused. This fits in with W. L. Craig's argument for a divine causality of the beginning of the universe, for Craig rejects the singularity as unreal and treats the Big Bang as the first physical state (Craig does not distinguish among the several Big Bangs) (Craig 1979). Craig's argument includes the steps

(4) We have reason to believe that all events have a cause
(5) The Big Bang is an event (or a set of events)
(6) Therefore, we have reason to believe that the Big Bang has a cause

Additional steps are introduced to show that the cause of the Big Bang probably is a personal Creator.

An argument of this sort avoids the problems in Newton-Smith's argument, for it does not argue from macroscopic events to something that is neither macroscopic nor an event, but argues from events in general to another event or set of events, the Big Bang. Furthermore, it does not violate the singularity theorems in supposing that the space-time manifold is extended beyond a singularity.

Nevertheless, the argument fails because its first premise, (4), is false. Craig writes of (4):

> Constantly verified and never falsified, the causal proposition may be taken as an empirical generalization enjoying the strongest support experience affords. (1979, p. 145)

However, quantum mechanical considerations show that the causal proposition is limited in its application, if applicable at all, and consequently that a probabilistic argument for a cause of the Big Bang cannot go through. It is not relevant to the demonstration of this fact whether the causal relation be analyzed in terms of physical necessity or in terms of the regular but nonnecessary conjunction of events of a certain kind. Either analysis may be assumed. It is sufficient to understand causality in terms of a law enabling single predictions to be deduced, precise predictions of individual events or states. That there are uncaused events in this sense follows from Heisenberg's uncertainty principle, which states that for conjugate magnitudes such as the position q and momentum p of a particle, it is impossible in principle to measure both simultaneously with precision. If p lies within a certain interval of length Δp, and q lies within a certain interval of length Δq, then if Δp is made very small (measured exactly), Δq cannot at the same time be made very small (measured exactly). Exactly put, the product of Δp and Δq cannot be made smaller than Planck's constant h divided by 4π, so that

$$\Delta p \cdot \Delta q \geq h/4\pi \tag{7}$$

Now if the initial conditions such as p and q of a particle x cannot all be known precisely at time t_1, then the subsequent conditions of x at time t_2 cannot be precisely predicted. The prediction of the conditions of the conjugate magnitudes of x at t_2 must be statistical and indeterministic. For example the position of x at t_2 is represented in terms of various possible positions each with a different probability value, such that none of these values is able to arbitrarily approach 1. These predictions are effected by means of the Schrodinger wave function ψ; the square of the amplitude of ψ at each point of ψ determines the probability distribution of condition q of x at t_2. If by $d(q, t_2)$ we mean the probability distribution of q at t_2, this can be calculated as

$$d(q,t_2) = |(q,t_2)|^2 \tag{8}$$

Equations (7) and (8) at most tend to show that acausal laws govern the *change of condition* of particles, such as the change of particle x's position from q_1 to q_2. They state nothing about the causality or acausality of absolute beginnings, of beginnings of the existence of particles. Consequently, supposing that with suitable additional premises we can draw inferences from (7) and (8) to the whole universe, the only relevant argument that we could show to be unsuccessful is

(9) We have reason to believe that all changes of condition are caused
(10) Therefore, we have reason to believe that all changes of condition of the universe as a whole are caused

and the failure of this argument does not entail the failure of

(11) We have reason to believe that all beginnings of existence are caused
(12) Therefore, we have reason to believe that the beginning of the universe's existence is caused.

Thus if the latter argument is to be refuted, it is necessary to find premises more relevant to absolute beginnings than (7) and (8).

Such premises can be obtained on the basis of Heisenberg's uncertainty relation

$$\Delta E \cdot \Delta t \geq h/4\pi \tag{13}$$

where $E \equiv$ energy, $t \equiv$ time and $h \equiv$ Planck's constant. This relation implies that if the energy of a particle is measured precisely, so that ΔE is made very small, the time at which the particle possesses this energy can be known only imprecisely, so that Δt is very large. Now if Δt is small enough, ΔE becomes so large that it becomes impossible in principle to determine if the law of energy conservation is violated. During the interval of time

$$\Delta t \simeq (h/4\pi)\Delta E \tag{14}$$

this law is inapplicable and consequently an amount of energy ΔE can spontaneously come into existence and then (before the interval has elapsed) cease to exist. There is observational evidence, albeit indirect, that this uncaused emergence of energy or particles (notably virtual particles) frequently occurs. It appears, then, that the argument (11)-(12) is unsuccessful and that the crucial step in the argument to a supernatural cause of the Big Bang, or more exactly, of the Big $Bang_1$, is faulty.

It might be objected that quantum acausality applies only on the microscopic level and not to macroscopic cosmological states or beginnings. This could be granted without detriment to my argument, for the physical processes constitutive of the Big $Bang_1$ (which I have defined to occur during the Planck era) are one and all microscopic and occur at dimensions where quantum mechanical principles unquestionably apply.

It might then be objected that the Big $Bang_1$ is the beginning of the existence of 4-d space-time itself, and that the uncaused events I have specified involve merely the beginnings of existence *within* 4-d space-time. Surely "There are some uncaused beginnings of existence within space-time" is irrelevant to and thus cannot increase the probability value of "The beginning of the existence of space-time itself is uncaused".

My response is that if this is so (and I will provide reasons for doubting that this is the case in Section 4 with my discussion of the vacuum fluctuations models of the universe), then the same holds for the parallel argument for a supernatural cause of 4-d space-time; for "There are some caused beginnings of existence within space-time" or even "All beginnings of existence within space-time are caused" would by the same token be irrelevant to and thus fail to increase the probability of "The beginning of the existence of 4-d space-time is caused".

I conclude that quantum mechanical considerations show that the argument to a divine cause of the Big $Bang_1$ based on the causal principle (4) is unsuccessful.

But this does not end the matter, for it is still open to a defender of the theistic argument to claim that I have no right to introduce quantum mechanical acausality into the discussion, since the Big Bang cosmological model is based on GTR and GTR presupposes a causal determinism to operate in the domain of its application.

I shall not decide whether this objection is valid, but will instead show that even if it is valid it still can be proven that we have no reason to think the Big $Bang_1$ is caused, supernaturally or otherwise. This can be demonstrated solely on the basis of the GTR Big Bang cosmological model itself.

Let us begin by assuming what I have already maintained to be the case, that the singularity at t_0 is real. Given that, we can note that the classical notions of space and time and all known laws of physics (since they are formulated on a classical space-time background) break down at the singularity, and consequently it is impossible to predict what will emerge from the singularity. This impossibility is not due to our ignorance of the correct theory but is a limitation upon possible knowledge that is similar but additional to the limitation entailed by the quantum mechanical uncertainty principle. The former limitation is due to the causal structure of space-time that is postulated by GTR; the interaction region postulated by GTR can be bounded not only by an initial surface on which data are given and a final surface on which measurements are made *but also by a hidden surface*. A hidden surface is one about which any possible observer can have only limited information, such as (in the case of black holes) the mass, angular momentum and charge. This surface "emits with equal probability all configurations of particles compatible with the observers limited knowledge" (Hawking 1976, p. 2460). A surface close to the Big Bang$_1$ singularity, a surface at the Planck time 10^{-43} second, is a hidden surface. The singularity hidden by this surface "would thus emit all configurations of particles with equal probability" (Hawking 1976, p. 2463). If we assume with Craig that the singularity is unreal, and that the first physical state is the Big Bang$_1$, then the hidden surface is not taken to be subsequent to the singularity; instead of the particles being regarded as randomly and spontaneously being emitted from the singularity, they are regarded as randomly and spontaneously being emitted from nothing at all. This means, precisely put, that if the Big Bang$_1$ is the first physical state, then every configuration of particles that does constitute or might have constituted this first state is as likely on *a priori* grounds to constitute it as every other configuration of particles. In either case, the constitution of the Big Bang$_1$ is impossible in principle to predict and thus is uncaused (for "uncaused" minimally means "in principle unpredictable").

The singularity itself is also regarded in the GTR-based Big Bang theory as uncaused, although for a different reason. It is defined as a point beyond which space-time curves cannot be extended, and thus which cannot have causal antecedents.

In sum, then, we may say that although the GTR-based Big Bang theory does suppose causality to operate in its domain of application, it also supposes that there is a limit to its operation; it represents causality as breaking down at the initial physical states, the singularity and the Big Bang. Consequently, this theory cannot be used to support the thesis that the initial physical states are probably caused and that this cause is God.

4. Quantum Gravity and the Uncaused Beginning of the Universe. There is a serious lacuna in my foregoing account of the beginning of the universe; I have been presuming that the GTR-based Big Bang theory has an unlimited application and therefore applies to the extreme conditions of the universe during the Big Bang$_1$. In fact, GTR fails to apply when quantum mechanical interactions predominate, and these predominate when the temperature is at or above $10^{32}K$, when the density is at or above 10^{94} gm cm^{-3}, and when the radius of curvature becomes of the order of 10^{-33} cm. Since these conditions obtain during the Planck era at the first 10^{-43} second after the singularity, the GTR-based Big Bang theory cannot be used as a reliable guide in reconstructing the physical processes that occurred during this time and *a fortiori* cannot be used as a reliable basis for predicting that the density, temperature and curvature reached infinite values prior to this time. Accordingly it seems that the foregoing probabilistic argument to an uncaused beginning of the universe is in jeopardy.

I believe, however, that there are three reasons for a continued support of the idea that the universe spontaneously began to exist. To comprehend these reasons, we must first observe that the reason why GTR is inapplicable during the Planck era is that the theory of gravity in GTR is unable to account for the quantum mechanical behavior of gravity during this era. A new *quantum theory of gravity* is needed. Although such a theory has not yet been developed, there are some general indications of what it may predict. It is in terms of these indications that our three reasons are to be understood.

First, it is thought that a quantum theory of gravity may show gravity to be repulsive rather than attractive under conditions that obtain during the Planck era. During this time regions of negative energy density may be created by the forces and particles present, and these regions lead to a gravitational repulsion. This suggests that any given finite set of past-directed timelike or null geodesics will not converge in a single point but will be pushed apart, as it were, by the repulsive gravitational force. This possibility is consistent with an oscillating universe, for as each contracting phase ends gravity becomes repulsive and prevents converging geodesics from terminating in a point; gravity repels them so that they enter a new expanding phase.

But this way of avoiding the singularity predicted by the Hawking-Penrose theorems does not give us a universe that is infinitely old. For—and this is the first of the three reasons I want to mention—this oscillating quantum-gravitational universe would still be subject to the same problems that were discussed in Section One, namely, increase in radius, length of cycle, radiation and entropy with each new cycle. Consequently, this theory does no more than push the cosmological singularity further into the past, at a time just before (or at) the beginning of the first cycle when the radius of the universe is zero (or near zero).

The second reason is that there is a way in which the Hawking-Penrose theorems' prediction of a singularity at the beginning of the present expansion can be made consistent with a quantum theory of repulsive gravity. These theorems do not *define* a singularity as that wherein curvature, density and temperature are infinite and the radius is zero. A singularity is defined as a point or series of points beyond which the space-time manifold cannot be extended. Consequently, if the effects of quantum gravity prevent a build up of temperature, density and curvature to infinite values, and a decrease of radius to zero, this need not mean there is no singularity at the beginning of the present expansion. The singularity could occur with *finite* and *nonzero* values.

The third reason is that the most theoretically developed attempts to account for the past of the universe on the basis of specifically quantum mechanical principles have represented the universe as spontaneously beginning at the onset of the present expansion. These theories are collectively known as the "vacuum fluctuation models of the universe". The models developed by Tryon (1973), Brout, Englert and Gunzig (1978), Grishchak and Zeldovich (1982), Atkatz and Pagels (1982), and Gott (1982) picture the universe as emerging spontaneously from an empty background space, and the model of Vilenkin (1982) depicts it as emerging without cause from nothing at all.

The first vacuum fluctuation model was developed by Edward Tryon in 1973. A vacuum fluctuation is an uncaused emergence of energy out of empty space that is governed by the uncertainty relation $\Delta E \cdot \Delta t \geq h/4\pi$, and which thus has zero net value for conserved quantities. Tryon argues that the universe is able to be a fluctuation from a vacuum in a larger space in which the universe is embedded since it does have a zero net value for its conserved quantities. Observational evidence (Tryon claims) supports or is consistent with the fact that the positive mass-energy of the universe is cancelled by its negative gravitational potential energy, and that the amount of matter created is equal to the amount of antimatter. (But this last point is inconsistent with current Grand Unified Theories.)

A disadvantage of Tryon's theory, and of other theories that postulate a background space from which the universe fluctuates, is that they explain the existence of the universe but only at the price of introducing another unexplained given, namely, the background space. This problem is absent from Vilenkin's theory, which represents the universe as emerging without a cause "from literally *nothing*" (1982, p. 26). The universe appears in a quantum tunneling from nothing at all to de Sitter space. Quantum tunneling is normally understood in terms of processes *within* space-time; an electron, for example, tunnels through some barrier if the electron lacks sufficient energy to cross it but nevertheless still does cross it. This is possible because the abovementioned uncertainty relation allows the electron to spontaneously acquire the additional energy for the short period of time required for it to tunnel through the barrier. Vilenkin applies this concept to space-time itself; in this case, there is not a state of the system before the tunneling, for the state of tunneling is the first state that exists. The state of tunneling thus is the analogue of the Big Bang_1 in the third definition of the beginning of the universe offered in Section 2, for it is the first state of the universe and there is no time before this state. The equation describing this state is a quantum tunneling equation, specifically the bounce solution of the Euclidean version of the evolutionary equation of a universe with a closed Robertson-Walker metric.[12] The universe emerged from the tunneling with a finite size ($a = H^{-1}$) and with a zero rate of expansion or contraction ($da/dt = 0$). It emerged in a symmetric vacuum state, which then decays and the inflationary era begins; after this era ends, the universe evolves according to the standard Big Bang model.

These quantum mechanical models of the beginning of the universe are explanatorily superior in one respect to the standard GTR-based Big Bang models; they do not postulate initial states at which the laws of physics break down but explain the beginning of the universe in accordance with the laws of physics. The GTR-based theory predicts a beginning of the universe by predicting initial states at which the laws of the theory that are used to predict these states break down. The singularity and the explosion of 4-d space-time from the singularity obey none of the laws of GTR that are obeyed by states within the universe or subsequent states of the universe. In contrast, the quantum mechanical theories represent the universe as coming into existence *via* the same laws that processes within the universe obey. Instead of an exploding singularity, there is a quantum fluctuation or tunneling that is analogous to the fluctuations or tunnelings within the universe and that obeys the same acausal laws as the latter fluctuations or tunnelings.[13]

This review of the role of quantum mechanics in accounts of the beginning of the universe strongly suggests that the probabilistic argument to an uncaused beginning of the universe, although more complicated than we had been supposing in Sections 1–3, still goes through. Its conclusion is summarized in this disjunctive statement: it is probably true that EITHER the universe began without cause at the beginning of this expansion (a) subsequent to a singularity of infinite density, temperature and curvature, and zero radius, or (b) at a singularity with finite and nonzero values, or (c) in a vacuum fluctuation from a larger space or a tunneling from nothing, OR the universe spontaneously began to exist at the beginning of some prior expansion phase under conditions described in (a), (b) or (c).

REFERENCES

Atkatz, D., and Pagels, H. (1982), "Origin of the Universe as a Quantum Tunneling Event", *Physical Review D 25*: 2065–2073.

[12]The bounce solution is $a(t) = H^{-1} \cos(Ht)$. See Vilenkin (1982, p. 26).

[13]For further discussion of the vacuum fluctuation theories of the beginning of the universe, see Smith (1986b, esp. pp. 81–84). Other pertinent cosmological discussions can be found in Smith (1985b) and Smith (1986a, chapter VI).

Barrow, J., and Silk, J. (1983), *The Left Hand of Creation*. New York: Basic Books.
Brout, R., Englert, F., and Gunzig, E. (1978), "The Creation of the Universe as a Quantum Phenomenon", *Annals of Physics* (N.Y.) *115*: 78–106.
Craig, W. L. (1979), *The Kalam Cosmological Argument*. New York: Harper and Row.
Fitzgerald, P. (1976), "Discussion Review: Swinburne's *Space and Time*", *Philosophy of Science 43*: 618–637.
Gott, J. R. (1982), "Creation of Open Universes from de Sitter Space", *Nature 295*: 304–307.
Grishchak, L. P., and Zeldovich, Y. B. (1982), "Complete Cosmological Theories", in M. J. Duff and C. J. Isham (eds.), *Quantum Structure of Space and Time*. Cambridge: Cambridge University Press, pp. 409–422.
Hawking, S. W. (1966), "Singularities in the Universe", *Physical Review Letters 17*: 444–445.
———. (1976), "Breakdown of Predictability in Gravitational Collapse", *Physical Review D 14*: 2460.
Hawking, S. W., and Ellis, G. F. R. (1973), *The Large Scale Structure of Space-Time*. Cambridge: Cambridge University Press.
Hawking, S. W., and Penrose, R. (1965), "Singularities in Homogenous World Models", *Physical Letters 17*: 246–247.
———. (1970), "The Singularities of Gravitational Collapse and Cosmology", *Proceedings of the Royal Society of London A 314*: 529–548.
Hubble, E. (1929), "A Relation Between Distance and Radial Velocity Among Extragalactic Nebulae", *Proceedings of the National Academy of Sciences 15*: 168–173.
Landsberg, P. T., and Park, D. (1975), "Entropy in an Oscillating Universe", *Proceedings of the Royal Society of London A 346*: 485–495.
Lifschitz, E. M., and Khalatnikov, I. M. (1963), "Investigations in Relativist Cosmology", *Advances in Physics 12*: 185–249.
Misner, C. W., Thorne, K. S., and Wheeler, J. A. (1973), *Gravitation*. San Francisco: W. H. Freeman and Co.
Newton-Smith, W. H. (1980), *The Structure of Time*. London: Routledge and Kegan Paul.
Penrose, R. (1965), "Gravitational Collapse and Space-Time Singularities", *Physical Review Letters 14*: 57–59.
———. (1974), "Singularities in Cosmology", in M. S. Longair (ed.), *Confrontation of Cosmological Theories with Observational Data*, Boston: D. Reidel, pp. 263–272.
Penzias, A., and Wilson, R. (1965), "A Measure of Excess Antenna Temperature at 4080 Mc/s", *Astrophysical Journal 142*: 419–421.
Schmidt, B. G. (1971), "A New Definition of Singular Points in General Relativity", *General Relativity and Gravity 1*: 269–280.
Silk, Joseph (1980), *The Big Bang*. San Francisco: W. H. Freeman and Co.
Smith, Q. (1985a), "Kant and the Beginning of the World", *The New Scholasticism 59*: 339–346.
———. (1985b), "The Anthropic Principle and Many-Worlds Cosmologies", *The Australasian Journal of Philosophy 63*: 336–348.
———. (1985c), "On the Beginning of Time", *Noûs 19*: 579–584.
———. (1986a), *The Felt Meanings of the World: A Metaphysics of Feeling*. West Lafayette, Indiana: Purdue University Press.
———. (1986b), "World Ensemble Explanations", *Pacific Philosophical Quarterly 67*: 73–86.
———. (1987), "Infinity and the Past", *Philosophy of Science 54*: 63–75.
Stromberg, G. (1925), "Analysis of Radial Velocities of Globular Clusters and Non-Galactic Nebulae", *Astrophysical Journal 61*: 353–362.
Swinburne, R. (1981), *Space and Time*, 2nd ed. New York: St. Martin's Press.
Tolman, R. C. (1934), *Relativity, Thermodynamics and Cosmology*. Oxford: Clarendon Press.
Tryon, E. P. (1973), "Is the Universe a Vacuum Fluctuation?", *Nature 246*: 396–397.
Vilenkin, A. (1982), "Creation of Universes from Nothing", *Physical Letters 117B*: 25–28.

Wheeler, J. (1973), "From Relativity to Mutability", in J. Mehra (ed.), *The Physicist's Conception of Nature*. Boston: D. Reidel, pp. 202–247.
Whitrow, J. G. (1980), *The Natural Philosophy of Time*, 2nd ed. Oxford: Clarendon Press.