3

The *Kalam* Cosmological Argument

WILLIAM LANE CRAIG AND JAMES D. SINCLAIR

Introduction

The cosmological argument is a family of arguments that seek to demonstrate the existence of a Sufficient Reason or First Cause of the existence of the cosmos. The roll of the defenders of this argument reads like a *Who's Who* of Western philosophy: Plato, Aristotle, ibn Sina, al-Ghazali, Maimonides, Anselm, Aquinas, Scotus, Descartes, Spinoza, Leibniz, and Locke, to name but some. Cosmological arguments can be conveniently grouped into three basic types: the *kalam* cosmological argument for a First Cause of the beginning of the universe; the Thomist cosmological argument for a sustaining Ground of Being of the world; and the Leibnizian cosmological argument for a Sufficient Reason why something exists rather than nothing.[1]

The *kalam* cosmological argument traces its roots to the efforts of early Christian theologians who, out of their commitment to the biblical teaching of *creatio ex nihilo*, sought to rebut the Aristotelian doctrine of the eternity of the universe. In his works *Against Aristotle* and *On the Eternity of the World against Proclus*, the Alexandrian Aristotelian commentator John Philoponus (d. 580?), the last great champion of *creatio ex nihilo* prior to the advent of Islam, initiated a tradition of argumentation in support of the doctrine of creation based on the impossibility of an infinite temporal regress of events (Philoponus 1987; Philoponus & Simplicius 1991). Following the Muslim conquest of North Africa, this tradition was taken up and subsequently enriched by medieval Muslim and Jewish theologians before being transmitted back again into Christian scholastic theology.[2]

In light of the central role played by this form of the cosmological argument in medieval Islamic theology, as well as the substantive contribution to its development by its medieval Muslim proponents, we use the word "*kalam*" to denominate this version of the argument. The Arabic word for speech, *kalam* was used by Muslim thinkers to denote a statement of

---

[1] This typology has become somewhat standard (*Routledge Encyclopedia of Philosophy* 1998; cf. *Stanford Encyclopedia of Philosophy* 2004a).

[2] For an exposition of the argument in its historical context, see Craig (1980), Wolfson (1966), Wolfson (1976), Davidson (1987), and Dales (1990).

The Blackwell Companion to Natural Theology Edited William Lane Craig and J. P. Moreland

ISBN: 978-1-405-17657-6

theological doctrine and eventually a statement of any intellectual position or an argument supporting such a position. According to the fourteenth-century Muslim theologian al-Idji, *kalam* is "the science which is concerned with firmly establishing religious beliefs by adducing proofs and banishing doubts" (al-Idji 1971). Ultimately, *kalam* became the name of the whole movement within Muslim thought that might best be described as Islamic scholasticism. A practitioner of *kalam* is called a *mutakallim* (pl. *mutakallimun*). Jewish theologians in Muslim Spain, who rubbed shoulders both with the Arabic East and the Latin West, were the means by which the *kalam* cosmological argument found its way back into Christian thought. The subject of extended debate, the argument pitted al-Ghazali against ibn Rushd, Saadia ben Gaon against Maimonides, and Bonaventure against Aquinas. The debate was eventually enshrined during the modern era in the thesis and antithesis of Kant's First Antinomy concerning time.

After suffering several centuries of eclipse, the argument has enjoyed a resurgence of interest in recent decades, doubtlessly spurred by the startling empirical evidence of contemporary astrophysical cosmology for a beginning of space and time. *Kalam* philosophical argumentation for the finitude of the past played a key role in the philosophy of time propounded by mathematician and cosmologist G. J. Whitrow (1980). As a piece of natural theology, the *kalam* argument was revived by Stuart Hackett in his little-noted *The Resurrection of Theism* (1957) and subsequently brought into philosophical prominence by his student William Lane Craig (1979). Noting the widespread debate over the argument today, Quentin Smith observes, "The fact that theists and atheists alike 'cannot leave [the] Kalam argument alone' suggests that it may be an argument of unusual philosophical interest or else has an attractive core of plausibility that keeps philosophers turning back to it and examining it once again" (Smith 2007, p. 183).

What is the *kalam* cosmological argument? In his *Kitab al-Iqtisad*, the medieval Muslim theologian al-Ghazali presented the following simple syllogism in support of the existence of a Creator: "Every being which begins has a cause for its beginning; now the world is a being which begins; therefore, it possesses a cause for its beginning" (al-Ghazali 1962, pp. 15–6). In defense of the second premise, Ghazali offered various philosophical arguments to show the impossibility of an infinite regress of temporal phenomena and, hence, of an infinite past. The limit at which the finite past terminates Ghazali calls "the Eternal" (al-Ghazali 1963, p. 32), which he evidently takes to be a state of timelessness. Given the truth of the first premise, the finite past must, therefore, "stop at an eternal being from which the first temporal being should have originated" (al-Ghazali 1963, p. 33).

The argument, then, is extremely simple:[3]

1.0. Everything that begins to exist has a cause.
2.0. The universe began to exist.
3.0. Therefore, the universe has a cause.

---

[3] In view of the bewildering variety of complex issues raised by cosmological arguments, Oppy cautions that *even before looking at it*, we should conclude that so simple an argument cannot plausibly be held to establish its conclusion (Oppy 2006b, p. 173). But the complexity of the issues involved in assessing the truth of an argument's premises does not require, *pace* Oppy, that the argument itself have "dozens of complex premises." The supporting arguments and responses to defeaters of the argument's two basic premises can proliferate in an almost fractal-like fashion.

As a final step one may explore the relevance of this conclusion for theism by means of a conceptual analysis of what it is to be a cause of the universe. In the sequel we shall examine each of the steps of the argument, beginning with Premise (2.0), since this is clearly the more controversial claim and since some attempts to subvert (1.0) are based upon cosmogonic theories – the discussion of which would be premature prior to their introduction in our treatment of (2.0).

2.0. Did the Universe Begin to Exist?

The crucial second premise of the *kalam* cosmological argument has been supported by both metaphysical and physical arguments. We shall examine two traditional philosophical arguments against the existence of an infinite temporal regress of events, as well as scientific evidence in support of an absolute beginning of the universe.

2.1. Argument from the impossibility of an actual infinite

One of the traditional arguments for the finitude of the past is based upon the impossibility of the existence of an actual infinite. It may be formulated as follows:

2.11. An actual infinite cannot exist.
2.12. An infinite temporal regress of events is an actual infinite.
2.13. Therefore, an infinite temporal regress of events cannot exist.

In order to assess this argument, we need to have a clear understanding of its key terms. First and foremost among these is "actual infinite." Prior to the revolutionary work of mathematicians Bernard Bolzano (1781–1848), Richard Dedekind (1831–1916), and, especially, Georg Cantor (1845–1918), there was no clear mathematical understanding of the actual infinite (Moore 1990, pt. I). Aristotle had argued at length that no actually infinite magnitude can exist (*Physics* 3.5.204^{b}1–206^{a}8). The only legitimate sense in which one can speak of the infinite is in terms of potentiality: something may be infinitely divisible or susceptible to infinite addition, but this type of infinity is potential only and can never be fully actualized (*Physics* 8.8. 263^{a}4–263^{b}3). The concept of a potential infinite is a dynamic notion, and strictly speaking, we must say that the potential infinite is at any particular time finite.

This understanding of the infinite prevailed all the way up to the nineteenth century. But although the majority of philosophers and mathematicians adhered to the conception of the infinite as an ideal limit, dissenting voices could also be heard. Bolzano argued vigorously against the then current definitions of the potential infinite (Bolzano 1950, pp. 81–4). He contended that infinite multitudes can be of different sizes and observed the resultant paradox that although one infinite might be larger than another, the individual elements of the two infinites could nonetheless be matched against each other in a one-to-one correspondence (Bolzano 1950, pp. 95–6).[4] It was precisely this paradoxical notion that Dedekind seized upon in his definition of the infinite: a system is said to be infinite if a part of

---

[4] Despite the one-to-one correspondence, Bolzano insisted that two infinites so matched might nevertheless be nonequivalent.

that system can be put into a one-to-one correspondence with the whole (Dedekind 1963, p. 63). According to Dedekind, the Euclidean maxim that the whole is greater than a part holds only for finite systems.

But it was undoubtedly Cantor who won for the actual infinite the status of mathematical legitimacy that it enjoys today. Cantor called the potential infinite a "variable finite" and attached the sign ∞ (called a lemniscate) to it; this signified that it was an "improper infinite" (Cantor 1915, pp. 55–6). The actual infinite he pronounced the "true infinite" and assigned the symbol $\aleph_0$ (aleph zero) to it. This represented the number of all the numbers in the series 1, 2, 3, . . . and was the first infinite or transfinite number, coming after all the finite numbers. According to Cantor, a collection or set is infinite when a part of it is equivalent to the whole (Cantor 1915, p. 108). Utilizing this notion of the actual infinite, Cantor was able to develop a whole system of transfinite arithmetic. "Cantor's . . . theory of *transfinite* numbers . . . is, I think, the finest product of mathematical genius and one of the supreme achievements of purely intellectual human activity," exclaimed the great German mathematician David Hilbert. "No one shall drive us out of the paradise which Cantor has created for us" (Hilbert 1964, pp. 139, 141).

Modern set theory, as a legacy of Cantor, is thus exclusively concerned with the actual as opposed to the potential infinite. According to Cantor, a set is a collection into a whole of definite, distinct objects of our intuition or of our thought; these objects are called elements or members of the set. Fraenkel draws attention to the characteristics *definite* and *distinct* as particularly significant (Fraenkel 1961, p. 10). That the members of a set are distinct means that each is different from the others. To say that they are definite means that given a set S, it should be intrinsically settled for any possible object *x* whether *x* is a member of S or not. This does not imply actual decidability with the present or even future resources of experience; rather a definition could settle the matter sufficiently, such as the definition for "transcendental" in the set of all transcendental numbers.

Unfortunately, Cantor's notion of a set as any logical collection was soon found to spawn various contradictions or antinomies within the naive set theory that threatened to bring down the whole structure. As a result, most mathematicians have renounced a definition of the general concept of set and chosen instead an axiomatic approach to set theory, by means of which the system is erected upon several given, undefined concepts formulated into axioms. An infinite set in the Zermelo–Fraenkel axiomatic set theory is defined as any set *R* that has a proper subset that is equivalent to *R*. A proper subset is a subset that does not exhaust all the members of the original set, that is to say, at least one member of the original set is not also a member of the subset. Two sets are said to be equivalent if the members of one set can be related to the members of the other set in a one-to-one correspondence, that is, so related that a single member of the one set corresponds to a single member of the other set and vice versa. Equivalent sets are regarded as having the same number of members. This convention has recently been dubbed as Hume's Principle[5] (on the basis of Hume 1978, bk, I, pt. iii, sec. 1, p. 71). An infinite set, then, is one that is such that the whole set has the same number of members as a proper subset. In contrast to this, a finite set is a set that is such that if *n* is a positive integer, the set has *n* members. Because set theory does not utilize the notion of potential infinity, a set containing a potentially infinite number of members is impossible. Such a collection would be one in which the

---

[5] The appellation is due to Boolos (1986–7).

membership is not definite in number but may be increased without limit. It would best be described as indefinite. The crucial difference between an infinite set and an indefinite collection would be that the former is conceived as a determinate whole actually possessing an infinite number of members, while the latter never actually attains infinity, although it increases perpetually. We have, then, three types of collection that we must keep conceptually distinct: finite, infinite, and indefinite.

When we use the word "exist," we mean "be instantiated in the mind-independent world." We are inquiring whether there are extratheoretical correlates to the terms used in our mathematical theories. We thereby hope to differentiate the sense in which existence is denied to the actual infinite in (2.11) from what is often called "mathematical existence." Kasner and Newman strongly differentiate the two when they assert, "'Existence' in the mathematical sense is wholly different from the existence of objects in the physical world" (Kasner & Newman 1940, p. 61). "Mathematical existence" is frequently understood as roughly synonymous with "mathematical legitimacy." Historically, certain mathematical concepts have been viewed with suspicion and, therefore, initially denied legitimacy in mathematics. Most famous of these are the complex numbers, which as multiples of $\sqrt{-1}$, were dubbed "imaginary" numbers. To say that complex numbers exist in the mathematical sense is simply to say that they are legitimate mathematical notions; they are in that sense as "real" as the real numbers. Even negative numbers and zero had to fight to win mathematical existence. The actual infinite has, similarly, had to struggle for mathematical legitimacy. For many thinkers, a commitment to the mathematical legitimacy of some notion does not bring with it a commitment to the existence of the relevant entity in the non-mathematical sense. For formalist defenders of the actual infinite such as Hilbert, mere logical consistency was sufficient for existence in the mathematical sense. At the same time, Hilbert denied that the actual infinite is anywhere instantiated in reality. Clearly, for such thinkers, there is a differentiation between mathematical existence and existence in the everyday sense of the word. We are not here endorsing two modes of existence but simply alerting readers to the equivocal way in which "existence" is often used in mathematical discussions, lest the denial of existence of the actual infinite in (2.11) be misunderstood to be a denial of the mathematical legitimacy of the actual infinite. A modern *mutakallim* might deny the mathematical legitimacy of the actual infinite in favor of intuitionistic or constructivist views of mathematics, but he need not. When Kasner and Newman say, "the infinite certainly does not exist in the same sense that we say, 'There are fish in the sea'" (Kasner & Newman 1940, p. 61), *that* is the sense of existence that is at issue in (2.11).

These remarks make clear that when it is alleged that an actual infinite "cannot" exist, the modality at issue is not strict logical possibility. Otherwise the presumed strict logical consistency of axiomatic set theory would be enough to guarantee that the existence of an actual infinite is possible. Rather what is at issue here is so-called metaphysical possibility, which has to do with something's being realizable or actualizable. This sort of modality, in terms of which popular possible worlds semantics is typically formulated, is often characterized as broadly logical possibility, but here a word of caution is in order. Insofar as by broadly logical possibility one means merely strict logical possibility augmented by the meaning of terms in the sentence within the scope of the modal operator, such a conception is still too narrow for the purposes of the present argument. Such a conception would enable us to see the necessity of analytic truths in virtue of logic and the meaning of sentential terms used in the expression of these truths (such as "All bachelors are unmarried"), but it will not capture the metaphysical necessity or impossibility of synthetic truths, whether these be known *a priori* (such as "Everything that has a shape has a size") or *a posteriori* (such as "This table could not have been made of ice"). Broad logical possibility, then, will not be broad enough for a proper understanding of the argument unless synthetic truths are among those truths that are classed as necessary.

The fact that the argument is framed in terms of metaphysical modality also has an important epistemic consequence. Since metaphysical modality is so much woollier a notion than strict logical modality, there may not be the sort of clean, decisive markers of what is possible or impossible that consistency in first-order logic affords for strict logical modality. Arguments for metaphysical possibility or impossibility typically rely upon intuitions and conceivability arguments, which are obviously much less certain guides than strict logical consistency or inconsistency. The poorly defined nature of metaphysical modality cuts both ways dialectically: on the one hand, arguments for the metaphysical impossibility of some state of affairs will be much more subjective than arguments concerning strict logical impossibility; on the other hand, such arguments cannot be refuted by facile observations to the effect that such states of affairs have not been demonstrated to be strictly logically inconsistent.

Premise (2.12) speaks of a temporal regress of events. By an "event," one means any change. Since any change takes time, there are no instantaneous events so defined. Neither could there be an infinitely slow event, since such an "event" would, in reality, be a changeless state. Therefore, any event will have a finite, nonzero duration. In order that all the events comprised by the temporal regress of past events be of equal duration, one arbitrarily stipulates some event as our standard and, taking as our point of departure the present standard event, we consider any series of such standard events ordered according to the relation *earlier than*. The question is whether this series of events comprises an actually infinite number of events or not. If not, then since the universe cannot ever have existed in an absolutely quiescent state, the universe must have had a beginning. It is therefore not relevant whether the temporal series had a beginning *point* (a first temporal instant). The question is whether there was in the past an event occupying a nonzero, finite temporal interval which was absolutely first, that is, not preceded by any equal interval.[6]

With these explications in mind, let us now turn to an examination of the argument's two premises.

2.11. Existence of an actual infinite

Premise (2.11) asserts that an actual infinite cannot exist in the real world. It is frequently alleged that this sort of claim has been falsified by Cantor's work on the actual infinite and by subsequent developments in set theory, which provide a convincing demonstration of the existence of actual infinites. But this allegation is far too hasty. It not only begs the question against denials of the mathematical legitimacy of the actual infinite on the part of certain mathematicians (such as intuitionists), but, more seriously, it begs the question against anti-Platonist views of mathematical objects. These are distinct questions, all too

---

6. This criterion allows that there may be events of shorter duration prior to the first standard event. By stipulating as one's standard event a shorter interval, these can be made arbitrarily brief.

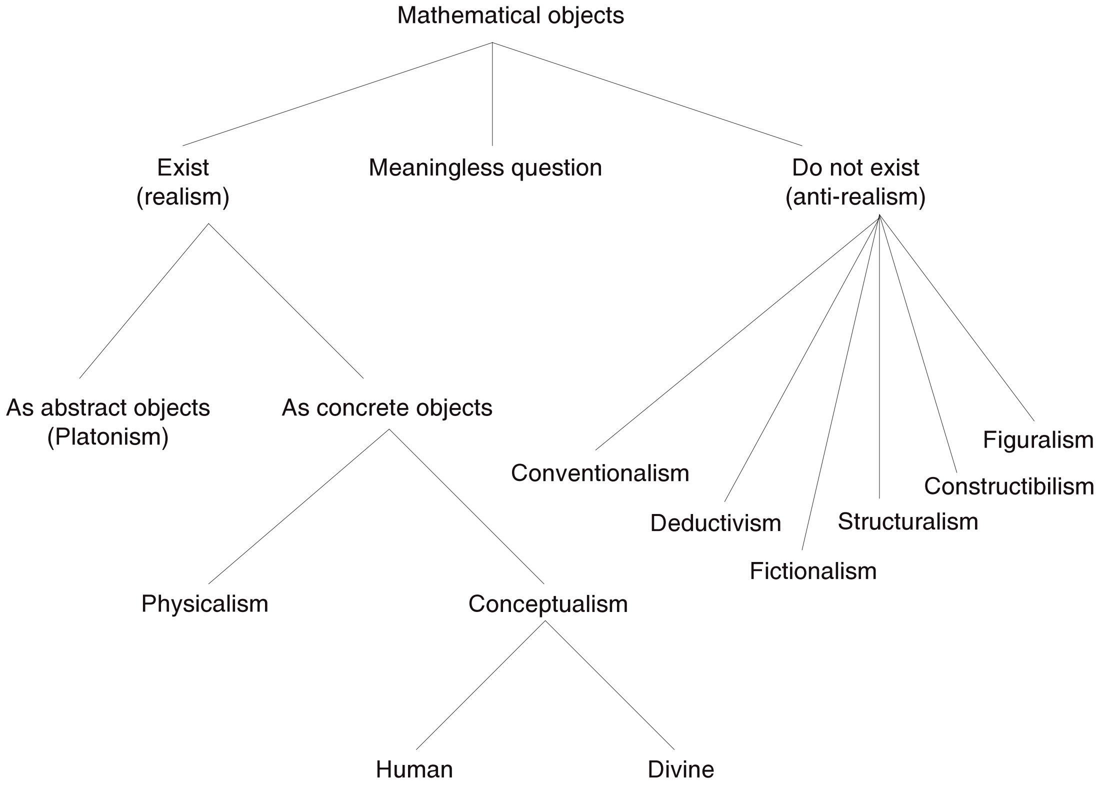


Figure 3.1 Some metaphysical options concerning the existence of abstract objects.

often conflated by recent critics of the argument (Sobel 2004, pp. 181–9, 198;9; Oppy 2006a, pp. 291–3; cf. Craig, 2008). Most non-Platonists would not go to the intuitionistic extreme of denying mathematical legitimacy to the actual infinite – hence, Hilbert's defiant declaration, "No one shall be able to drive us from the paradise that Cantor has created for us" (Hilbert 1964, p. 141) – rather they would simply insist that acceptance of the mathematical legitimacy of certain notions does not imply an ontological commitment to the reality of various objects. Thus, in Hilbert's view, "The infinite is nowhere to be found in reality. It neither exists in nature nor provides a legitimate basis for rational thought. . . . The role that remains for the infinite to play is solely that of an idea" (Hilbert 1964, p. 151). Cantor's system and axiomatized set theory may be taken to be simply a universe of discourse, a mathematical system based on certain adopted axioms and conventions, which carries no ontological commitments. In view of the plethora of alternatives to Platonism (Figure 3.1), critics of the argument cannot justifiably simply assume that the language of mathematics commits us ontologically to mind-independent entities, especially to such obscure objects as sets.

On antirealist views of mathematical objects such as Balaguer's fictionalism (Balaguer 1998, pt. II; 2001, pp. 87–114; *Stanford Encyclopedia of Philosophy* 2004b), Yablo's figuralism (Yablo 2000, pp. 275–312; 2001, pp. 72–102; 2005, pp. 88–115), Chihara's constructibilism (Chihara 1990, 2004; 2005, pp. 483–514), or Hellman's Modal structuralism (Hellman 1989; 2001, pp. 129–57; 2005, pp. 536–62), mathematical discourse is not in any way abridged, but there are, notwithstanding, no mathematical objects at all, let alone an infinite number of them. The abundance of nominalist (not to speak of conceptualist) alternatives to Platonism renders the issue of the ontological status of mathematical entities at least a moot question. The Realist, then, if he is to maintain that mathematical objects furnish a decisive counterexample to the denial of the existence of the actual infinite, must provide some overriding argument for the reality of mathematical objects, as well as rebutting defeaters of all the alternatives consistent with classical mathematics – a task whose prospects for success are dim, indeed. It is therefore open to the *mutakallim* to hold that while the actual infinite is a fruitful and consistent concept within the postulated universe of discourse, it cannot be transposed into the real world.

The best way to support (2.11) is by way of thought experiments that illustrate the various absurdities that would result if an actual infinite were to be instantiated in the real world.[7] Benardete, who is especially creative and effective at concocting such thought experiments, puts it well: "Viewed *in abstracto*, there is no logical contradiction involved in any of these enormities; but we have only to confront them *in concreto* for their outrageous absurdity to strike us full in the face" (Benardete 1964, p. 238).[8]

Let us look at just one example: David Hilbert's famous brainchild "Hilbert's Hotel."[9] As a warm-up, let us first imagine a hotel with a finite number of rooms. Suppose,

---

[7] Ludwig Wittgenstein nicely enunciated this strategy when, in response to Hilbert's solemn declaration, he quipped, "I wouldn't dream of trying to drive anyone from this paradise. I would do something quite different: I would try to show you that it is not a paradise – so that you'll leave of your own accord. I would say, 'You're welcome to this; just look about you' . . ." (Wittgenstein 1976, p. 103). But here the strategy is employed on behalf of metaphysical, not mathematical, finitism. Oppy objects that such puzzles show, at most, that certain kinds of actual infinities cannot exist, but that this conclusion cannot be generalized (Oppy 2006b, p. 140). The difficulty with this attempt to blunt the force of the absurdities is twofold: (i) nothing in the various situations seems to be metaphysically impossible apart from the assumption of an actual infinite and (ii) the absurdities are not tied to the particular kinds of objects involved.

[8] He has in mind especially what he calls paradoxes of the serrated continuum, such as the following:

> Here is a book lying on the table. Open it. Look at the first page. Measure its thickness. It is very thick indeed for a single sheet of paper – 1/2 inch thick. Now turn to the second page of the book. How thick is this second sheet of paper? 1/4 inch thick. And the third page of the book, how thick is this third sheet of paper? 1/8 inch thick, &c. *ad infinitum*. We are to posit not only that each page of the book is followed by an immediate successor the thickness of which is one-half that of the immediately preceding page but also (and this is not unimportant) that each page is separated from page 1 by a finite number of pages. These two conditions are logically compatible: there is no certifiable contradiction in their joint assertion. But they mutually entail that there is no last page in the book. Close the book. Turn it over so that the front cover of the book is now lying face down upon the table. Now – slowly – lift the back cover of the book with the aim of exposing to view the stack of pages lying beneath it. *There is nothing to see*. For there is no last page in the book to meet our gaze. (Benardete 1964, pp. 236–7).

To our mind this conclusion itself is evidently metaphysically absurd. Although Oppy, following Hazen (1993), offers expansions of the story so that someone opening the book will have some sort of visual experience, rather than as it were, a blank (Oppy 2006a, pp. 83–5), that does not negate the conclusion that there is nothing there to see since there is no last page. Benardete imagines what would happen if we tried to touch the last page of the book. We cannot do it. Either there will be an impenetrable barrier at $\omega + 1$, which seems like science fiction, or else our fingers will penetrate through an infinity of pages without first penetrating a page, which recalls Zeno's paradoxes in spades, since the pages are actual entities. What makes paradoxes such as these especially powerful, as Benardete points out, is that no process or supertask is involved here; each page is an actual entity having a finite thickness (none is a degenerate interval) which could be unbound from the others and all the pages scattered to the four winds, so that an actual infinity of pages would exist throughout space. If such a book cannot exist, therefore, neither can an actual infinite.

[9] The story of Hilbert's Hotel is related in Gamow (1946, p. 17).

furthermore, that all the rooms are occupied. When a new guest arrives asking for a room, the proprietor apologizes, "Sorry, all the rooms are full," and that is the end of the story. But now let us imagine a hotel with an infinite number of rooms and suppose once more that *all the rooms are occupied.* There is not a single vacant room throughout the entire infinite hotel. Now suppose a new guest shows up, asking for a room. "But of course!" says the proprietor, and he immediately shifts the person in room #1 into room #2, the person in room #2 into room #3, the person in room #3 into room #4, and so on out to infinity. As a result of these room changes, room #1 now becomes vacant, and the new guest gratefully checks in. But remember, before he arrived, all the rooms were occupied! Equally curious, there are now no more persons in the hotel than there were before: the number is just infinite. But how can this be? The proprietor just added the new guest's name to the register and gave him his keys – how can there not be one more person in the hotel than before?

But the situation becomes even stranger. For suppose an infinity of new guests show up at the desk, asking for a room. "Of course, of course!" says the proprietor, and he proceeds to shift the person in room #1 into room #2, the person in room #2 into room #4, the person in room #3 into room #6, and so on out to infinity, always putting each former occupant into the room number twice his own. Because any natural number multiplied by two always equals an even number, all the guests wind up in even-numbered rooms. As a result, all the odd-numbered rooms become vacant, and the infinity of new guests is easily accommodated. And yet, before they came, all the rooms were occupied! And again, strangely enough, the number of guests in the hotel is the same after the infinity of new guests check in as before, even though there were as many new guests as old guests. In fact, the proprietor could repeat this process *infinitely many times* and yet there would never be a single person more in the hotel than before.

But Hilbert's Hotel is even stranger than the German mathematician made it out to be. For suppose some of the guests start to check out. Suppose the guest in room #1 departs. Is there not now one fewer person in the hotel? Not according to infinite set theory! Suppose the guests in rooms #1, 3, 5, . . . check out. In this case an infinite number of people has left the hotel, but by Hume's Principle, there are no fewer people in the hotel. In fact, we could have every other guest check out of the hotel and repeat this process infinitely many times, and yet there would never be any fewer people in the hotel. Now suppose the proprietor does not like having a half-empty hotel (it looks bad for business). No matter! By shifting guests in even-numbered rooms into rooms with numbers half their respective room numbers, he transforms his half-vacant hotel into one that is completely full. In fact, if the manager wanted double occupancy in each room, he would have no need of additional guests at all. Just carry out the dividing procedure when there is one guest in every room of the hotel, then do it again, and finally have one of the guests in each odd-numbered room walk next door to the higher even-numbered room, and one winds up with two people in every room!

One might think that by means of these maneuvers the proprietor could always keep this strange hotel fully occupied. But one would be wrong. For suppose that the persons in rooms #4, 5, 6, . . . checked out. At a single stroke the hotel would be virtually emptied, the guest register reduced to three names, and the infinite converted to finitude. And yet it would remain true that as many guests checked out this time as when the guests in rooms #1, 3, 5, . . . checked out! Can anyone believe that such a hotel could exist in reality?

Hilbert's Hotel is absurd. But if an actual infinite were metaphysically possible, then such a hotel would be metaphysically possible. It follows that the real existence of an actual infinite is not metaphysically possible.

Partisans of the actual infinite might concede the absurdity of a Hilbert's Hotel but maintain that this case is somehow peculiar and, therefore, its metaphysical impossibility warrants no inference that an actual infinite is metaphysically impossible. This sort of response might seem appropriate with respect to certain absurdities involving actual infinites; for example, those imagining the completion of a so-called supertask, the sequential execution of an actually infinite number of definite and discrete operations in a finite time. But when it comes to situations involving the simultaneous existence of an actually infinite number of familiar macroscopic objects, then this sort of response seems less plausible.[10] If a (denumerably) actually infinite number of things could exist, they could be numbered and manipulated just like the guests in Hilbert's Hotel. Since nothing hangs on the illustration's involving a hotel, the metaphysical absurdity is plausibly attributed to the existence of an actual infinite. Thus, thought experiments of this sort show, in general, that it is impossible for an actually infinite number of things to exist in reality.

At this point, the actual infinitist has little choice but, in Oppy's words, simply to "embrace the conclusion of one's opponent's *reductio ad absurdum* argument" (Oppy 2006a, p. 48). Oppy explains, "these allegedly absurd situations are just what one ought to expect if there were . . . physical infinities" (Oppy 2006a, p. 48).

Oppy's response, however, falls short: it does nothing to prove that the envisioned situations are not absurd but only serves to reiterate, in effect, that if an actual infinite could exist in reality, then there could be a Hilbert's Hotel, which is not in dispute. The problem cases would, after all, not be problematic if the alleged consequences would not ensue! Rather the question is whether these consequences really are absurd.

Sobel similarly observes that such thought experiments bring into conflict two "seemingly innocuous" principles, namely,

(i) There are not more things in a multitude M than there are in a multitude M′ if there is a one-to-one correspondence of their members.

and

(ii) There are more things in M than there are in M′ if M′ is a proper submultitude of M.

We cannot have both of these principles along with

(iii) An infinite multitude exists.

---

[10] Oppy, for example, makes the point that having a hotel with an infinite number of occupied rooms does not commit one to the possibility of accommodating more guests by shifting guests about – maybe the hotel's construction hinders the guests' movements or the guests die off before their turn to move comes round. But as a *Gedankenexperiment* Hilbert's Hotel can be configured as we please without regard to mere physical possibilities.

For Sobel, the choice to be taken is clear: "The choice we have taken from Cantor is to hold on to (i) while restricting the proper submultiplicity condition to finite multiplicities. In this way we can 'have' comparable infinite multitudes" (Sobel 2004, pp. 186–7; cf. Mackie 1982, p. 93).

But the choice taken from Cantor of which Sobel speaks is a choice on the part of the mathematical community to reject intuitionism and finitism in favor of axiomatic infinite set theory. Finitism would too radically truncate mathematics to be acceptable to most mathematicians. But, as already indicated, that choice does not validate metaphysical conclusions. The metaphysician wants to know why, in order to resolve the inconsistency among (i)–(iii), it is (ii) that should be jettisoned (or restricted). Why not instead reject or restrict to finite multiplicities (i), which is a mere set-theoretical convention? More to the point, why not reject (iii) instead of the apparently innocuous (i) or (ii)? It certainly lacks the innocuousness of those principles, and giving it up would enable us to affirm both (i) and (ii). Remember: we can "have" comparable infinite multiplicities in mathematics without admitting them into our ontology.

Sobel thus needs some *argument* for the falsity of (ii). Again, it is insufficient merely to point out that if (i) and (iii) are true, then (ii) is false, for that is merely to reiterate that if an actual infinite were to exist, then the relevant situations would result, which is not in dispute.

Take Hilbert's Hotel. Sobel says that the difficulties with such a hotel are practical and physical; "they bring out the physical impossibility of this particular infinity of concurrent real things, not its logical impossibility" (Sobel 2004, p. 187). But the claim is not that such a hotel is logically impossible but metaphysically impossible. As an illustrative embodiment of transfinite arithmetic based on the axiomatic set theory, Hilbert's Hotel will, of necessity, be as logically consistent as that system; otherwise it would be useless as an illustration. But it also vividly illustrates the absurd situations to which the real existence of an infinite multitude can lead. The absurdity is not merely practical and physical; it is ontologically absurd that a hotel exist which is completely full and yet can accommodate untold infinities of new guests just by moving people around.

Oppy is prepared, if need be, simply to bite the bullet: "There can, after all, be a hotel in which infinitely many new guests are accommodated, even though all the rooms are full, via the simple expedient of moving the guests in room N to room $2N$ (for all N)" (Oppy 2006a, p. 53). So asserting does nothing to alleviate one's doubts that such a hotel is absurd. And would Oppy say something similar about what would happen when an infinite number of guests depart?[11] In transfinite arithmetic, inverse operations of subtraction and division with infinite quantities are prohibited because they lead to contradictions; as Sobel says, "Of course, as operations and properties are extended from finite to transfinite cardinals, some arithmetic principles are left confined to the finite" (Sobel 2007). But in reality, one cannot stop people from checking out of a hotel if they so desire! In this case, one does wind up with logically impossible situations, such

---

[11] Oppy suggests using J. Conway's recently developed constructions called surreal numbers to define operations of subtraction and division of transfinite numbers (Oppy 2006b, p. 140), but he explicitly denies that such non-canonical theories can be applied "to real-world problems, if one wishes to treat one's models with full ontological seriousness" (Oppy 2006a, p. 272). Oppy does not show, nor does he think, that the results of operations on surreals would be any less counterintuitive when translated into the concrete realm.

as subtracting identical quantities from identical quantities and finding nonidentical differences.[12]

In response to the absurdities springing from performing inverse operations with infinite quantities, David Yandell has insisted that subtraction of infinite quantities does not yield contradictions. He writes,

> Subtracting the even positive integers from the set of positive integers leaves an infinite set, the odd positive integers. Subtracting all of the positive integers greater than 40 from the set of positive integers leaves a finite (forty-membered) set. Subtracting all of the positive integers from the set of positive integers leaves one with the null set. But none of these subtractions could possibly lead to any other conclusion than each leads to. This alleged contradictory feature of the infinite seems not to generate any actual contradictions. (Yandell 2003, p. 132)

It is, of course, true that every time one subtracts all the even numbers from all the natural numbers, one gets all the odd numbers, which are infinite in quantity. But that is not where the contradiction is alleged to lie. Rather the contradiction lies in the fact that one can subtract equal quantities from equal quantities and arrive at different answers. For example, if we subtract all the even numbers from all the natural numbers, we get an infinity of numbers, and if we subtract all the numbers greater than three from all the natural numbers, we get only four numbers. Yet in both cases we subtracted the *identical number* of numbers from the *identical number* of numbers and yet did not arrive at an identical result. In fact, one can subtract equal quantities from equal quantities and get any quantity between zero and infinity as the remainder. For this reason, subtraction and division of infinite quantities are simply prohibited in transfinite arithmetic – a mere stipulation which has no force in the nonmathematical realm.

Sometimes it is said that we can find concrete counterexamples to the claim that an actually infinite number of things cannot exist, so that Premise (2.11) must be false. For example, Walter Sinnott-Armstrong asserts that the continuity of space and time entails the existence of an actually infinite number of points and instants (Craig & Sinnott-Armstrong 2003, p. 43). This familiar objection gratuitously assumes that space and time are composed of real points and instants, which has never been proven. Mathematically, the objection can be met by distinguishing a potential infinite from an actual infinite. While one can continue indefinitely to divide conceptually any distance, the series of subintervals thereby generated is merely potentially infinite, in that infinity serves as a limit that one endlessly approaches but never reaches. This is the thoroughgoing Aristotelian position on the infinite: only the potential infinite exists. This position does not imply that minimal time atoms, or chronons, exist. Rather time, like space, is infinitely divisible in the sense that division can proceed indefinitely, but time is never actually infinitely divided, neither does one arrive at an instantaneous point. If one thinks of a geometrical line as logically

---

[12] It will not do, in order to avoid the contradiction, to assert that there is nothing in transfinite arithmetic that forbids using set difference to form sets. Indeed, the thought experiment assumes that we can do such a thing. Removing all the guests in the odd-numbered rooms always leaves an infinite number of guests remaining, and removing all the guests in rooms numbered greater than four always leaves three guests remaining. That does not change the fact that in such cases identical quantities minus identical quantities yields nonidentical quantities, a contradiction.

prior to any points which one may care to specify on it rather than as a construction built up out of points (itself a paradoxical notion[13]), then one's ability to specify certain points, like the halfway point along a certain distance, does not imply that such points actually exist independently of our specification of them. As Grünbaum emphasizes, it is not infinite divisibility as such which gives rise to Zeno's paradoxes; the paradoxes presuppose the postulation of an actual infinity of points *ab initio.* ". . . [A]ny attribution of (infinite) 'divisibility' to a Cantorian line must be based on the fact that *ab initio* that line and the intervals are already 'divided' into an actual dense infinity of point-elements of which the line (interval) is the aggregate. Accordingly, the Cantorian line can be said to be already actually *infinitely divided*" (Grünbaum 1973, p. 169). By contrast, if we think of the line as logically prior to any points designated on it, then it is not an ordered aggregate of points nor actually infinitely divided. Time as duration is then logically prior to the (potentially infinite) divisions we make of it. Specified instants are not temporal intervals but merely the boundary points of intervals, which are always nonzero in duration. If one simply assumes that any distance is *already* composed out of an actually infinite number of points, then one is begging the question. The objector is assuming what he is supposed to prove, namely that there is a clear counterexample to the claim that an actually infinite number of things cannot exist.

Some critics have charged that the Aristotelian position that only potential, but no actual, infinites exist in reality is incoherent because a potential infinite presupposes an actual infinite. For example, Rudy Rucker claims that there must be a "definite class of possibilities," which is actually infinite in order for the mathematical intuitionist to regard the natural number series as potentially infinite through repetition of certain mathematical operations (Rucker 1980, p. 66). Similarly, Richard Sorabji asserts that Aristotle's view of the potentially infinite divisibility of a line entails that there is an actually infinite number of positions at which the line could be divided (Sorabji 1983, pp. 210–3, 322–4).

---

[13] See Craig (1985). Consider, for example, the many variations on the Grim Reaper Paradox (Benardete 1964, pp. 259–61; Hawthorne 2000; Oppy 2006a, pp. 63–6, 81–3). There are denumerably infinitely many Grim Reapers (whom we may identify as gods, so as to forestall any kinematic objections). You are alive at 12:00 p.m. Grim Reaper 1 will strike you dead at 1:00 p.m. if you are still alive at that time. Grim Reaper 2 will strike you dead at 12:30 p.m. if you are still alive then. Grim Reaper 3 will strike you dead at 12:15 p.m., and so on. Such a situation seems clearly conceivable but leads to an impossibility: you cannot survive past 12:00 p.m. and yet you cannot be killed at any time past 12:00 p.m. Oppy's solution to a similar paradox concerning infinitely many deafening peals, viz. that there is no particular peal responsible for your deafness but that the collective effect of infinitely many peals is to bring about deafness (Oppy 2006a, p. 83), not only involves a most bizarre form of retro-causation (Benardete 1964, p. 259) but is also in any case inapplicable to the Grim Reaper version since once you are dead no further Grim Reaper will swing his scythe, so that collective action is out of the question. The most plausible way to avert such paradoxes is by denying that time and space are constructions out of an actually infinite number of points. (My thanks to Alexander Pruss for drawing my attention to this version of the paradox.)

Moreover, on an A-Theory of time, according to which temporal becoming is an objective feature of reality, treating time as composed of instants (degenerate temporal intervals of zero duration) seems to land one in Zeno's clutches since temporal becoming would require the actualization of consecutive instants, which is incoherent. For a good discussion, see Grünbaum (1950–1, pp. 143–86). Grünbaum succeeds in defending the continuity of time only at the expense of sacrificing temporal becoming, which his interlocutors James and Whitehead would not do. See further Craig (2000c).

If this line of argument were successful, it would, indeed, be a *tour de force* since it would show mathematical thought from Aristotle to Gauss to be not merely mistaken or incomplete but incoherent in this respect. But the objection is not successful. For the claim that a physical distance is, say, potentially infinitely divisible does not entail that the distance is potentially divisible *here* and *here* and *here* and. . . . Potential infinite divisibility (the property of being susceptible of division without end) does not entail actual infinite divisibility (the property of being composed of an infinite number of points where divisions can be made). The argument that it does is guilty of a modal operator shift, inferring from the true claim

(1) Possibly, there is some point at which x is divided

to the disputed claim

(2) There is some point at which x is possibly divided.

But it is coherent to deny the validity of such an inference. Hence, one can maintain that a physical distance is potentially infinitely divisible without holding that there is an infinite number of positions where it could be divided.

Rucker also argues that there are probably, in fact, physical infinities (Rucker 1980, p. 69). If the *mutakallim* says, for example, that time is potentially infinite, then Rucker will reply that the modern, scientific worldview sees the past, present, and future as merely different regions coexisting in space-time. If he says that any physical infinity exists only as a temporal (potentially infinite) process, Rucker will rejoin that it is artificial to make physical existence a by-product of human activity. If there are, for example, an infinite number of bits of matter, this is a well-defined state of affairs which obtains right now regardless of our apprehension of it. Rucker concludes that it seems quite likely that there is some form of physical infinity.

Rucker's conclusion, however, clearly does not follow from his arguments. Time and space may well be finite. But could they be potentially infinite? Concerning time, even if Rucker were correct that a tenseless four-dimensionalism is correct, that would provide no reason at all to think the space-time manifold to be temporally infinite: there could well be finitely separated initial and final singularities. In any case, Rucker is simply incorrect in saying that "the modern, scientific worldview" precludes a theory of time, according to which temporal becoming is a real and objective feature of reality. Following McTaggart, contemporary philosophers of space and time distinguish between the so-called A-Theory of time, according to which events are temporally ordered by tensed determinations of past, present, and future, and temporal becoming is an objective feature of physical reality, and the so-called B-Theory of time, according to which events are ordered by the tenseless relations of *earlier than*, *simultaneous with*, and *later than*, and temporal becoming is purely subjective. Although some thinkers have carelessly asserted that relativity theory has vindicated the B-Theory over against its rival, such claims are untenable. One could harmonize the A-Theory and relativity theory in at least three different ways: (1) distinguish metaphysical time from physical or clock time and maintain that while the former is A-Theoretic in nature, the latter is a bare abstraction therefrom, useful for scientific purposes and quite possibly B-Theoretic in character, the element of becoming having been abstracted out; (2) relativize becoming to reference frames, just as is done with simultaneity; and (3) select a privileged reference frame to define the time in which objective becoming occurs, most plausibly the cosmic time, which serves as the time parameter for hypersurfaces of homogeneity in space-time in the General Theory of Relativity. And concerning space, to say that space is potentially infinite is not to say, with certain constructivists, that it depends on human activity (nor again, that there are actual places to which it can extend), but simply that space expands limitlessly as the distances between galaxies increase with time. As for the number of bits of matter, there is no incoherence in saying that there is a finite number of bits or that matter is capable of only a finite number of physical subdivisions, although mathematically one could proceed to carve up matter potentially *ad infinitum*. The sober fact is that there is just no evidence that actual infinities are anywhere instantiated in the physical world. It is therefore futile to seek to rebut (2.11) by appealing to clear counterexamples drawn from physical science.

2.12. An infinite regress of events as an actual infinite

The second premise states that *an infinite temporal regress of events is an actual infinite*. The point seems obvious enough, for if there has been a sequence composed of an infinite number of events stretching back into the past, then the set of all events in the series would be an actually infinite set.

But manifest as this may be to us, it was not always considered so. The point somehow eluded Aristotle himself, as well as his scholastic progeny, who regarded the series of past events as a potential infinite. Aristotle contended that since things in time come to exist sequentially, an actual infinite never exists at any one moment; only the present thing actually exists (*Physics* 3.6.206^{a}25–206^{b}1). Similarly, Aquinas, after confessing the impossibility of the existence of an actual infinite, nevertheless proceeded to assert that the existence of an infinite regress of past events is possible (*Summa Theologiae* 1.a.7.4.). This is because the series of past events does not exist in actuality. Past events do not now exist, and hence do not constitute an infinite number of actually existing things. The series is only potentially infinite, not actually infinite, in that it is constantly increasing by the addition of new events.

These Aristotelian thinkers are clearly presupposing an A-Theory of time and an ontology of presentism, according to which the only temporal items which exist are those that presently exist. On a B-Theory of tenseless time, since there is ontological parity among all events, there can be no question that an infinite temporal regress of events is composed of an actually infinite number of events.[14] Since all events are equally real, the fact that they exist (tenselessly) at different times loses any significance. The question, then, is whether events' temporal distribution over the past on a presentist ontology precludes our saying that the number of events in a beginningless series of events is actually infinite.

Now we may take it as a datum that the presentist can accurately count things that have existed but no longer exist. He knows, for example, how many US presidents there have been up through the present incumbent, what day of the month it is, how many shots Oswald squeezed off, and so forth. He knows how old his children are and can reckon how many billion years have elapsed since the Big Bang, if there was such an event. The

---

[14] Some philosophers of time, such as C. D. Broad and Michael Tooley, have defended a sort of hybrid A/B-Theory, according to which the past and present are on an ontological par, the past being a growing space-time block. On such a view, a beginningless series of past events is also, uncontroversially, actually infinite.

nonexistence of such things or events is no hindrance to their being enumerated. Indeed, any obstacle here is merely epistemic, for aside from considerations of vagueness there must be a certain number of such things. So in a beginningless series of past events of equal duration, the number of past events must be infinite, for it is larger than any natural number. But then the number of past events must be $\aleph_0$, for ∞ is not a number but an ideal limit. Aquinas' own example of a blacksmith working from eternity who uses one hammer after another as each one breaks furnishes a good example of an actual infinite, for the collection of all the hammers employed by the smith is an actual infinite. The fact that the broken hammers still exist is incidental to the story; even if they had all been destroyed after being broken, the number of hammers broken by the smith is the same. Similarly, if we consider all the events in an infinite temporal regress of events, they constitute an actual infinite.

The question arises whether on the A-Theory the series of future events, if time will go on forever, is not also actually infinite. Intuitively, it seems clear that the situation is not symmetrical, but this is notoriously difficult to express. It might rightly be pointed out that on presentism there are no future events and so no series of future events. Therefore, the number of future events is simply zero, not $\aleph_0$. (By this statement, one means not that there are future events, and that their number is 0, but that there just are no future events.) But on presentism, the past is as unreal as the future and, therefore, the number of past events could, with equal justification, be said to be zero. It might be said that at least there *have been* past events, and so they can be numbered. But by the same token there *will be* future events, so why can they not be numbered? Accordingly, one might be tempted to say that in an endless future there *will be* an actually infinite number of events, just as in a beginningless past there *have been* an actually infinite number of events. But in a sense that assertion is false; for there never will be an actually infinite number of events since it is impossible to count to infinity. The only sense in which there will be an infinite number of events is that the series of events will go toward infinity as a limit. But that is the concept of a potential infinite, not an actual infinite. Here the objectivity of temporal becoming makes itself felt. For as a result of the arrow of time, the series of events later than any arbitrarily selected past event is properly to be regarded as potentially infinite, that is to say, finite but indefinitely increasing toward infinity as a limit. The situation, significantly, is not symmetrical: as we have seen, the series of events earlier than any arbitrarily selected future event cannot properly be regarded as potentially infinite. So when we say that the number of past events is infinite, we mean that prior to today, $\aleph_0$ events have elapsed. But when we say that the number of future events is infinite, we do not mean that $\aleph_0$ events will elapse, for that is false. Ironically, then, it turns out that the series of future events cannot be actually infinite regardless of the infinity of the past or the metaphysical possibility of an actual infinite, for it is the objectivity of temporal becoming that makes the future potentially infinite only.

Because the series of past events is an actual infinite, all the absurdities attending the existence of an actual infinite apply to it. For example, if the series of past events is actually infinite, then the number of events that have occurred up to the present is no greater than the number that have occurred *at any point in the past.* Or again, if we number the events beginning in the present, then there have occurred as many odd-numbered events as events. If we mentally take away all the odd-numbered events, there are still an infinite number of events left over; but if we take away all the events greater than three, there are only four events left, even though in both cases we took away the same number of events.

2.13. Conclusion

Since an actual infinite cannot exist and an infinite temporal regress of events is an actual infinite, we may conclude that an infinite temporal regress of events cannot exist. Therefore, since the temporal regress of events is finite, the universe began to exist.

2.2. Argument from the impossibility of the formation of an actual infinite by successive addition

We now turn to a second philosophical argument in support of the premise that the universe began to exist, the argument from the impossibility of the formation of an actual infinite by successive addition. The argument may be simply formulated as follows:

2.21 A collection formed by successive addition cannot be an actual infinite.
2.22 The temporal series of events is a collection formed by successive addition.
2.23 Therefore, the temporal series of events cannot be an actual infinite.

This second argument is independent of the foregoing argument, for its conclusion is not incompatible with the existence of an actual infinite. It rather denies that a collection containing an actually infinite number of things can be *formed* by adding one member after another. If an actual infinite cannot be formed by successive addition, then the series of past events must be finite since that series is formed by successive addition of one event after another in time.

2.21. Formation of an actual infinite

Quite independent of the absurdities arising from the existence of an actually infinite number of things are the further difficulties arising as a result of the temporal formation of such a multitude through a process of successive addition. By "successive addition," one means the accrual of one new element at a (later) time. The temporality of the process of accrual is critical here. For while it is true that $1 + 1 + 1 + \ldots$ equals $\aleph_0$, the operation of addition signified by "+" is not applied successively but simultaneously or, better, timelessly. One does not add the *addenda* in temporal succession: $1 + 1 = 2$, then $2 + 1 = 3$, then $3 + 1 = 4, \ldots,$ but rather all together. By contrast, we are concerned here with a temporal process of successive addition of one element after another.

The impossibility of the formation of an actual infinite by successive addition seems obvious in the case of beginning at some point and trying to reach infinity.[15] For given any finite number n, $n + 1$ equals a finite number. Hence, $\aleph_0$ has no immediate predecessor; it is not the terminus of the natural number series but stands, as it were, outside it and is the

---

[15] This despite the speculation concerning the possibility of supertasks, various thought experiments involving the completion of an infinite number of tasks in a finite time by performing each successive task during half the time taken to perform its immediate predecessor. The fatal flaw in all such scenarios is that the state at $\omega + 1$ is causally unconnected to the successive states in the ω series of states. Since there is no last term in the ω series, the state of reality at $\omega + 1$ appears mysteriously from nowhere. The absurdity of such supertasks underlines the metaphysical impossibility of trying to convert a potential into an actual infinite.

number of all the members in the series. Notice that the impossibility of forming an actual infinite by successive addition has nothing to do with the amount of time available. Sometimes it is wrongly alleged that the only reason an actual infinite cannot be formed by successive addition is because there is not enough time.[16] But this is mistaken. While we can imagine an actually infinite series of events mapped onto a tenselessly existing infinite series of temporal intervals, such that each consecutive event is correlated with a unique consecutive interval, the question remains whether such a sequence of intervals can be instantiated, not tenselessly, but one interval after another. The very nature of the actual infinite precludes this. For regardless of the time available, a potential infinite cannot be turned into an actual infinite by any amount of successive addition since the result of every addition will always be finite. One sometimes, therefore, speaks of the impossibility of counting to infinity, for no matter how many numbers one counts, one can always count one more number before arriving at infinity. One sometimes speaks instead of the impossibility of traversing the infinite. The difficulty is the same: no matter how many steps one takes, the addition of one more step will not bring one to a point infinitely distant from one's starting point.

The question then arises whether, as a result of time's asymmetry, an actually infinite collection, although incapable of being formed by successive addition by beginning at a point and adding members, nevertheless could be formed by successive addition by never beginning but ending at a point, that is to say, ending at a point after having added one member after another from eternity. In this case, one is not engaged in the impossible task of trying to convert a potential into an actual infinite by successive addition. Rather at every point the series already is actually infinite, although allegedly successively formed.

Although the problems will be different, the formation of an actually infinite collection by never beginning and ending at some point seems scarcely less difficult than the formation of such a collection by beginning at some point and never ending. If one cannot count *to* infinity, how can one count down *from* infinity? If one cannot traverse the infinite by moving in one direction, how can one traverse it by moving in the opposite direction? In order for us to have "arrived" at today, temporal existence has, so to speak, traversed an infinite number of prior events.[17] But before the present event could occur, the event immediately prior to it would have to occur; and before that event could occur, the event immediately prior to it would have to occur; and so on *ad infinitum*. One gets driven back and back into the infinite past, making it impossible for any event to occur. Thus, if the series of past events were beginningless, the present event could not have occurred, which is absurd.

---

[16] For example, Oppy's discussion of counting forward to infinity is predicated upon Dretske's assumption that if one never stops counting, then one does count to infinity (Oppy 2006a, p. 61; cf. Dretske 1965). Oppy fails so much as to mention, much less take account, of the difference between an actual and a potential infinite in this case. One who, having begun, never stops counting counts "to infinity" only in the sense that one counts potentially infinitely.

[17] Richard Gale protests, "This argument depends on an anthropomorphic sense of 'going through' a set. The universe does not go through a set of events in the sense of planning which to go through first, in order to get through the second, and so on" (Gale 2007, pp. 92–3). Of course not; but on an A-Theory of time, the universe does endure through successive intervals of time. It arrives at its present event-state only by enduring through a series of prior event-states. Gale's framing the argument in terms of a "set of events" is maladroit since we are not talking about a set but about a series of events which elapse one after another.

It is unavailing to say that an infinite series of past events cannot be formed only in a finite time but that such formation is possible given infinite time, for that riposte only pushes the question back a notch: how can an actually infinite series of congruent temporal intervals successively elapse? Tenseless correlations are irrelevant here. Granted that the series of past events, if infinite, can be mapped one-to-one onto an equally infinite series of past temporal intervals, the question remains how such a temporal series can be lived through so as to arrive at the present.

The arguments against the formation of an actual infinite by successive addition bear a clear resemblance to Zeno's celebrated paradoxes of motion, in particular the Stadium and Dichotomy paradoxes, the Stadium in the case of beginning at some point and never ending and the Dichotomy in the case of never beginning and ending at some point. In the Dichotomy Paradox, Zeno argued that before Achilles could cross the stadium, he would have to cross halfway; but before he could cross halfway, he would have to cross a quarter of the way; but before he could cross a quarter of the way, he would have to cross an eighth of the way, and so on to infinity. It is evident that Achilles could not arrive at any point. In the case of the infinite past, we cannot speak meaningfully of halfway through the past or a quarter of the way, and so on since there is no beginning point, as there is in Achilles' case. But the metrical distances traversed are not essential to the conundrum insofar as the series of past events is concerned since the essential point holds that before traversing any interval there will always be a prior interval to be traversed first.

Now although Zeno's paradoxes have proved very stubborn, scarely anybody has really believed that motion is impossible. Is the argument against the impossibility of traversing an infinite past, as some critics allege, no more plausible than Zeno's paradoxes? This cannot be said because the allgation fails to reckon with two crucial disanalogies of the case of an infinite past to Zeno's paradoxes: whereas in Zeno's thought experimentsm the intervals traversed are *potential* and *unequal*, in the case of an infinite past the intervals are *actual* and *equal*. The claim that Achilles must pass through an infinite number of halfway points in order to cross the stadium already assumes that the whole interval is a composition of an infinite number of points, whereas Zeno's opponents, like Aristotle, take the line as a whole to be conceptually prior to any divisions which we might make in it. Moreover, Zeno's intervals, being unequal, sum to a merely finite distance, whereas the intervals in an infinite past sum to an infinite distance. The question is not whether it is possible to traverse infinitely many (progressively shorter) distances but whether it is possible to traverse an infinite distance. Thus, the problem of traversing an infinite distance comprising an infinite number of equal, actual intervals to arrive at our present location cannot be dismissed on the basis of the argument's resemblance in certain respects to Zeno's puzzles.

It is surprising that a number of critics, such as Mackie and Sobel, have objected that the argument illicitly presupposes an infinitely distant starting point in the past and then pronounces it impossible to travel from that point to today. But if the past is infinite, they say, then there would be no starting point whatever, not even an infinitely distant one. Nevertheless, from any given point in the past, there is only a finite distance to the present, which is easily "traversed" (Mackie 1982, p. 93; Sobel 2004, p. 182). But, in fact, no proponent of the *kalam* argument of whom we are aware has assumed that there was an infinitely distant starting point in the past. The fact that there is *no beginning* at all, not even an infinitely distant one, seems only to make the problem worse, not better. To say that the infinite past could have been formed by successive addition is like saying that someone has just succeeded in writing down all the negative numbers, ending at −1. And how is the claim that from any given moment in the past there is only a finite distance to the present even relevant to the issue? For the question is how the *whole* series can be formed, not a finite portion of it. Do Mackie and Sobel think that because every *finite* segment of the series can be formed by successive addition the whole *infinite* series can be so formed? That is as logically fallacious as saying that because every part of an elephant is light in weight, the whole elephant is light in weight, or in other words, to commit the fallacy of composition. The claim that from any given moment in the past there is only a finite distance to the present is simply irrelevant.

Wholly apart from these Zenonian arguments, the notion that the series of past events could be actually infinite is notoriously difficult. Consider, for example, al-Ghazali's thought experiment involving two beginningless series of coordinated events. He envisions our solar system's existing from eternity past, the orbital periods of the planets being so coordinated that for every one orbit which Saturn completes Jupiter completes 2.5 times as many. If they have been orbiting from eternity, which planet has completed the most orbits? The correct mathematical answer is that they have completed precisely the same number of orbits. But this seems absurd, for the longer they revolve, the greater becomes the disparity between them, so that they progressively approach a limit at which Jupiter has fallen infinitely far behind Saturn. Yet, being now actually infinite, their respective completed orbits are somehow magically identical. Indeed, they will have "attained" infinity from eternity past: the number of completed orbits is always the same. Moreover, Ghazali asks, will the number of completed orbits be even or odd? Either answer seems absurd. We might be tempted to deny that the number of completed orbits is either even or odd. But post-Cantorian transfinite arithmetic gives a quite different answer: the number of orbits completed is both even and odd! For a cardinal number n is even if there is a unique cardinal number m such that $n = 2m$, and n is odd if there is a unique cardinal number m such that $n = 2m + 1$. In the envisioned scenario, the number of completed orbits is (in both cases!) $\aleph_0$, and $\aleph_0 = 2\aleph_0 = 2\aleph_0 + 1$. So Jupiter and Saturn have each completed both an even and an odd number of orbits, and that number has remained equal and unchanged from all eternity, despite their ongoing revolutions and the growing disparity between them over any finite interval of time. This seems absurd.[18]

Or consider the case of Tristram Shandy, who, in the novel by Sterne, writes his autobiography so slowly that it takes him a whole year to record the events of a single day. Tristram Shandy laments that at this rate he can never come to an end.

According to Russell, if Tristram Shandy were immortal and did not weary of his task, "no part of his biography would have remained unwritten," since by Hume's Principle to each day there would correspond 1 year, and both are infinite (Russell, 1937, p. 358). Such an assertion is misleading, however. The fact that every part of the autobiography will be eventually written does not imply that the whole autobiography will be eventually written, which was, after all, Tristram Shandy's concern. For every part of the autobiography there

---

[18] Oppy's discussion of al-Ghazali's problem just fails to connect with the problem as we understand it (Oppy 2006a, pp. 49–51), probably because Oppy takes its point to be that there is a logical contradiction with respect to the number of orbits completed (Oppy 2006a, p. 8), so that he spends most of his space arguing that given Cantorian assumptions there is no unequivocal sense in which the number of orbits both is and is not same. Temporal becoming is left wholly out of account.

is some time at which it will be completed, but there is not some time at which every part of the autobiography will be completed. Given an A-Theory of time, though he write forever, Tristram Shandy would only get farther and farther behind, so that instead of finishing his autobiography, he would progressively approach a state in which he would be infinitely far behind.

But now turn the story about: suppose Tristram Shandy has been writing from eternity past at the rate of 1 day per year. Should not Tristram Shandy now be infinitely far behind? For if he has lived for an infinite number of years, Tristram Shandy has recorded an equally infinite number of past days. Given the thoroughness of his autobiography, these days are all consecutive days. At any point in the past or present, therefore, Tristram Shandy has recorded a beginningless, infinite series of consecutive days. But now the question arises: Which days are these? Where in the temporal series of events are the days recorded by Tristram Shandy at any given point? The answer can only be that *they are days infinitely distant from the present*. For there is no day on which Tristram Shandy is writing which is finitely distant from the last recorded day.

This may be seen through an incisive analysis of the Tristram Shandy Paradox given by Robin Small (1986, pp. 214–5). He points out that if Tristram Shandy has been writing for 1 year's time, then the most recent day he could have recorded is 1 year ago. But if he has been writing for 2 years, then that same day could not have been recorded by him. For since his intention is to record consecutive days of his life, the most recent day he could have recorded is the day immediately after a day at least 2 years ago. This is because it takes a year to record a day, so that to record 2 days he must have 2 years. Similarly, if he has been writing 3 years, then the most recent day recorded could be no more recent than 3 years and 2 days ago. In other words, the longer he has written the further behind he has fallen. In fact, the recession into the past of the most recent recordable day can be plotted according to the formula (present date – n years of writing) + n – 1 days. But what happens if Tristram Shandy has, *ex hypothesi*, been writing for an infinite number of years? The most recent day of his autobiography recedes to infinity, that is to say, to a day infinitely distant from the present. Nowhere in the past at a finite distance from the present can we find a recorded day, for by now Tristram Shandy is infinitely far behind. The beginningless, infinite series of days which he has recorded are days which lie at an infinite temporal distance from the present. This is not in itself a contradiction. The infinite past must have in this case, not the order type of the negative numbers ω^*, but the order type $\omega^* + \omega^*$, the order type of the series . . . , –3, –2, –1, . . . , –3, –2, –1. But there is no way to traverse the temporal interval from an infinitely distant event to the present, or, more technically, for an event which was once present to recede to an infinite temporal distance. Since the task of writing one's autobiography at the rate of 1 year per day seems obviously coherent, what follows from the Tristram Shandy story is that an infinite series of past events is absurd.[19]

But suppose that such an infinite task could be completed by the present day. Suppose we meet a man who claims to have been counting down from infinity and who is now finishing: . . . , –3, –2, –1, 0. We could ask, why did he not finish counting yesterday or the

---

[19] Oppy rightly observes that it is the whole scenario that is impossible, which includes the requirement that consecutive days be recorded (Oppy 2006a, p. 57, n. 3). But given that the task of writing one's autobiography at the rate of 1 day per year seems obviously coherent, it seem to us that the blame can be placed on the infinity of the past.

day before or the year before? By then an infinite time had already elapsed, so that he has had ample time to finish. Thus, at no point in the infinite past should we ever find the man finishing his countdown, for by that point he should already be done! In fact, no matter how far back into the past we go, we can never find the man counting at all, for at any point we reach he will already have finished. But if at no point in the past do we find him counting, this contradicts the hypothesis that he has been counting from eternity. This shows again that the formation of an actual infinite by never beginning but reaching an end is as impossible as beginning at a point and trying to reach infinity.

Conway and Sorabji have responded that there is no reason to think that the man would at any point have already finished (Sorabji 1983, pp. 219–22; Conway 1984). Sorabji thinks the argument confuses counting an *infinity* of numbers with counting *all* the numbers. At any given point in the past, the man will have already counted an infinity of negative numbers, but that does not entail that he will have counted all the negative numbers. Similarly, in Conway's analysis, the nub of the argument lies in the conditional

(*) If an infinite number of numbers had been counted by yesterday, then the man will have finished by yesterday.

But Conway's conditional is quite ambiguous, and the arguments that he suggests in support of it have no apparent relevance to the reasoning behind the paradox. The *mutakallim* is not making the obviously false claim that to count infinitely many negative numbers is to count all the negative numbers! Rather, the conditional at the heart of the paradox is a counterfactual conditional like:

(**) If the man would have finished his countdown by today, then he would have finished it by yesterday,

and the truth of this conditional seems plausible in light of Hume's Principle. It is on the basis of this principle that the defender of the infinite past seeks to justify the intuitively impossible feat of someone's counting down all the negative numbers and ending at 0. Since the negative numbers can be put into a one-to-one correspondence with the series of, say, past hours, someone counting from eternity would have completed his countdown. But by the same token, the man at any point in the past should have already completed his countdown, since by then a one-to-one correspondence exists between each negative number and a past hour. In this case, having infinite time does seem to be a sufficient condition of finishing the job. Having had infinite time, the man should have already completed his task.

Such reasoning in support of the finitude of the past and the beginning of the universe is not mere armchair cosmology. P. C. W. Davies, for example, utilizes this reasoning in explaining two profound implications of the thermodynamic properties of the universe:

> The first is that the universe will eventually die, wallowing, as it were, in its own entropy. This is known among physicists as the 'heat death' of the universe. The second is that the universe cannot have existed for ever, otherwise it would have reached its equilibrium end state an infinite time ago. Conclusion: the universe did not always exist. (Davies 1983, p. 11)

The second of these implications is a clear application of the reasoning that underlies the current paradox: even if the universe had infinite energy, it would in infinite time come to an equilibrium since at any point in the past infinite time has elapsed, a beginningless universe would have already reached an equilibrium, or as Davies puts it, it would have reached an equilibrium an infinite time ago. Therefore, the universe began to exist, *quod erat demonstrandum*.[20]

Oppy's response to the problem at hand is to say that the man's finishing his countdown when he does rather than earlier is just "a brute feature of the scenario, that is, a feature that has no explanation" (Oppy 2006a, p. 59; cf. p. 63; Oppy 2006b, pp. 141–2). It has always been the case that he will finish when he does, but why the man finishes when he does rather than at some other time is just inexplicable. Resting with inexplicability may seem unsatisfactory, however, especially in light of the respectable role such reasoning plays in scientific cosmological discussions. Oppy justifies his response on the basis that principles of sufficient reason requiring that there be an explanation in such a case are highly contentious. Oppy presents the typical objections to various versions of the Principle of Sufficient Reason such as the impossibility of providing an explanation of what has been called the "Big Contingent Conjunctive Fact" (BCCF), which is the conjunction of all the contingent facts there are, or of libertarian free choices (Oppy 2006a, pp. 279–80). The problem with this justification, however, is twofold. First, plausible defenses of the Principle of Sufficient Reason can be given.[21] Second, and more to the point, there is no reason to think that requiring the need for an explanation in the present case demands for its acceptability or plausibility the enunciation and defense of some general Principle of Sufficient Reason. Indeed, any such principle is apt to be tested inductively for its adequacy by whether cases like this constitute plausible counterexamples. The exceptions offered by Oppy, such as the inexplicability of the BCCF and libertarian choices, are simply irrelevant to the present case, for the BCCF is not at stake nor can a person counting from eternity at a constant rate choose arbitrarily when to finish his countdown. In the case under discussion, we have a good reason to think that the man should have finished his countdown earlier than any

---

[20] See the similar reasoning of Barrow and Tipler (1986, p. 601–8) against inflationary steady-state cosmologies on the ground that any event which would have happened by now would have already happened before now if the past were infinite.

[21] See Pruss' article in this volume. We shall leave to him the defense of principles of sufficient reason. Oppy himself thinks that it is "very plausible" that there are acceptable instances of the following schema for a Principle of Sufficient Reason:

> **O** (for every **FG** of kind K, there is an **F′G′** that *partly* explains why the **GF**s rather than **Q** possible alternatives),

where **O** is an operator like "necessarily," "it is knowable *a priori*," etc., **G** is an ontological category such as a proposition, state of affairs, etc., **F** is a restriction such as true, contingent, etc., and **Q** is a quantifier like "any," "every," etc. (Oppy 2006a, p. 285, cf. pp. 275–6). But he thinks that it is not at all clear that there are acceptable instances of this schema that can be used to rule out scenarios like counting down from infinity. Although it is not clear what Oppy means by "**GF**s," the following principle would seem to be an instance of his schema: Necessarily, for any contingent state of affairs involving concrete objects there is a contingent state of affairs that partly explains why that state of affairs obtains rather than any other. Such a principle would require that there be some partial explanation for why the man finishes his countdown today rather than at some other time. But not even a partial explanation can be given, for regardless of how we vary such factors as the rate of counting, they will be the same regardless of the time that he finishes and so do not furnish even a partial explanation of why he finishes today. So why is this instance of the schema not acceptable?

time that he does, namely, he has already had infinite time to get the job done.[22] If we deny that infinite time is sufficient for completing the task, then we shall wonder why he is finishing today rather than tomorrow or the day after tomorrow, or, indeed, at any time in the potentially infinite future. It is not unreasonable to demand some sort of explanation for why, if he finishes today, he did not already finish yesterday. By contrast, if such a countdown is metaphysically impossible, then no such conundrum can arise. But clearly, there is no metaphysical impossibility in counting backward for all time, unless time is past eternal. It follows that the past cannot be infinite.

For all of these reasons, the formation of an actual infinite by successive addition is a notoriously difficult notion, even more so than the static existence of an actual infinite.

2.22. Successive formation of the series of past events

Premise (2.22) may seem rather obvious. The past did not spring into being whole and entire but was formed sequentially, one event occurring after another. Notice, too, that the direction of this formation is "forward," in the sense that the collection grows with time. Although we sometimes speak of an "infinite regress" of events, in reality an infinite past would be an "infinite progress" of events with no beginning and its end in the present.

As obvious as this premise may seem at first blush, it is, in fact, a matter of great controversy. It presupposes once again an A-Theory of time. On such a theory, the collection of all past events prior to any given event is not a collection whose members all tenselessly coexist. Rather it is a collection that is instantiated sequentially or successively in time, one event coming to pass on the heels of another. Since temporal becoming is an objective feature of the physical world, the series of past events is not a tenselessly existing manifold, all of whose members are equally real. Rather the members of the series come to be and pass away one after another.

Space does not permit a review of the arguments for and against the A- and B-Theories of time respectively. But on the basis of a case such as is presented by Craig (2000a,b), we take ourselves to be justified in affirming the objective reality of temporal becoming and, hence, the formation of the series of temporal events by successive addition. It is noteworthy that contemporary opponents of Zenonian arguments such as Grünbaum resolve those puzzles only by denying the objective reality of temporal becoming and treating time as a continuum of tenselessly existing point-instants. If moments of time and, hence, events really do come to be and elapse, then it remains mysterious how an infinite number of such event-intervals can be traversed or manage successively to elapse.

2.23. Conclusion

It follows, then, that the temporal series of events cannot be actually infinite. The only way a collection to which members are being successively added could be actually infinite would

---

[22] Notice, too, that if there is *any* probability of his finishing in infinite time, then he will have already finished.

be for it to have an infinite tenselessly existing "core" to which additions are being made. But then, it would not be a collection *formed* by successive addition, for there would always exist a surd infinite, itself not formed successively but simply given, to which a finite number of successive additions have been made. Clearly, the temporal series of events cannot be so characterized, for it is by nature successively formed throughout. Thus, prior to any arbitrarily designated point in the temporal series, one has a collection of past events up to that point which is successively formed and completed and cannot, therefore, be actually infinite.

2.3. Scientific confirmation

The sort of philosophical problems with the infinity of the past, which have been the object of our discussion, are now being recognized in scientific papers by leading cosmologists and philosophers of science.[23] For example, Ellis, Kirchner, and Stoeger ask, "Can there be an infinite set of really existing universes? We suggest that, on the basis of well-known *philosophical* arguments, the answer is No" (Ellis, Kirchner, & Stoeger 2003, p. 14; emphasis added). Similarly, noting that an actual infinite is not constructible and, therefore, not actualizable, they assert, "This is precisely why a realized past infinity in time is not considered possible from this standpoint – since it involves an infinite set of completed events or moments" (Ellis, Kirchner, & Stoeger 2003, p. 14). These misgivings represent endorsements of both the *kalam* arguments defended earlier. Ellis and his colleagues conclude, "The arguments against an infinite past time are strong – it's simply not constructible in terms of events or instants of time, besides being conceptually indefinite" (Ellis, Kirchner, & Stoeger 2003, p. 14).

Apart from these philosophical arguments, there has emerged during the course of the twentieth century provocative empirical evidence that the universe is not past eternal. This physical evidence for the beginning of the universe comes from what is undoubtedly one of the most exciting and rapidly developing fields of science today: astronomy and astrophysics. Prior to the 1920s, scientists had always assumed that the universe was stationary and eternal. Tremors of the impending earthquake that would topple this traditional cosmology were first felt in 1917, when Albert Einstein made a cosmological application of his newly discovered gravitational theory, the General Theory of Relativity (Einstein 1917, pp. 177–88). In so doing, he assumed that the universe is homogeneous and isotropic and that it exists in a steady state, with a constant mean mass density and a constant curvature of space. To his chagrin, however, he found that General Relativity (GR) would not permit such a model of the universe unless he introduced into his gravitational field equations a certain "fudge factor" Λ in order to counterbalance the gravitational effect of matter and so ensure a static universe. Einstein's universe was balanced on a razor's edge, however, and the least perturbation – even the transport of matter from one part of the universe to another – would cause the universe either to implode or to expand. By taking this feature of Einstein's model seriously, the Russian mathematician Alexander Friedmann and the Belgian astronomer Georges Lemaître were able to formulate independently in the 1920s solutions to the field equations which predicted an expanding universe (Friedmann 1922; Lemaître 1927).

---

[23] Besides the paper by Ellis et al., see Vaas (2004).

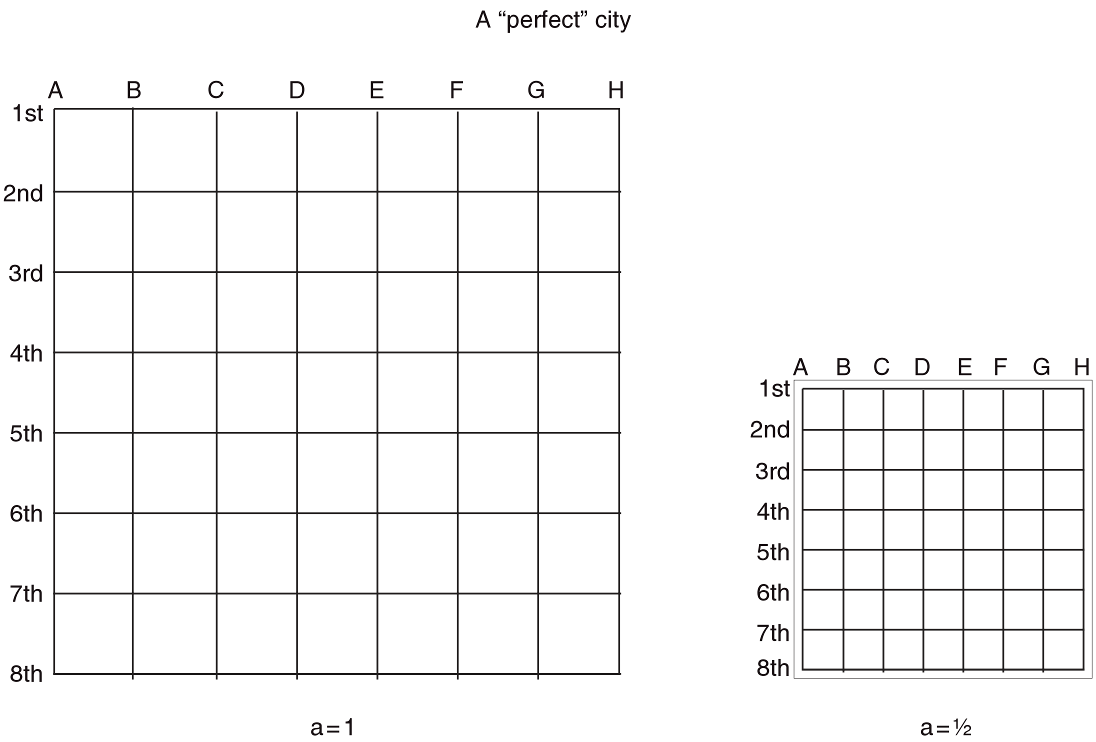


Figure 3.2 Analogy of the universe as a city laid out in a grid.

Friedmann's first equation is:

$$H^2 = \left(\frac{\mathring{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{k}{a^2},$$

where H = Hubble parameter, a = scale factor, G = gravitational constant, ρ = mass density of universe, Λ = cosmological constant, k = curvature parameter.[24] By way of explanation, the scale factor, "*a*," of the universe is a global multiplier to universe size. Imagine the universe as a perfectly laid out city with streets that travel only north-south and east-west. Streets are spaced at equal-distance intervals. Street intersections then define perfectly symmetric city blocks. One could go further and think of buildings in the city as analogous with galaxies in the universe.

The distance from one city block to another is a function of two values: the originally laid out distance (call that the "normalized" distance) and the scale factor multiplier "*a*." Note that, as in Figure 3.2, when one multiplies by a scale factor of 1/2, one still has precisely the same city with the same number of city blocks. The only thing that has changed is the distance interval between the city blocks.

Now consider buildings within the city. If the city block distance were reduced to the size of buildings, clearly something must give. The buildings would be squeezed together and destroyed. This is analogous to what happens with matter in the real universe. The sizes of nonelementary particle mass structures such as protons, neutrons, atomic nuclei, and so on are fixed; they do not change with the scale factor. Other physical structures, such as massless particles, *do* adjust with the scale factor. The wavelength of radiation

---

[24] This form of the equation assumes a unit system where the speed of light, "*c*," is equal to one. Otherwise the cosmological constant term and the curvature parameter term would be seen as multipled by c^2.

adjusts and, hence, would gain (in the case of contraction) or lose (in the case of expansion) energy as a result[25]. When one crowds these particles in upon themselves, one will see a transition to different physics.

Recall that the full city is always present regardless of the value of the scale factor. So now consider two additional situations. First, suppose that the scale factor were to shrink to zero. Space (and time) would disappear. Any structure that could not transform to zero size would be destroyed. If there were no physical process that would allow such a thing to happen, we should seem to have a paradox. Either there must be an undiscovered physical process or the scale factor cannot, in reality, assume a null value.

Second, imagine that the city is of infinite size. Conceptually, there is no problem with extending the streets north-south and east-west to infinity in each direction. What does it mean to scale the universe's size in such a situation? No matter what scale factor one adopts, the size of the universe remains infinite. Nevertheless, the idea of scaling still retains coherency in that we can apply a multiplier to the finite distance between city blocks. Yet what would be the meaning of applying a zero scale factor in this situation? Now it would appear that the size of the full universe is "zero times infinity," which, in general, can be any finite number (Barrow 2005, p. 160). What does this mean, given that the distance between *any* spot in the universe to *any* other spot must still be zero? GR simply breaks down at zero scale factor.

Whether or not the full universe is of infinite or finite size is given in the Friedmann equation by the curvature parameter "k." A positive k indicates that the universe, much like the surface of the Earth, is unbounded yet of finite size. Going back to the analogy, imagine that the city is laid out over Earth's entire curved surface. A traveler on 1st street would never come to the end; rather he would eventually come back to the location where he started. A positive k yields positive curvature and a closed universe. This is one type of "compact metric" within GR.

A zero value for "k" yields a "flat" universe. The 1st street is unbounded and of infinite length (in both directions). A similar situation obtains for a negative k value. Here one has negative, or "saddle-shaped," curvature. Two travelers moving east and side-by-side up 1st and 2nd streets would actually get laterally farther apart as the curvature of the surface causes the streets to diverge from each other. The latter case gives an infinitely sized "open" universe.

The components of the universe (all the energy, keeping in mind that $E = mc^2$) determine what type of curvature the universe possesses. The "strength" of gravity, included in the equation via the parameter "G," affects the magnitude of the curvature.

The parameters ρ and Λ indicate the type and magnitude of the different types of energy that cause the curvature. The parameter "ρ" represents the density (that is, the energy per unit volume) of the two types of "ordinary" energy: matter and radiation. It is "ordinary" in the sense that we are familiar with it in daily life and it is of a form that makes gravity an attractive force. Λ represents an exotic type of energy density which can transform gravity from an attractive to a repulsive force.

Friedmann's first equation tells us how the scale factor changes as time elapses. Mathematically, this is the first derivative of the scale factor "a," known as "a-dot," or $\dot{a}$. One can see that the increase (or decrease) in the scale factor is strongly a function of the universe's energy content. Now the "ordinary" energy density ρ will become smaller as the universe

---

[25] Hence, the temperature of the universe would be seen to rise as one looked back in time.

expands, since one has the same amount of energy spread out over a greater volume. So its causal impact on the expansion will progressively diminish at ever later times (this works in reverse for contraction). By contrast, Λ, which represents the dark energy density, is constant. The dark energy does not become more dilute during expansion or concentrated during contraction. Hence, early in the life of an expanding universe, Λ is unimportant compared to ρ. But its impact "snowballs" as time goes on. As long as the impact of ρ in the early universe is not enough to overturn an expansion and begin a contraction, the effect of Λ will eventually lead to a runaway expansion of the universe. There will appear a moment in the history of the universe when the dark energy will begin to dominate the ordinary energy, and the universe's expansion will begin to accelerate. Recent observations, in fact, seem to show precisely this effect in our own universe, with a transition age at 9 billion years (Overbye 2006).

Friedmann's second equation gives the rate of change of the expansion rate:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}$$

It determines whether the expansion itself is slowing or accelerating. This acceleration is referred to as "a-double-dot," or $\ddot{a}$. A new term "p" appears in the equation. This is the pressure (similar to the pressure of a gas inside a balloon). Pressure itself can produce gravitational force. Pressure is normally negligible for ordinary matter, although it can play a role in radiation dominated universes. Pressure, however, can have a tremendous impact given a universe dominated by dark energy. As Friedmann's second equation shows, the rate at which the expansion of the universe accelerates is proportional to: $(-\rho - 3p)$. But the pressure in the vacuum is just equal to the negative of the energy density; this is called the equation of state. Hence, overall, the acceleration is positive (which will produce expansion) and proportional to twice the energy density.

Ordinary matter will exert positive pressure (which will keep a balloon inflated, for example). This type of pressure will produce an attractive gravitational force, which supplements the attractive gravity that accrues from mass. Dark energy has the bizarre property that it generates *negative* pressure. But dark energy, while it has a positive energy density (which contributes to attractive gravity) will, *on net*, produce a repulsive gravitational effect. Looking at Friedmann's second equation, one sees that an attractive gravitational contribution tends to slow down expansion (or accelerate contraction), while repulsive gravity will do the opposite.

The monumental significance of the Friedmann–Lemaître model lay in its historization of the universe. As one commentator has remarked, up to this time the idea of the expansion of the universe "was absolutely beyond comprehension. Throughout all of human history the universe was regarded as fixed and immutable and the idea that it might actually be changing was inconceivable" (Naber 1988, pp. 126–7). But if the Friedmann–Lemaître model is correct, the universe can no longer be adequately treated as a static entity existing, in effect, timelessly. Rather the universe has a history, and time will not be a matter of indifference for our investigation of the cosmos.

In 1929, the American astronomer Edwin Hubble showed that the redshift in the optical spectra of light from distant galaxies was a common feature of all measured galaxies and was proportional to their distance from us (Hubble 1929, pp. 168–73). This redshift, first observed by Vesto Slipher at the Lowell Observatory,[26] was taken to be a Doppler effect indicative of the recessional motion of the light source in the line of sight. Incredibly, what Hubble had discovered was the isotropic expansion of the universe predicted by Friedmann and Lemaître on the basis of Einstein's GR. It was a veritable turning point in the history of science. "Of all the great predictions that science has ever made over the centuries," exclaims John Wheeler, "was there ever one greater than this, to predict, and predict correctly, and predict against all expectation a phenomenon so fantastic as the expansion of the universe?" (Wheeler 1980, p. 354).

2.31. The standard Hot Big Bang model

According to the Friedmann–Lemaître model, as time proceeds, the distances separating the ideal particles of the cosmological fluid constituted by the matter and energy of the universe become greater. It is important to appreciate that as a GR-based theory, the model does not describe the expansion of the material content of the universe into a preexisting, empty, Newtonian space, but rather the expansion of space itself. The ideal particles of the cosmological fluid are conceived to be at rest with respect to space but to recede progressively from one another as space itself expands or stretches, just as buttons glued to the surface of a balloon will recede from one another as the balloon inflates. As the universe expands, its density progressively declines.

This has the astonishing implication that as one reverses the expansion and extrapolates back in time, the universe becomes progressively denser until one arrives at a state of infinite density at some point in the finite past. This state represents a singularity at which space-time curvature, along with temperature, pressure, and density, becomes infinite. To be more correct, the volume of the universe *approaches* zero in the limit as the *scale factor* of the universe approaches zero. The Friedmann–Lemaître model does not, in fact, describe what happens at the singularity, since Einstein's GR breaks down at this limit.

The initial cosmological singularity is, therefore, not in space-time but constitutes an edge or boundary to space-time itself. Robert Wald describes how singular space-times are to be properly characterized:

> By far the most satisfactory idea proposed thus far is basically to use the 'holes' left behind by the removal of singularities as the criterion for their presence. These 'holes' should be detectable by the fact that there will be geodesics which have finite affine length; that is, more precisely there should exist geodesics which are inextendible in at least one direction but have only a finite range of affine parameter. Such geodesics are said to be *incomplete*. (For timelike and spacelike geodesics, finite affine 'length' is equivalent to finite proper time or length so the use of affine parameter simply generalizes the notion of 'finite length' to null geodesics.) Thus, we could define a spacetime to be singular if it possesses at least one incomplete geodesic.
>
> Nevertheless, there is a serious physical pathology in any spacetime which is timelike or null geodesically incomplete. In such a spacetime, it is possible for at least one freely falling particle or photon to end its existence within a finite 'time' (that is, affine parameter) or to have begun its existence a finite time ago. Thus, even if one does not have a completely satisfactory general notion of singularities, one would be justified in calling such spacetimes

---

[26] Slipher's early papers are now available online at http://www.roe.ac.uk/~jap/slipher/.

> physically singular. It is this property that is proven by the singularity theorems to hold in a wide class of spacetimes. (Wald 1984, pp. 215–6)[27]

The existence of a boundary to space-time implies, not merely that the "stuff" of the universe begins to exist but that space and time do as well (for in the Friedmann–Lemaître model all past-directed geodesics terminate at the singularity). P. C. W. Davies comments:

> If we extrapolate this prediction to its extreme, we reach a point when all distances in the universe have shrunk to zero. An initial cosmological singularity therefore forms a past temporal extremity to the universe. We cannot continue physical reasoning, or even the concept of spacetime, through such an extremity. For this reason most cosmologists think of the initial singularity as the beginning of the universe. On this view the big bang represents the creation event; the creation not only of all the matter and energy in the universe, but also of spacetime itself. (Davies 1978, pp. 78–9)

The term "Big Bang," originally a derisive expression coined by Fred Hoyle to characterize the beginning of the universe predicted by the Friedmann–Lemaître model, is thus potentially misleading, since the expansion cannot be visualized from the outside (there being no "outside," just as there is no "before" with respect to the Big Bang).[28]

The standard Hot Big Bang model, as the Friedmann–Lemaître model came to be called, thus describes a universe which is not eternal in the past, but which came into being a finite time ago. Moreover – and this deserves underscoring – the origin it posits is an absolute origin *ex nihilo*. For not only all matter and energy but also space and time themselves come into being at the initial cosmological singularity. As Barrow and Tipler emphasize, "At this singularity, space and time came into existence; literally nothing existed before the singularity, so, if the Universe originated at such a singularity, we would truly have a creation *ex nihilo*" (Barrow and Tipler 1986, p. 442). On such a model the universe originates *ex nihilo* in the sense that it is false that something existed prior to the singularity.

---

[27] A geodesic is the path that a freely falling particle traces out through space and time. A timelike geodesic is traveled by a massive particle. A null geodesic is traveled by a massless particle such as the photons that make up visible light.

[28] As Gott et al. write:

> The universe began from a state of infinite density about one Hubble time ago. Space and time were created in that event and so was all the matter in the universe. It is not meaningful to ask what happened before the big bang; it is somewhat like asking what is north of the North Pole. Similarly, it is not sensible to ask where the big bang took place. The point-universe was not an object isolated in space; it was the entire universe, and so the only answer can be that the big bang happened everywhere. (1976, p. 65)

The Hubble time is the time since the singularity if the rate of expansion has been constant. The singularity is a point only in the sense that the distance between any two points in the singularity is zero. Anyone who thinks that there must be a place in the universe where the Big Bang occurred still has not grasped that it is space itself which is expanding; it is the two-dimensional *surface* of an inflating balloon which is analogous to three-dimensional space. The spherical surface has no center and so no location where the expansion begins. The analogy of the North Pole with the beginning of time should not be pressed, since the North Pole is not an edge to the surface of the globe; the beginning of time is more like the apex of a cone. But the idea is that just as one cannot go further north than the North Pole, so one cannot go earlier than the initial singularity.

2.32. Evidence for GR

The earliest evidence in favor of the Big Bang came from the consonance of theory and experiment. Einstein's early papers proposed two tests that could be performed immediately. It had been known for some time that Newton's gravitational theory could not adequately describe the orbit of the planet Mercury. The real orbit precessed around the sun (i.e. the ellipse itself rotates over time). In contrast to Newton's theory of gravity, GR "predicted" that this precession should take place. Einstein's theory also predicted that, since matter bends space, light rays should have their paths noticeably bent when they pass close to massive objects. A solar eclipse in 1919 provided the opportunity for a test of this prediction. An expedition led by Arthur Eddington confirmed that light rays were indeed deflected.

These tests were not sufficiently accurate[29] to ensure that small deviations from GR were not possible. It was also suspected that, since the real universe is not completely homogeneous and isotropic at all scales, Friedmann and Lemaître's prediction of a true singular beginning to the universe would ultimately fail. Perhaps a slight anisotropy could result in matter's "sling-shotting" past itself at a minimum (but nonzero) radius condition, so that the present expansion was preceded by a cosmic contraction, thereby avoiding the absolute beginning of the universe. In 1970, however, Stephen Hawking and Roger Penrose proved that the homogeneity/isotropy assumption was irrelevant. The Hawking–Penrose singularity theorems showed that so long as the universe is governed by GR (with a few technical exceptions, which will become prominent in our discussion later into the chapter), our past must include a singularity (Hawking & Penrose 1970). Wald comments:

> [Hawking–Penrose 1970] gives us strong reason to believe that our universe is singular . . . the observational evidence strongly suggests that our universe – or, at least, the portion of our universe within our causal past – is well described by a Robertson-Walker model [standard hot Big Bang theory] at least back as far as the decoupling time of matter and radiation. However, in these models, the expansion of the past directed null geodesics emanating from the event representing us at the present time becomes negative at a much more recent time than the decoupling time. Thus there is strong reason to believe that condition 4c of [Hawking–Penrose 1970] is satisfied in our universe. Since we expect that conditions (1)-(3) also are satisfied, it appears that our universe must be singular. Thus, it appears that we must confront the breakdown of classical general relativity expected to occur near singularities if we are to understand the origin of our universe. (Wald 1984, p. 241)

The conditions Wald mentions are:

1. Satisfaction of the strong energy condition (typically obeyed by "normal" types of matter).

---

[29] However, further experiments *did* definitively establish the correspondence between GR and nature. The 1993 Nobel Prize for Physics was awarded to two astronomers: Russell A. Hulse and Joseph H. Taylor, Jr. The award was given for their study of a distant solar system consisting of a binary pulsar – two neutron stars orbiting each other. GR predicted that the orbit would shrink over time due to the emission of gravitational waves. They proved that GR is accurate to a startling degree of one part in 10^{14}. This makes GR perhaps the best-proved theory in all of physics.

2. Satisfaction of the generic energy condition (there is no exotic property of the space-time that prevents gravitational focusing).
3. No closed time loops (the future does not bend back and become one's own past).
4. There is a point *p* such that past-directed worldlines emanating from *p* have negative expansion; that is, they are focused back on each other (the worldlines of observers do not trace into an infinite past given sufficiently concentrated matter but are "focused," as by a lens, into a singular condition within a finite time).

2.33. Exceptions to the Hawking–Penrose theorems

Four possible exceptions to the Hawking–Penrose singularity theorems conveniently distinguish four classes of nonstandard models that provide possible alternatives to the standard Big Bang model (Figure 3.3). The Hawking–Penrose Theorem also has the obvious, but implicit, condition that GR is fundamental; that is, it is a complete as well as correct description of conditions within our universe (thereby defining a 5th condition).

The first option (closed time loops) has been the subject of some exploration in cosmological circles. The next two – eternal inflation and quantum gravity – represent areas of fertile cosmological investigation which merit our attention. The last two exception conditions are not expected to be part of "reasonable" physical models of the universe. Hawking explains:

> Between 1965 and 1970 Penrose and I used the techniques I have described to prove a number of singularity theorems. These theorems had three kinds of conditions. First there was an energy condition such as the weak, strong or generic energy conditions. Then there was some global condition on the causal structure such as that there shouldn't be any closed time like curves. And finally was some condition that gravity was so strong in some region that nothing could escape. . . .

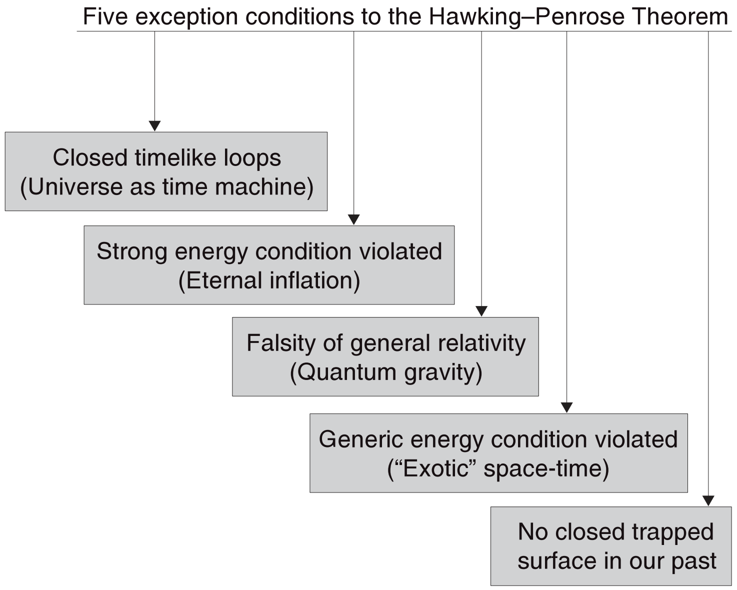


Figure 3.3 Model classes based on exceptions to the Hawking–Penrose singularity theorems.

> [The generic energy condition] says that first the strong energy condition holds. Second, every timelike or null geodesic encounters some point where there is some curvature that is not specially aligned with the geodesic. The generic energy condition is not satisfied by a number of known exact solutions. But these are rather special. One would expect it to be satisfied by a solution that was "generic" in an appropriate sense [that is, a reasonable physical model]. If the generic energy condition holds, each geodesic will encounter a region of gravitational focusing. (Hawking & Penrose 1996, pp. 14–5)

We do expect our past to feature a closed, trapped surface, and there is no reason to postulate an exotic space-time construction that would just happen to have perfect defocusing characteristics so as to counter the effects of gravity. Hence, our discussion will revolve around the first three options.

I. Closed timelike curves (CTCs)

A first, exotic exception to the Hawking–Penrose theorems is the possible existence of CTCs. Permitted by Einstein's GR, CTCs represent an observer tracing out a circular path through space and time.

J. Richard Gott and Li-Xin Li have proposed a model according to which the early universe (only) is a closed time loop that occasionally gives "birth" to a universe like ours (Figure 3.4). They maintain that Alexander Vilenkin's "tunneling from nothing" model (see section IVc) should properly be taken as tunneling from a previous state. As will be seen in this chapter, most cosmological models assert that the past terminates at a boundary a finite time ago. One then wishes to explain what exists at that boundary. Vilenkin (and, independently, the team of Stephen Hawking and James Hartle) believes that the universe sprang into being "out of nothing." Gott and Li believe, instead, that there is a CTC at this boundary.

Figure 3.4 A Gott–Li universe time machine.

The region of closed timelike curves (CTC) exists at the bottom of the diagram. This region is separated from the future evolution of the universe by a Cauchy horizon (a boundary that separates space-time into CTC regions and non-CTC regions). The four branches to the top of the diagram can be thought of as inflationary bubbles undergoing a de Sitter-like expansion (see sections II and IIIa for a discussion of inflation and de Sitter spaces). (source: Gott & Li 1998)

The Gott–Li model could be an example of a universe timeline looping back upon itself to allow the universe, in Gott and Li's words, "to become its own mother." They explain:

> In this paper, we consider instead the notion that the Universe did not arise out of nothing, but rather created itself. One of the remarkable properties of the theory of general relativity is that in principle it allows solutions with CTCs. Why not apply this to the problem of the first-cause?
>
> Many inflationary models allow creation of baby inflationary universes inside black holes, either by tunneling across the Einstein-Rosen bridge, or by formation as one approaches the singularity. *If one of these baby universes simply turns out to be the universe we started with, then a multiply connected model with early CTCs bounded by a Cauchy horizon is produced*.
>
> . . . Then the Universe neither tunneled from nothing, nor arose from a singularity; it created itself. (Gott & Li 1998, p. 39; emphasis added)

Some histories in our past were circular in nature (forming a multiply connected spacetime) and recreating the Big Bang. Further, this is not a cyclic universe; *it is the same Big Bang*.

The CTC scenario raises interesting philosophical questions about the nature of time (see below, p. 191). But here, our interest is in the model's physical viability. The primary physical problem confronting CTC models in general is their violation of the so-called Chronology Protection Conjecture (CPC).

Gott and Li indicate that "the region of CTCs . . . should be in a pure vacuum state containing no real particles or Hawking radiation and no bubbles"(Gott & Li 1998, p. 39). This is so because this stray radiation would destroy the CTC. The reason for this curious feature of a CTC model was discussed by Stephen Hawking (1992), where he formally suggested a "CPC." His theory was that a time machine (CTC) would have characteristics that were so unstable that it would quickly destroy itself. Hence, nature conspires to prevent time machines. A popular level (and entertaining) description of this effect is given by GR theorist Kip Thorne. He constructs a scenario that allows a local time machine to exist with one end on a spaceship departing Earth with his wife Carole and the other end on Earth with him in his living room (this is an exotic general relativistic structure called a wormhole).

> Imagine that Carole is zooming back to Earth with one wormhole mouth in her spacecraft, and I am sitting at home on Earth with the other. When the spacecraft gets to within 10 light-years of Earth, it suddenly becomes possible for radiation (electromagnetic waves) to use the wormhole for time travel: any random bit of radiation that leaves our home in Pasadena traveling at the speed of light toward the spacecraft can arrive at the spacecraft after 10 years' time (as seen on Earth), enter the wormhole mouth there, travel back in time by 10 years (as seen on Earth), and emerge from the mouth on Earth at precisely the same moment as it started its trip. The radiation piles right on top of its previous self, not just in space but in spacetime, doubling its strength. What's more, during the trip each quantum of radiation (each photon) got boosted in energy due to the relative motion of the wormhole mouths (a "Doppler-shift" boost).
>
> After the radiation's next trip out to the spacecraft then back through the wormhole, it again returns at the same time as it left and again piles up on itself, again with a Doppler-boosted energy. Again and again this happens, making the beam of radiation infinitely strong.

> In this way, beginning with an arbitrarily tiny amount of radiation, a beam of infinite energy is created, coursing through space between the two wormhole mouths. As the beam passes through the wormhole . . . it will produce infinite spacetime curvature [i.e. a singularity] and probably destroy the wormhole, thereby preventing [a time machine from coming into being in the first place]. (Thorne 1994, pp. 505–6)

Of interest to us is the general applicability of this effect to the Gott–Li model. Gott and Li are sensitive to this problem and have developed a solution[30]. They and others have found some specially constructed space-times that appear to elude Hawking's CPC. To avoid the CPC, they have constructed a special initial state for the universe: a zero-temperature empty space called an "adapted Rindler vacuum." It is specially built and balanced such that it does not develop the destructive effect suggested by Thorne earlier.

After the publication of Gott and Li's paper, William Hiscock developed a defense of the CPC that still appears to stand (Hiscock 2000). First, Hiscock argues that the Gott–Li choice of initial conditions is highly fine-tuned. In fact, Gott–Li's vacuum is of "measure zero" in the set of all possible Rindler vacuums. This means that the scenario is just about as unlikely as is possible without ruling it out summarily. D. H. Coule agrees in his summary of quantum gravity models, referring to the Gott–Li model as "rather contrived" (Coule 2005).[31] Second, Hiscock argues that the Gott–Li vacuum is not stable, given more realistic physical force fields. He writes:

> . . . the (Rindler) vacuum stress-energy of a nonconformally coupled scalar field, or a conformally coupled massless field with a . . . self-interaction will diverge on the chronology horizon for all values of the Misner identification scale [this is the parameter that Gott–Li have fine-tuned]. In addition, the vacuum polarization of [the scalar field considered in the Gott–Li model] diverges in all cases [leading to the Thorne effect cited earlier], even for the conformally invariant case examined by Li and Gott. Hence, the regular behavior found by Cassidy and Li and Gott holds only for a conformally invariant, non-interacting field, and only for the stress-energy tensor. While some fields in nature (e.g., the electromagnetic field, before interactions are added) are conformally invariant, others – notably gravity itself–are not; and interactions are the rule, not the exception. (Hiscock 2000, p. 4)

---

[30] Gott indicates:

> I think no one has been able to rule out CTC's. There have been no significant changes in quantum gravity since our paper. To understand whether one can create a time machine one may have to understand quantum gravity and we do not yet. Several loopholes in chronology protection have been found. Li Xin-Li and I and Cassidy, Hawking's student have found examples of quantum vacuum states that do not blow up on the cauchy horizon. Li Xin-Li's paper on the correct renormalization procedure [Phys. Rev. D. 084016 (1999)] showed that the vacuum field did not blow up for electromagnetic fields and other fields as well as for scalar fields. This means the action or entropy does not blow up either–solving a trouble Hawking and Cassidy thought existed. For matter obeying the weak energy condition instabilities are cured if the time loop occurs at the beginning of the universe, as we are proposing. (pers. comm., March 1, 2008)

[31] Coule has some additional objections of his own. For example, he criticizes the nature of the vacuum, indicating that thermal fluctuations need to be of a precise form in order to avoid the radiation backreaction described earlier by Thorne. This expectation is inconsistent with the Planck scale physics employed by Gott and Li.

Coule adds: "... in Misner space this state [Gott–Li model] was only possible with identification scale $b = 2\pi$, or $b = 2\pi r0$ for the multiple de Sitter case. Such an exact value is itself inconsistent with notions of quantum uncertainty" (Coule 2005, p. 31). So the Heisenberg uncertainty principle of quantum mechanics (QM) would guarantee that the relevant parameter could not be "just-so." But if it is not "just-so," then the universe collapses into a singular condition in the presence of a time machine. Coule also suggests that this parameter, called the "Misner identification scale" is not a constant. Rather it is likely to change dynamically as a function of matter couplings or energy potentials. As soon as it does, the CTC will destabilize.

Interestingly, Gott and Li used similar objections when arguing for their model at the expense of the "creation from nothing" approach of Vilenkin and Hartle and Hawking [section IVc]. Gott and Li criticize the "creation from nothing" approach on the grounds of the uncertainty principle and the fact that their competitors are not using realistic force fields; that is to say, the Vilenkin approach is not close enough to what we expect for the real universe. Yet their own model appears to break down when similar objections are leveled against *it*.[32,33]

CTC physics is interesting, and while some theorists still pursue it, it occupies only a small minority of ongoing cosmological investigation. While it is true that no one has been able definitively to rule out CTCs, the evidentiary burden lies upon those defending the viability of such space-times and models predicated upon their reality.

II. *Eternal inflation*

Motivation

A more serious exception to the Hawking–Penrose singularity theorems is afforded by inflationary theory. Although the Friedmann–Lemaître model had a great deal of evidential support, there were, nonetheless, observational anomalies which suggested that there was more to the story. There were also theoretical reasons to think that the description was not quite complete. These difficulties, especially the horizon, flatness, and cosmic relic problems, prompted theorists to propose a modification of the standard Big Bang picture called "inflation."

With respect to the horizon problem, cosmologists lacked an explanation as to *why* the universe should be so homogeneous and isotropic.[34] Alan Guth explains:

> The initial universe is assumed to be homogeneous, yet it consists of at least ~10^{83} separate regions which are causally disconnected (*i.e.*, these regions have not yet had time to commu-

---

[32] In fairness to Gott and Li, it should be noted that Hiscock's criticisms are based on a semiclassical approach (an approximation of quantum gravity), and it is possible that a full theory of quantum gravity could vindicate their idea.

[33] Vilenkin has also criticized the Gott–Li model (see Vilenkin 2006, p. 219). He indicates that the Gott–Li space-time contains incomplete histories, so, "This means that the spacetime itself is past-incomplete, and therefore does not provide a satisfactory model of a universe without a beginning."

[34] The universe does *appear* different at various distances as we look at it. But this is due to the fact that we observe distant galaxies as they were in the past, given the time it takes for their light to reach us.

> nicate with each other via light signals) . . . Thus, one must assume that the forces which created these initial conditions were capable of violating causality. (Guth 1981, p. 347)

Cosmology had an appropriate "organizing" principle at hand – thermodynamic equilibrium – yet mathematics showed that, in the limit, as one looked backward at the Big Bang, the different parts of the universe would lose causal communication. Without causal communication, all the parts of the observable universe could not have cooperated in energy transfer so as to make all parts look the same (in the present). Physicist Brian Greene describes the horizon problem:

> Physicists define a region's *cosmic horizon* (or *horizon* for short) as the most distant surrounding regions of space that are close enough to the given region for the two to have exchanged light signals in the time since the [Big] bang. . . . The *horizon problem*, then is the puzzle, inherent in the observations, that regions whose horizons have always been separate – regions that could never have interacted, communicated, or exerted any kind of influence on each other – somehow have nearly identical temperatures. (Greene 2004, p. 289; emphasis in the original)
>
> . . . imagine running the cosmic film in reverse while focusing on two regions of space currently on opposite sides of the observable universe – regions that are so distant that they are beyond each other's spheres of influence. If in order to halve their separation we have to roll the cosmic film more than halfway back toward the beginning, then even though the regions of space were closer together, communication between them was still impossible: they were half as far apart, but the time since the bang was *less* than half of what it is today, so light could travel only *less* than half as far. Similarly, if from that point in the film we have to run more than halfway back to the beginning in order to halve the separation between the regions once again, communication becomes more difficult still. With this kind of cosmic evolution, even though regions were closer together in the past, it becomes more puzzling – not less –that they somehow managed to equalize their temperatures. Relative to how far light can travel, the regions become increasingly cut off as we examine them ever farther back in time. This is exactly what happens in the standard big bang theory. (Greene 2004, p. 288; emphasis in the original)

A second problem was that the universe appears to be "flat" (i.e. space is Euclidian: the angles of a triangle add up to 180 degrees; parallel lines do not intersect), while GR predicts that that is a wildly improbable outcome.

> A typical closed universe will reach its maximum size on the order [of the Planck scale of 10^{-44} sec], while a typical open universe will dwindle to a ρ [density] much less than ρ_{cr} [critical density; the density for a long-lived universe]. A universe can only survive $\sim 10^{10}$ years [approximately the age of our universe] only by extreme fine tuning. . . . For [the likely initial conditions for our universe] the value of H_0 [the initial expansion rate of the universe] must be fine tuned to an accuracy of one part in 10^{55}. In the standard model this incredibly precise initial relationship must be assumed without explanation. (Guth 1981, p. 348)

The third problem was that the supposition of an initial disorganized state of the universe led to the prediction of the presence of bizarre cosmic relics. Magnetic monopoles should appear in our universe at a density amenable to detection with our present means. In Guth's original paper on inflation, he indicates that standard particle physics predicts a monopole concentration 14 orders of magnitude greater than the upper bound observed in our universe. To date, we have seen none of these exotic structures.[35]

Guth's solution to these three problems was to postulate a period of exponential expansion very early in the history of the universe. Again Greene:

> In inflationary cosmology, there was a brief instant during which gravity was repulsive and this drove space to expand faster and faster. During this part of the cosmic film, you would have to wind the film less than halfway back in order to halve the distance between the two regions. . . . the increasingly rapid separation of any two regions of space during inflationary expansion implies that halving their separation requires winding the cosmic film less – *much less* – than halfway back toward the beginning. As we go farther back in time, therefore, it becomes *easier* for any two regions of space to influence each other, because, proportionally speaking, there is more time for them to communicate. Calculations show that if the inflationary expansion phase drove space to expand by at least a factor of 10^{30}, an amount that is readily achieved in specific realizations of inflationary expansion, all the regions in space that we currently see . . . were able to communicate . . . and hence efficiently come to a common temperature in the earliest moments of the universe. (Greene 2004, pp. 289–90; emphasis in the original)

This inflationary period began and ended in a fraction of a second, yet a typical inflationary event could lead to 70 "e-folds." An e-fold is a logarithmic measure of how large the universe grows during an inflationary event. Here, N is the number of e-folds, and $a(t)$ represents the scale factor of the universe at the beginning and end of inflation.[36]

$$N(t) \equiv \ln\left[a(t_{end}) / a(t_{beginning})\right]$$

E-folds are a shorthand way of expressing the huge increase in size of the universe during an inflationary event (recall Greene's factor of 10^{30}).

Hence, prior to inflation all parts of the present, observable universe could be in causal communication with one another. Inflationary expansion would also smooth the curvature of the present-day universe to be flat or nearly flat, similar to the way the curvature of a basketball would appear to vanish if it suddenly grew to the size of the Earth. Further, since our present observable universe would be only a microscopic part of the original generic manifold, the density of exotic cosmic relics would be expected to be so small that we should not see them.

Inflation was a remarkable fix to a set of serious anomalies; but it also had one more feature in store. The Hawking–Penrose singularity theorems had as one of their requirements that gravity is always attractive – just as it is for ordinary matter. But the

---

[35] To understand what these exotic structures represent, consider the analogy of a pond freezing in wintertime. If the pond starts freezing in one place and the ice simply grows until it encompasses the whole pond, you will have a smooth surface. But if different parts of the pond start freezing separately, ultimately, these growing "icebergs" must meet at a boundary. (Imagine taking big rocks and cracking holes in the ice; then letting it refreeze. The boundaries will be rough.) The early universe was similar. These boundaries are called "defects" and can be zero-, one-, or two-dimensional. Zero-dimensional defects are called magnetic monopoles. One-dimensional defects are called cosmic strings. Two-dimensional boundaries are called domain walls.

[36] Definition of e-fold available at http://astro.uchicago.edu/~cunha/inflation/node4.html.

most likely physical candidate that could account for an inflationary event was a type of energy similar to the original cosmological constant that Einstein had proposed (Einstein 1917). This bizarre type of energy would act like repulsive gravity. This led to a philosophically desired outcome. If this "repulsive gravity" was present in the early universe and could dominate attractive gravity, then the possibility arises that the Hawking–Penrose singularity theorems did not apply to the real universe. Perhaps the universe is past eternal after all.

Inflationary theorizing eventually led to a yet grander theory, according to which the gravitationally repulsive material may, in fact, be the norm in the universe rather than the exception. Cosmologist Andrei Linde explains:

> This process, which I have called eternal inflation, keeps going as a chain reaction, producing a fractal-like pattern of universes. In this scenario the universe as a whole is immortal. *Each particular part of the universe may stem from a singularity somewhere in the past, and it may end up in a singularity somewhere in the future.* There is, however, no end for the evolution of the entire universe.
>
> The situation with the very beginning is less certain. There is a chance that all parts of the universe were created simultaneously in an initial big bang singularity. The necessity of this assumption, however, is no longer obvious.
>
> Furthermore, the total number of inflationary bubbles on our 'cosmic tree' grows exponentially in time. Therefore, most bubbles (including our own part of the universe) grow indefinitely far away from the trunk of this tree. Although this scenario makes the existence of the initial big bang almost irrelevant, for all practical purposes, one can consider the moment of formation of each inflationary bubble as a new 'big bang.' From this perspective, inflation is not a part of the big bang theory, as we thought 15 years ago. On the contrary, the big bang is a part of the inflationary model. (Linde 1998, p. 103; emphasis added)

Linde's chaotic inflation was one of two competing views for the theory. The competitor, called "new" inflation, featured the idea that there is a "false vacuum," which represents (meta)stable vacuum with a high Einstein-like cosmological constant (compared with the "true vacuum" in which we live). In "new inflation," this eternally expanding false vacuum regionally decays into the slower expanding true vacuum. It expands faster than it decays, so the process never stops.

Chaotic inflation is rather a different idea. Here, the universe starts from a "generic manifold," which is a state of maximal entropy chaos. An energy field of different regional values pervades this manifold. Where the field is large, inflation occurs. The field can undergo quantum fluctuation to high values as well; thereby giving onset to inflation. Locally, a region will have a tendency to seek out the minimum allowed value of the energy field. This leads to a process called "reheating," which creates the ordinary matter and energy that we see around us. Meanwhile, in locations of the universe where the field energy density is high, the quantum fluctuations will tend to offset the minimizing tendency and make perpetual (globally, but not regionally) the inflationary process (Figure 3.5).

Eternal inflationary models

The 1980s and 1990s witnessed a proliferation of inflationary models that theoretically allowed for a projection into an eternal past (Linde 2005). The inflationary phase mentioned earlier was not viewed in these models as an isolated event. Theorists began to describe the

Figure 3.5 In chaotic inflation, an initial generic manifold (a global space with random characteristics; that is, inhomogeneous and anisotropic in energy density and curvature) undergoes regional inflation. An observer sitting on an inflating spot will eventually see inflation come to a stop, but conditions are always suitable – somewhere – for inflation to proceed.

exotic energy that produces inflation as a field that pervades otherwise empty space. A key assumption was that the density[37] of the energy throughout space never changes, so that it resembles Einstein's cosmological constant. It does not depend on space or time, that is, it is constant. In that case, as space expands, more energy must continually be produced in order to maintain a constant energy density (where this energy comes from is still a matter of controversy). Space "clones" itself.[38] Occasionally, parts of this rapidly expanding space decay (convert) into the type of "empty" space that we live in. This space has a much lower energy density, so there is now a great deal of excess energy that pervades our new "bubble." This excess energy is thought to convert into the normal matter that we see around us.

In the latest version of inflationary theory (Figure 3.6), these decays take the form of quantum tunneling events. Every state that possesses a positive cosmological constant is "metastable." That means that, similar to a radioactive isotope, the state lasts for a while and then changes to a different (usually lower) allowed value of the cosmological constant. This lower-energy state is initially confined to a tiny portion of space, but given that the cosmological constant makes space expand, it becomes a rapidly growing bubble that is nested within the original space.

But what happens to the original space, part of which decayed to form our universe? It is still there, continuing to expand at enormous speed. Since it (usually) has a cosmological constant larger than the new bubble, its growth outpaces that of the new bubble. Since the false vacuum expands faster than it decays, inflation is eternal into the future. New bubbles of low-energy vacuum will continue to decay out of the expanding space.

---

[37] In Linde's chaotic inflation, the energy field does feature quantum fluctuations that are critical to the onset of new inflationary patches.

[38] Since only the "material" cause is missing, this process is an example of genuine *creatio ex nihilo* seen by physical theorists in the present day. Such recognition of efficient causation in the absence of material causation may serve to mute objections to theistic *creatio ex nihilo* as featured in the *kalam* cosmological argument.

Publisher's Note:
Permission to reproduce this image online was not granted by the copyright holder. Readers are kindly requested to refer to the printed version of this chapter.

Figure 3.6 The String Landscape inflationary model. The Big Bang is just a regional event within a larger multiverse. There are different kinds of "empty space," which feature different values of the cosmological constant (different bubbles). The larger the constant, the faster the universe expands. Our universe decayed from one of these "false vacuum" regions. (inset adapted from the National Aeronautics and Space Administration, available at: http://map.gsfc.nasa.gov/news/index.html)

Theorists wondered whether this process could be infinitely extended into the past. Interestingly, Guth himself, along with collaborators Alexander Vilenkin and Arvind Borde, has likely closed the door on that possibility. In 2003, Borde, Guth, and Vilenkin published an updated singularity theorem far grander in scope than the Hawking–Penrose theorems. They explain,

> Our argument shows that null and time-like geodesics are, in general, past-incomplete in inflationary models, whether or not energy conditions hold, provided only that the averaged expansion condition $H_{av} > 0$ holds along these past-directed geodesics. (Borde, Guth, & Vilenkin 2003, p. 3)[39]
>
> A remarkable thing about this theorem is its sweeping generality. We made no assumptions about the material content of the universe. We did not even assume that gravity is described by Einstein's equations. So, if Einstein's gravity requires some modification, our conclusion will still hold. The only assumption that we made was that the expansion rate of the universe

---

[39] H_{av} refers to the average value of the Hubble constant throughout history.

> never gets below some nonzero value, no matter how small. This assumption should certainly be satisfied in the inflating false vacuum. The conclusion is that past-eternal inflation without a beginning is impossible. (Vilenkin 2006, p. 175)

Vilenkin affirms that any universe (including universes modeled by higher dimensional cosmology, pre–Big Bang cosmology, and so forth,) which, on average, expands has to connect, in a finite time, to a past boundary (pers. comm., March 4, 2004).

Intuitively, the reason that the universe must have a beginning in the finite past is that, in an expanding space, an observer tracing out a worldline (to the future) slows down. This is the redshift. Vilenkin explains:

> Let us now introduce another observer who is moving relative to the spectators [each of whom is motionless except for the expansion of space]. We shall call him the space traveler. He is moving by inertia, with the engines of his spaceship turned off, and has been doing so for all eternity. As he passes the spectators, they register his velocity.
>
> Since the spectators are flying apart [i.e. the universe is expanding], the space traveler's velocity relative to each successive spectator will be smaller than his velocity relative to the preceding one. Suppose, for example, that the space traveler has just zoomed by the Earth at the speed of 100,000 kilometers per hour and is now headed toward a distant galaxy, about a billion light years away. That galaxy is moving away from us at a speed of 20,000 kilometers per second, so when the space traveler catches up with it, the observers there will see him moving at 80,000 kilometers per second.
>
> If the velocity of the space traveler relative to the spectators gets smaller and smaller into the future, then it follows that his velocity should get larger and larger as we follow his history into the past. In the limit, his velocity should get arbitrarily close to the speed of light. (Vilenkin 2006)[40]

So, looking into the past, the observer must be seen to speed up. But one cannot exceed the speed of light. The implication of this is that the past worldline of this observer has a finite length. This is the symptom of singularity; the "pathology" that Robert Wald referred to earlier. The observer will have "begun its existence a finite time ago."

The Borde–Vilenkin–Guth (BVG) singularity theorem is now widely accepted within the physics community. As of this writing, it has gone largely unchallenged.[41] Instead a new round of model building has resulted based on exceptions to *this* theorem. Four alternatives present themselves (Figure 3.7).

---

[40] Alan Guth, in a 2003 lecture at the University of California Santa Barbara's Kavli Institute, says: "If we follow the observer backwards in an expanding universe, she speeds up. But the calculation shows that if $H_{\text{average}} > 0$ in the past, then she will reach the speed of light in a finite proper time." (See http://online.kitp.ucsb.edu/online/strings_c03/guth/pdf/KITPGuth_2up.pdf.)

[41] Andrei Linde has offered a critique, suggesting that BVG imply that all the individual parts of the universe have a beginning, but perhaps the WHOLE does not. This seems misconstrued, however, since BVG are *not* claiming that *each* past inextendible geodesic is related to a *regional* singularity. Rather, they claim that Linde's universe description contains an internal contradiction. As we look backward along the geodesic, it *must* extend to the infinite past if the universe is to be past eternal. But it does not (for the observer comoving with the expansion). Rather, past inextendible geodesics are the "symptom," not the "disease." As Robert Wald (1984, p. 216) says, "Unfortunately, the singularity theorems give virtually no information about the nature of the singularities of which they prove existence." So we do not know the nature of the singularity that the BVG Theorem indicates; we know only that Linde's description of an infinite past is in error.

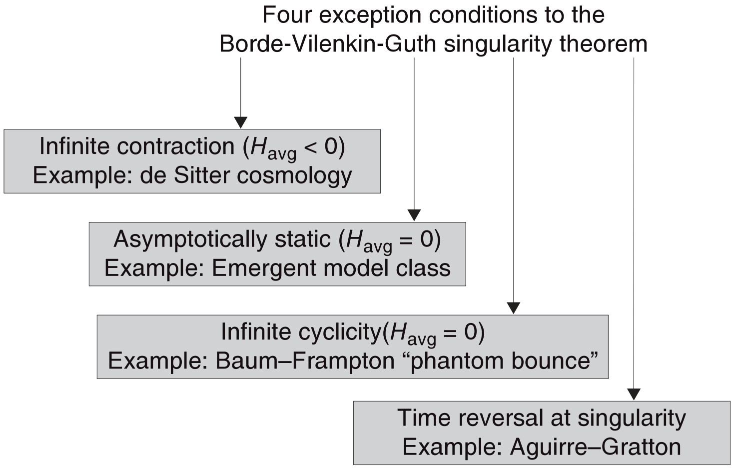


Figure 3.7 Post-2003 cosmological model building based on finding exceptions to the Borde–Vilenkin–Guth Theorem.

IIIa. Infinite contraction

Assume that a spatially infinite universe contracted down to a singularity and then "bounced" into our present expansion. In such a case, the universe cannot be said to be, on average, in a state of cosmic expansion throughout its history since the expansion phase, even if infinite, is canceled out by the contraction phase. While permissible under the BVG Theorem, this option is not, however, a popular option among contemporary cosmologists. George Ellis identifies two problems that bedevil such an approach:

> The problems are related: first, initial conditions have to be set in an extremely special way at the start of the collapse phase in order that it is a Robertson-Walker universe collapsing; and these conditions have to be set in an acausal way (in the infinite past). It is possible, but a great deal of inexplicable fine tuning is taking place: how does the matter in widely separated causally disconnected places at the start of the universe know how to correlate its motions (and densities) so that they will come together correctly in a spatially homogeneous way in the future??
>
> Secondly, if one gets that right, the collapse phase is unstable, with perturbations increasing rapidly, so only a very fine-tuned collapse phase remains close to Robertson-Walker even if it started off so, and will be able to turn around as a whole (in general many black holes will form locally and collapse to a singularity).
>
> So, yes, it is possible, but who focused the collapse so well that it turns around nicely? (pers. comm., January 25, 2006)

So there is a significant problem of *acausal* fine-tuning. One asserts not just brute contingency but also a rather curious form of it. In the face of apparent fine-tuning, physicists usually prefer to offer some type of explanation. Consider, for example, multiverse models as an explanation of the apparent fine-tuning of the fundamental physical constants, or Guth's inflationary resolution of the horizon problem (past thermodynamic equilibrium).

Second, there is the problem that the collapse becomes chaotic as it approaches the singularity. This will produce a preexpansion start condition that is known to be dramatically different from our actual "Big Bang." This phenomenon is referred to as "BKL chaos" after its discoverers (see Belinsky, Khalatnikov, & Lifshitz 1970).[42] This problem will appear for all attempts at a past-eternal timeline that seek to introduce a pre–Big Bang phase that "bounces" into the present expansion. In fact, the true implication of BKL may well be that it is physically impossible to "bounce" *through* a singularity.

In stating that the initial conditions "have to be set in an acausal way (in the infinite past)," Ellis puts his finger on a nettlesome philosophical issue in cosmological models featuring an infinite past, namely, they often *seem* to treat the infinite past as though it featured an infinitely distant beginning point. Several of these models are discussed in this essay. But, as we have already seen in our discussion of philosophical *kalam* arguments, such a supposition is illicit, since such an infinitely distant point is merely an ideal limit characteristic of the potential infinite, not a moment that actually once was present.[43] If we are allowed to speak of the condition of the universe at past infinity, then Zenonian paradoxes (see p. 119) are unavoidable.

If the past condition of the universe is acausal, then of course there was no "setting" of the condition; it just "is." Ellis is referring merely to the construction of the mathematical model. But suppose we do imagine that the boundary conditions were literally set at past infinity. Something like this was a feature of Charles Misner's old "Mixmaster" universe:

> In reality we don't expect universes to expand at exactly the same rate in every direction, and when they become asymmetrical like this they behave in a very complicated way. Although they expand in volume, one direction tends to contract while the other two expand, tending to create an expanding 'pancake'. But soon the contracting direction switches to expansion and one of the other two expanding directions switches into contraction. Over a long period of time, the effect is a sequence of oscillations . . . The striking thing about the sequence of oscillations of the volume of the universe as it shrinks to zero, when one runs its history back into the Big Bang at time-zero, or on into the Big Crunch at crunch-time, is that an infinite number of oscillations occur. . . . The difference between the Mixmaster Universe and Zeno's paradox is that an infinite number of physically distinct, real events happen in any finite interval of time that includes time-zero or crunch time. Measured by a clock that 'ticks' on this oscillatory time, the Mixmaster Universe would be judged to be infinitely old, because an infinite number of things have happened to the past in this time, and it will 'live' forever because an infinite number of things are still to happen in the future. (Barrow 2005, pp. 242–3)

The Mixmaster universe is interesting in that it appears to offer a past infinite timeline that nonetheless features a clear past boundary to that timeline; that is, an infinitely distant

---

[42] Also, see Damour and Henneaux (2000): ". . . our findings suggest that the spatial inhomogeneity continuously increases toward a singularity, as all quasi-uniform patches of space get broken up into smaller and smaller ones by the chaotic oscillatory evolution. In other words, the spacetime structure tends to develop a kind of 'turbulence'."

[43] In response to the question, "Are c-boundaries [see Figure 3.8 and the following discussion for explanation of these terms] such as past and future timelike infinity and scri+ physically real edges to spacetime (real, as a black hole is an ontologically real entity) or are they merely mathematical conveniences? But if infinity is 'actual' and reachable, then a c-boundary must be an actual edge to spacetime, physically real in its ontology," Ellis (pers. comm.) responds curtly:

1. no
2. maths – after all a spacetime diagram is just a representation of physical reality
3. in my view infinity is neither actual nor reachable.

beginning point. There *is* a question of judging the most physically appropriate measure of time. By proper time, Mixmaster arose a finite time ago from a singularity and will end its existence a finite time to the future. Time measured by oscillatory "ticks" would report a timeline that is infinite to the past and future. Barrow and Tipler elucidate:

> It is always possible to find a conformal transformation which will convert an infinite universe to a finite one and vise-versa. One can always find a time coordinate in which a universe that exists for a finite proper time . . . exists for an infinite time in the new time coordinate, and a time coordinate in which a universe that exists for an infinite proper time . . . exists for only a finite time. The most appropriate *physical* time may or may not be the proper time coordinate. (Barrow & Tipler 1986, p. 636)

Physicists routinely consider an infinitely distant past "beginning" point, in effect bringing infinity into their physical models through a process called a conformal transformation. Consider Barrow and Tipler, here explaining simple Friedmann–Robertson–Walker (FRW) cosmological models using a device called a Penrose diagram.

> The boundaries of a Penrose diagram represent what are termed c-boundaries of the cosmological models. The c-boundaries are composed of the singularities and the points at infinity; the c-boundary of a cosmology is the edge of space-time, the 'place' at which space and time begin. By convention, singularities are represented by double lines in Penrose diagrams. [For example] the initial and final singularities are the only c-boundaries in a closed Friedmann universe. An open Friedmann universe, on the other hand, has four distinct c-boundary structures: an initial singularity out of which the entire space-time arose, a single point i^0 representing spatial infinity, a 45° line ϑ^+ (called 'scri plus') representing 'null infinity' which are the points at infinity that light rays (null curves) reach after infinite time, and a single point i^+ which all timelike curves approach for all finite times, and reach after infinite time (with the exception of those timelike curves that accelerate forever and thus approach arbitrarily close to the speed of light. These curves hit scri plus rather that i^+ at temporal infinity).
>
> A Penrose diagram allows us to define rigorously 'an achieved infinity', a concept whose logical consistency philosophers have been doubtful about for thousands of years. Using the c-boundary, it is possible to discuss the topology of the 'achieved infinity' and the 'beginning of time' in cosmological models. (Barrow & Tipler 1986, pp. 635–6)

Models such as Mixmaster, the problem of supertasks[44] in general, and the meaning of conformal transformations raise the question of whether an infinite past implies the absence of a past boundary.

Figure 3.8 shows a Penrose diagram for the universe type under consideration in this section; that is, one that contracts from infinite size down to a singularity, and then bounces into an expanding universe (see right side of the diagram). Figure 3.9 shows another type of simplified model (a de Sitter universe) that has this type of behavior. The de Sitter model includes "dark energy," while the contracting model in Figure 3.8 includes only ordinary matter. A more realistic physical model would include both ordinary matter and "dark energy." The behavior of the universe at large size would be dominated

---

[44] A supertask is an infinite series of subtasks that can be completed in a finite time.

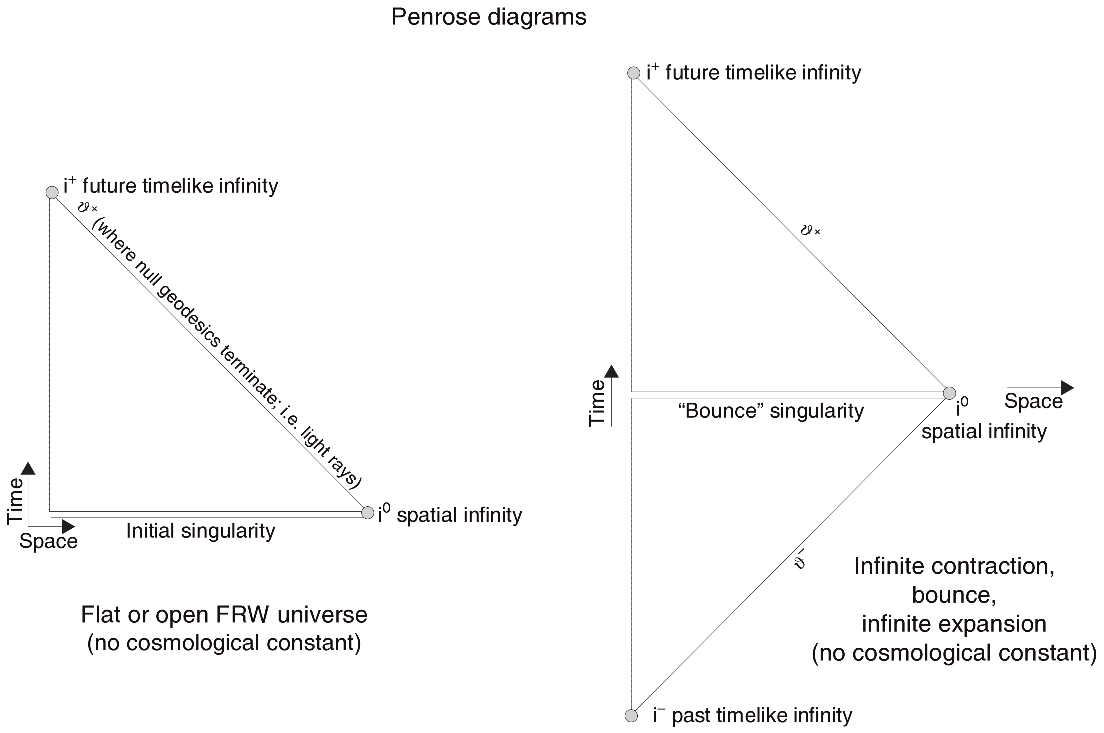
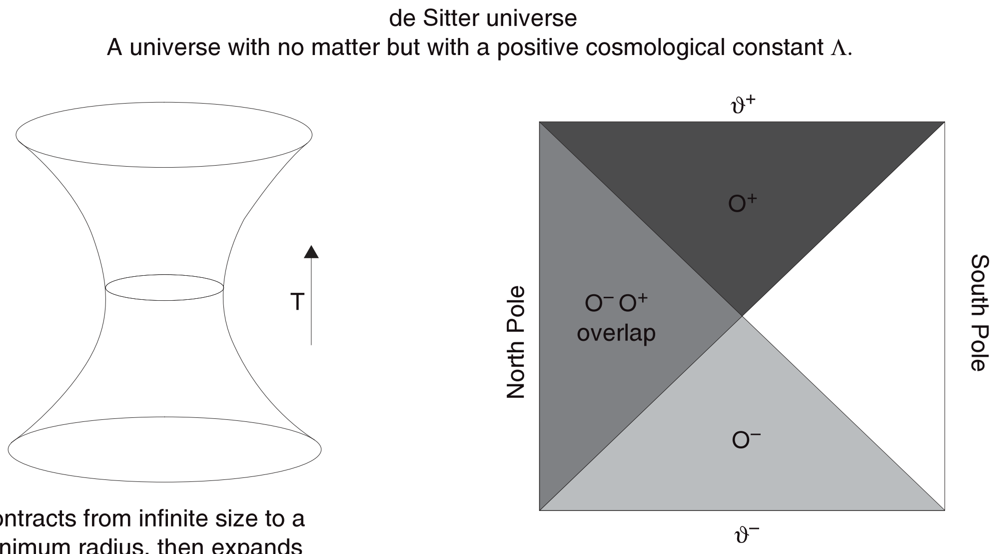


Figure 3.8 Penrose depiction of Friedmann–Robertson–Walker (FRW) cosmology.



Contracts from infinite size to a minimum radius, then expands exponentially. A more realistic model would include radiation and matter, which would cause a singularity at time zero, while the distant past and future would behave like the de Sitter solution.

Penrose diagram: South and North "Pole" represent observers tracing out timelike paths. O$^-$ indicates limits of causal past of North Pole observer. O$^+$ indicates causal future. No observer sees the full space-time; hence the mystery of the acausal fine-tuning.

Figure 3.9 A more realistic rendering of a universe that infinitely collapses down to a Big Bang, and then expands. The de Sitter cosmology takes into account the dominant behavior of a cosmological constant for a universe of large size.

by the dark energy, and so a de Sitter model yields good insight into the behavior of an infinite contraction model for the asymptotic past. By contrast, the behavior of the universe at the Big Bang would be best described by the Friedmann–Lemaître models in Figure 3.8.

It appears there is a dilemma. On the one hand, one could have the reality of a past infinite timeline without a beginning. But then one must assert brute contingency. "Things are as they are because they were as they were."[45] Further, one must do this with respect to apparent fine-tuning. This seems implausible. One can at least say that it is unpopular, given that cosmologists are avidly seeking an *explanation* for apparent fine tuning in the form of a multiverse or a superdeterministic Theory of Everything. If we are going to give up *explanation*, then what was wrong with leaving cosmology as it was prior to 1980, namely, the standard Hot Big Bang model (with associated breakdown of physics at the singularity)?[46]

The other horn involves an infinitely distant beginning point and allows the possibility of an explanation. But it opens the door for a supernatural explanation of the beginning. The *kalam* cosmological argument's second premise would be upheld, for that premise does not require that the beginning lie in the finite past.

IIIb. Asymptotically static space-time

An asymptotically static space is one in which the average expansion rate of the universe over its history is equal to zero, since the expansion rate of the universe "at" infinity is zero. Hence, the universe, perhaps in the asymptotic past, is in a static state (neither expanding nor contracting). This allows the model to escape the BVG singularity theorem.

At first blush, it would seem that the universe could hardly be said to have zero average expansion throughout its history if, as we know from observation, it has indeed been expanding! Would not the average expansion rate have to be greater than zero? No, not when we include "infinity" in the average. Consider an analogy in which the local government decides that, henceforth, everyone will pay property taxes according to the average value of property (per acre) in the county instead of on one's individual assessment. This might be good or bad for you, depending on whether you live in the high end district. But suppose that your county suddenly expanded to include the Sahara Desert. The Sahara is worthless and big, hence the average value of property, by the square mile, dives precipitously. Further, the larger the Sahara is, the closer to zero one's property taxes will be. In the limit as the Sahara grows to infinite size, one's property taxes will go to zero. In a similar way, a zero expansion condition at infinity would have the same impact on the average expansion rate. And the BVG Theorem only applies to a positive *average* expansion universe. George Ellis and his colleagues have been active in this type of model building. Models of this sort belong to what is called the "Emergent" model class. They rehabilitate Einstein's static model by postulating that the universe initially existed in such a phase and then transitioned via an inflationary phase into the universe we see around us today. Ellis and Maartens explain:

---

[45] Barrow and Tipler (1986, p. 408), attributed to cosmologist Thomas Gold.

[46] See, for example, Earman and Mosterin (1999) for a related argument.

> We show here that when K = +1 [recall the curvature parameter from Friedmann's equation] there are closed inflationary models that do not bounce, but inflate from a static beginning, and then reheat in the usual way. [Recall Guth's inflation] The inflationary universe emerges from a small static state that has within it the seeds for the development of the macroscopic universe, and we call this the "Emergent Universe" scenario. (This can be seen as a modern version and extension of the Eddington universe.) *The universe has a finite initial size, with a finite amount of inflation occurring over an infinite time in the past*, and with inflation then coming to an end via reheating in the standard way. (Ellis & Maartens 2004; emphasis added)[47]
>
> As such, it is a manifestly nonsingular closed inflationary cosmology that *begins from a meta-stable Einstein static state* and decays into a de Sitter phase and subsequently into standard hot Big Bang evolution. (Ellis, Murugan, & Tsagas 2004; emphasis added)
>
> A second, equally intriguing, possibility is that the initial Einstein static universe is created from "nothing" by some quantum tunneling process. Indeed, finiteness of the tunneling action requires that the universe created through instantonic tunneling be closed. It is not implausible, then, that through spontaneous quantum fluctuations, *a closed universe could be created in a long lived but transient Einstein static state* which then makes a transition to a finite lifetime de-Sitter and subsequent marginally closed FRW phase along the lines described above. (Ellis, Murugan, & Tsagas 2004; emphasis added)[48]

Now the question that interests us is whether the past of this model is perceived as eternal. A certain amount of ambiguity attends the answer. In some accounts (such as the above), it seems pretty clear that the Emergent models do have a beginning, namely, the Einstein static state (ESS). It is also stated explicitly that the model can be constructed with ESS occurring a finite time to the past (Ellis & Maartens 2004, sec. V). However, in the relevant papers ESS is usually described as asymptotically approached for past infinite time: "Here . . . we consider a universe filled with a dynamical scalar field, which is past asymptotic to an Einstein static model with a radius determined by the field's kinetic energy" (Ellis & Maartens 2004, p. 1). Some philosophers who have written on the topic have a problem with the contrived nature of the past infinity in models of this type. For example, Rüdiger Vaas characterizes the Emergent models as "soft-bang/pseudobeginning" in nature.

---

[47] Now we just showed in the previous section that Ellis has a philosophical problem with models that suggest an infinitely distant beginning point and even, in fact, the notion of a realized infinity in nature. Yet here we have a family of models developed by Ellis et al. that seem to suggest precisely that. This is explained via the following:

1. Infinity is so deeply ingrained in GR that pure pragmatism demands that one include the concept within one's work.
2. It is not cognitively dissonant to consider that one might be wrong and research accordingly. In fact, good scientific procedure includes an attempt to falsify one's own theories.
3. Ellis's collaborators may not have the same philosophical commitments.
4. An infinity that appears due to a technical interpretation of GR can disappear given a generalization of the theory (say, by considering quantum gravity).

A full look at Ellis's recent work indicates a bias toward models with compact spaces (i.e. spatially finite either through closed curvature or topology), a skepticism with regard to infinite multiverses, and openness toward the idea of a "pseudobeginning" in the finite past. In short, the pseudobeginning idea is that there is timeless reality where time "switches on," producing our present state of affairs.

[48] This is related to "creation from nothing" models; see section IVc.

He views the asymptotic approach toward ESS as something of a mathematical artifact (Vaas 2004, p. 18).

It is worth focusing on the issue of the instability of ESS. The Einstein static universe itself was originally viewed as past eternal. But there are obvious problems with this interpretation. The reason Einstein himself originally dropped the model was its feature of unstable equilibrium. Although, in pure nonquantum GR, one can consider a static state with worldlines that trace to negative infinite time, in reality we know that gravity is a quantum force. As Vilenkin notes, "Small fluctuations in the size of the universe are inevitable according to the quantum theory, and thus Einstein's universe cannot remain in balance for an infinite time" (Vilenkin 2006, p. 209).[49] On the other hand, the current observable universe is demonstrably *not* in a static state. A quantum (or perhaps a thermal) fluctuation is necessary to force a transition to an expanding universe. A fluctuation is, in fact, necessary for the two phase model to work. But this very mechanism implies that the *initial state* is not past eternal.

The best that can be done is the latest version of the Emergent model, which uses a "low-energy" solution of loop quantum gravity (LQG) to make the Einstein state stable against perturbations of a limited size (Figure 3.10). In response to the question, "Is the initial state metastable and therefore finite in its lifetime?," Ellis answers that the Einstein state can persist at most for a "long" but apparently *finite* time.[50]

LQG theorist Martin Bojowald explains that *any* perturbation, even if not of sufficient initial size to cause the system to escape the metastable potential, will *eventually* cause the system to escape it:

> Static solutions do not evolve, and so are clearly ill-suited as a model for the Universe. *But by introducing a perturbation to a static solution, one can slightly change it and thereby start a more interesting history*. Unfortunately, the classical solution [ESS] is unstable: any disturbance grows rapidly, leaving little of the initial state behind. The insight of Mulryne and colleagues is that quantum effects could supply all the necessary ingredients where classical solutions do not. Within the framework of loop quantum gravity, repulsion also implies static solutions at small size, but these – in contrast to the classical case – are stable. *According to the authors' model, perturbing such a state leads to small cycles of interchanging expansion and contraction.* During this process, matter will evolve slowly, and the cycles will gradually change their behavior. By itself, this perpetual recurrence and incremental change seems to lack the spark necessary for so momentous an event as the birth of the Universe. And indeed, Mulryne and colleagues identify one final theoretical ingredient that lights this spark: mediated through repulsive effects, *potential energy is gradually pushed into the matter during its slow evolution. At the point when potential energy starts to dominate kinetic energy, the mundane cycling is broken by a sudden, dramatic inflationary explosion – the emergent Universe*. (Bojowald 2005, pp. 920–1; emphasis added)

---

[49] We note, as well, that a perturbation to a near-ESS state should be just as effective at disrupting the universe as a perturbation to a genuine ESS. Hence, a model which is only past asymptotic ESS does not escape the problem. In fact, given past infinite time, and the variety of exotic quantum universe transitions postulated throughout the cosmological literature, it seems inconceivable that any universe could possibly maintain a conserved structure over time periods "long" compared with the interval since the Big Bang.

[50] He says, "note the later version of our model (astro-ph/0502589) based in the semi-classical approximation to loop quantum gravity where the static model is stable for a long while" (Private communication, January 24, 2006).

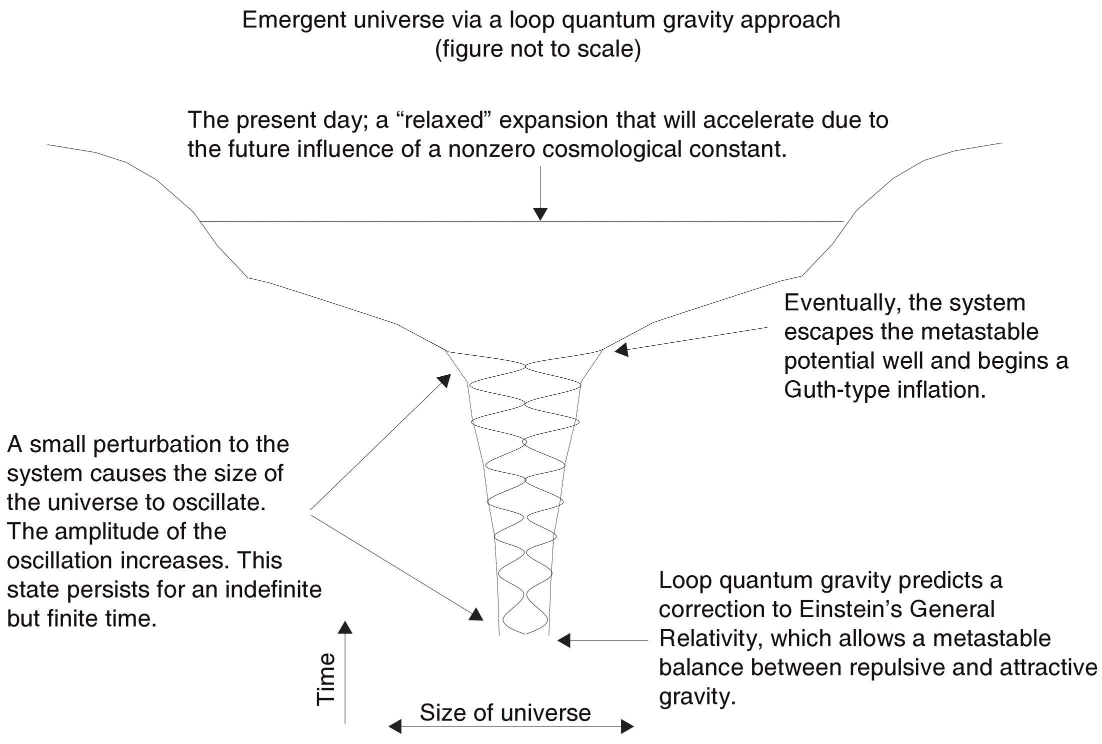


Figure 3.10 Evolution of an Emergent universe from a metastable loop quantum gravity state.

Metastability implies a finite life for the associated state. Either something must have come before it or it was "created." This issue of metastability is a general problem across a wide array of model classes. Vaas elaborates:

> Meta-stable states have a local, but not a global minimum in their potential landscape and, hence, they can decay; ground states might also change due to quantum uncertainty, *i.e.*, due to local tunneling events. Some still speculative theories of quantum gravity permit the assumption of such a global, macroscopically time-less ground state (*e.g.*, quantum or string vacuum, spin networks, twistors). *Due to accidental fluctuations, which exceed a certain threshold value, universes can emerge out of that state.* Due to some also speculative physical mechanism (like cosmic inflation) they acquire – and, thus, are characterized by – directed non-equilibrium dynamics, specific initial conditions, and, hence, an arrow of time. (Vaas 2004, p. 10; emphasis added)

It therefore seems that metastable (and, for that matter, unstable) states must have a merely finite lifetime. Metastable states leave unexplained how they came to exist. Universes with a metastable initial state must therefore have a beginning, consistent with the second premise of the *kalam* cosmological argument.

IIIc. Cyclic universe

According to these models, the universe goes through a cycle in which it grows from zero (or near-zero) size to a maximum and then contracts back to its starting condition. The universe itself is periodic, in the sense that it undergoes many such cycles, perhaps an infinite number. The average expansion of the universe would be zero in a "pure" cyclic model since cycle by cycle, one always experiences precisely equal amounts of expansion and contraction. Hence, a cyclic model evades the BVG Theorem. The past is featureless. Unlike the previous two model classes, it is not the case that the universe asymptotically approaches some particular state in the infinite past.

As Vilenkin indicates, however, cyclic models face a thermodynamic problem: "A truly cyclic universe has a problem with entropy increase: it should have reached thermodynamic equilibrium by now" (pers. comm., January 19, 2007). Our observation of the present day universe indicates that we are not at a condition of thermodynamic equilibrium – a good thing for us, as life requires nonequilibrium conditions to exist! As one looks into the past, the size of each cycle is also thought to decrease (due to radiation effect on entropy). Eventually the cycles are so small that one ends up with a different physics – which would preclude the cycling and imply a beginning to the universe.

So how does one overcome this problem? Paul Frampton and Lauris Baum have recently proposed an ingenious mechanism that breaks genuinely new ground in cosmological studies. It is surprising that they base their model on a scenario that is generally thought to imply quite the opposite of cycling. They assume that a type of dark energy pervades the universe where its equation of state (the ratio between pressure and energy density) is less than −1. This would be different from the cosmological constant mentioned earlier (equation of state equal to −1). This type of expansion is thought to lead to an event called the Big Rip. Dark energy (also called phantom energy in this context) causes the acceleration in the expansion of the universe to become so great that our visible horizon shrinks over time. Eventually, this causal horizon shrinks so much that cosmological objects of smaller and smaller size become causally unbound. Galaxies, solar systems, planets, and, eventually, even atoms get ripped apart as the expansion rate of the universe tends toward infinity. This would stop at a spatial singularity in the finite future. Baum and Frampton propose a "mosaic" model[51] to overcome the problem of entropy buildup in a single universe:

> We consider a model where, as we approach the rip, expansion stops due to a brane contribution just short of the big rip and there is a turnaround time $t = t_T$ when the scale factor is deflated to a very tiny fraction (f) of itself and only one causal patch is retained, while the other $1/f^3$ patches contract independently to separate universes. Turnaround takes place an extremely short time ($<10^{-27}$ s) before the big rip would have occurred, at a time when the Universe is fractionated into many independent causal patches. (Baum & Frampton 2007, p. 1)

What happens in the Baum–Frampton approach is that very close to the Big Rip event, the universe splits into noninteracting (causally disconnected) patches. The universe has expanded so much at this point that nearly all of these patches are empty of (normal) matter and radiation. They contain only phantom[52] energy. It turns out that the entropy content of the universe (which is what interferes with cycling) is contained within the thinly spread matter and radiation. Those patches that contain only phantom energy are supposed never to undergo the Big Rip. Instead they separately undergo a *deflation* event; contracting

---

[51] Mosaic model: undesirable features of a model universe may be regional in scope; consideration of a multiverse may remove those features.

[52] Phantom energy: dark energy with a supernegative equation of state, that is, $p/\rho < -1$.

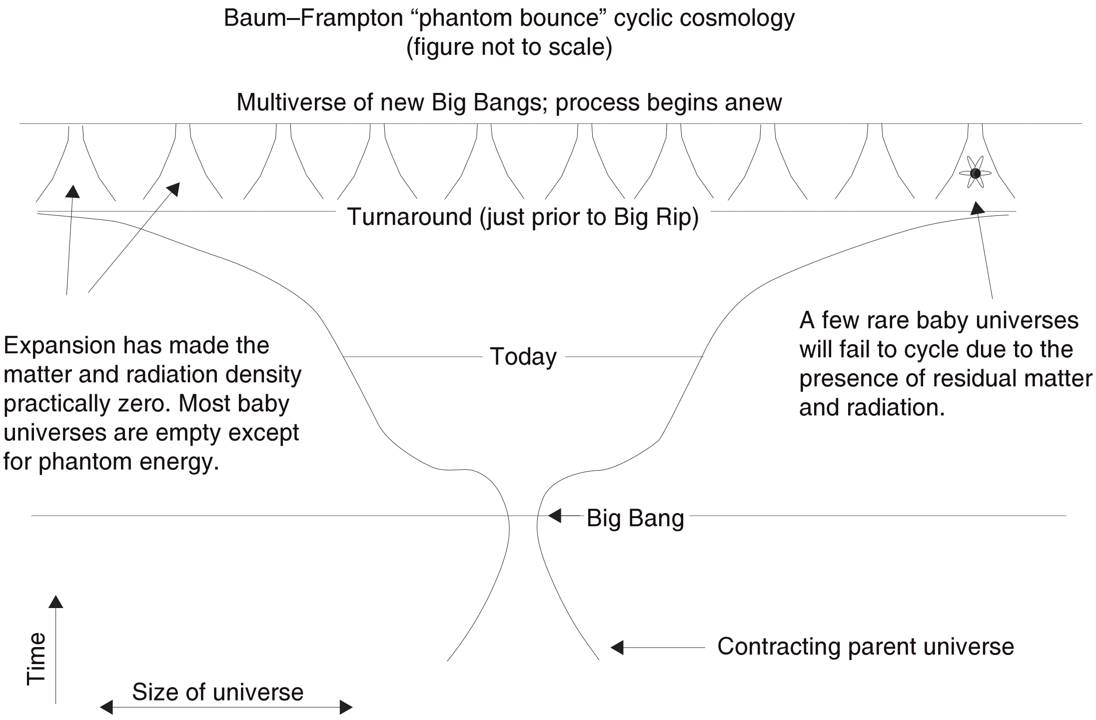


Figure 3.11 Baum–Frampton phantom bounce model.

an amount exactly equal to the expansion that the universe experienced since the Big Bang (thereby avoiding the implications of the BVG Theorem). Prior to reaching a singularity, the contracting patch rebounds due to the effects of phantom energy. It then repeats the same cycle endlessly. Every patch that undergoes this fractionates into new "universes" that themselves begin to propagate (think of dandelions spreading over your lawn). Hence, the Baum–Frampton model is said to feature an infinite multiverse in addition to a beginningless cyclic behavior (Figure 3.11). But is the model viable?

Several challenges remain to be addressed if the model is to be a viable option. First, in order to avoid the BVG singularity theorem, the average contraction must equal exactly the average expansion (for every geodesic). But how is this to be done without introducing explicit fine-tuning? There is no reason that deflation of scale factor will exactly match post–Big Bang expansion. Frampton admits:

> I have no idea why it [the BVG constraint] is satisfied because it does relate expansion with dark matter to contraction without dark matter. . . . I suspect it is not fine tuning but that may ultimately depend on one's perspective: if BVG were not satisfied, infinite cyclicity is impossible. (pers. comm., February 5, 2007)

The problem has not yet been addressed.[53]

---

[53] In June 2007, Paul Frampton did put a paper on the Web preprint (Frampton 2007a) that partially addressed the issue. But it appears to show only that a generic cyclic model avoids the BVG Theorem. That point was not in dispute. The original question still seems to persist.

Second, globally, entropy should have already grown to an infinite value. How is it that the various regions remain causally disconnected? The key factor in this model is the method for jettisoning the universe's entropy. As Baum and Frampton emphasize, if matter is retained during a contraction phase:

> . . . presence of dust or matter would require that our universe go in reverse through several phase transitions (recombination, QCD and electroweak to name a few) which would violate the second law of thermodynamics. We thus require that our universe comes back empty! (Baum & Frampton 2007, p. 4)

Now, globally, over infinite past time, the model posits that an infinite amount of matter and radiation (hence, infinite entropy) has been produced. How is it, then, that the entropy density avoids achieving an infinite quantity? Frampton responds:

> It is true that if we retain all the separate patches entropy continues to increase. Our key idea is to retain only one causal patch for our universe whose entropy thereby drops precipitously in what we for obvious reasons call deflation. (pers. comm., February 5, 2007)

The sole mention of where this entropy goes occurs in the following passage: "The old problem confronting Tolman is avoided by removing entropy to an unobservable exterior region; one may say in hindsight that the problem lay in considering only one universe" (Frampton 2007a, p. 4).

Simply shoving the entropy into other "universes" raises the question whether, given infinite time, a static space and a countable infinity of realms within the multiverse, the realms must not eventually collide. Frampton responds:

> . . . the 'causal patches' remain disjoint spawning separate universes which do not interact or collide again as they start out and remain causally disconnected. This more definite answer . . . results from our better understanding of the turnaround from subsequent technical calculations. (pers. comm., October 10, 2007)

In our initial communication (February 2007), Frampton indicated that the problem of collision was a valid one and would be investigated. These calculations have not yet been published, hence the issue remains troubling.

Cosmologist Xin Zhang, in a recent paper (Zhang 2007a), argues that the causal disconnection mechanism at turnaround does not work precisely because the disconnected patches do come back into causal contact. Frampton has pointed out a possible error in Zhang's critique (Frampton 2007b). This appears, however, to have been rectified in a new communication (Zhang 2007b). Zhang's (new) critique is the following: Frampton uses the following form of the modified Friedmann equation:[54]

$$H^2 = \frac{8\pi G}{3}\rho\left(1 - \frac{\rho}{\rho_c}\right)$$

---

[54] Phantom bounce models operate on the assumption that the standard Friedmann equation is no longer valid for universe conditions where the phantom energy density is near a critical value.

Here, "*H*" is the Hubble parameter, which is defined as the time derivative of the scale factor divided by the scale factor. Recall that the scale factor is the factor by which one multiplies the size of the universe in order to represent expansion or contraction.

Zhang clarifies:

> . . . ρ_c is the critical energy density set by quantum gravity, which is the maximal density of the universe. Such a modified Friedmann equation with a phantom energy component leads to a cyclic universe scenario in which the universe oscillates through a series of expansions and contractions. In the usual universe [that is one governed by the ordinary Friedmann equation], the phantom dark energy leads to a "big rip" singularity; however, in this peculiar cyclic universe, the big-rip singularity can be avoided because when ρ reaches ρ_c the universe will turn around due to [the modified Friedmann equation]. (Zhang 2007b)

So the rate of change of the size of the universe is governed by the density of this special type of "phantom" energy. Zhang continues:

> When the universe approaches the turnaround point ($\rho \to \rho_c$), we have $H \to 0$. Therefore, obviously, at the turnaround, we have $H^{-1} \to \infty$. This implies that the Hubble radius becomes infinity at the turnaround point, because at that time the universe becomes static instantaneously (that is, it ceases to expand at the turnaround). Obviously, at the turnaround, the universe would not be fragmented into many disconnected causal patches. (Zhang 2007b)

The inverse of the Hubble parameter (H^{-1}, called the Hubble radius) governs the scale at which microphysics can act, that is, the scale of causal connection. While this is near zero as one nears a "Big Rip" event, it seems clear that as one approaches Baum and Frampton's turnaround, the Hubble radius grows again to an infinite value. This means that all the separate patches of the universe are not disconnected (i.e. light signals can now propagate among them and allow them to interact). Thus, a subsequent collapse phase should include all of the matter and radiation then present, which essentially prevents a working cyclic scenario (since all the ordinary matter and radiation from the previous expansion would now be included). Frampton offers the following comment with respect to Zhang's criticism: "deflation must occur at a time before turnaround when the Hubble radius is small, before it reaches its minimum value ($x = 1/2$). Deflation remains a plausible conjecture which still requires further technical calculation to confirm" (pers. comm., Feb 7, 2008).

This seems difficult to square, as it would seem deflation = contraction, and turnaround is the time at which expansion ends and contraction begins. Baum and Frampton, in fact, issued a preprint in 2006 entitled "Deflation at Turnaround for Oscillatory Cosmology," in which they state:

> A key ingredient in our cyclic model is that at turnaround $t = t_T \pmod{\tau}$ our universe deflates dramatically with scale factor $a(t_T)$ shrinking to $\hat{a}(t_T) = fa(t_T)$ where $f < 10^{-28}$. This jettisoning of almost all, a fraction (1-f), of the accumulated entropy is permitted by the exceptional causal structure of the universe. (Baum & Frampton 2006, p. 4)[55]

Frampton clarifies:

> The beginning of contraction is the turnaround. Deflation is where the causal patches separate and our entropy drops to zero.

---

[55] Here τ indicates periodicy; it indicates which cycle the universe is presently in.

> By the way the time difference between deflation and later turnaround is a trillion trillionth of a second or less!!!
>
> It [Hubble radius divergence at turnaround] does have the significance that each spawned universe is separately one causal patch at turnaround. (pers. comm., February 7, 2008)

So deflation is not contraction. Instead it refers to causal disconnection. The Hubble radius *does* diverge at turnaround. But for some reason the causal horizons frozen in at deflation remain intact. This seems problematic. After all, causal horizons are not *real* physical barriers. They are observer dependent. For example, we on Earth have a causal horizon that stretches out some 46 billion light years. So does a space traveler in orbit around Alpha Centauri. But the space traveler sees a different portion of the universe than we do here on Earth. There is not actually a physical barrier 46 billion light years away from each of us. This is distinct from the event horizon of a black hole, which *is* an objective physical barrier. How exactly do Baum and Frampton understand this causal disconnection? A study of Frampton's early work could perhaps clarify the situation:

> . . . the time when a system becomes gravitationally unbound corresponds approximately to the time when the growing dark energy density matches the mean density of the bound system. For a "typical" object like the Earth (or a hydrogen atom where the mean density happens to be about the density of water $\rho_{H2O} = 1$ g/cm^3 since 10^{-24} g/(10^{-8} cm)$^3 = 1$ g/cm^3) water's density ρ_{H2O} is an unlikely but practical unit for cosmic density in the oscillatory universe.
>
> . . . the unimaginable dark energy density at turnaround of $\rho_\Lambda(tT) > 10^{27}\rho_{H2O}$. By the time the dark energy density reaches such a value, according to the Big Rip analysis,[56] the smallest known bound systems of particles have become unbound. Additionally the constituents will be causally disconnected, meaning that if *the expansion had, instead, continued to the Big Rip* the particles could no longer causally communicate. (Baum & Frampton 2006, pp. 3–4; emphasis added)

This is the key. If the density of the phantom energy is $\rho_\Lambda(tT) > 10^{27}\rho_{H2O}$, then *in a Big Rip scenario*,[57] the universe would be causally unbound. But we are not in a Big Rip scenario. Instead, according to Baum and Frampton, the laws of physics have changed (we now have a *modified* Friedmann equation) where *something* (perhaps extradimensional brane dynamics) acts to stop the expansion and leads to a contraction. In this case, it is hard to see why Frampton and Takahashi's analysis of causally unbound systems still applies. It is the runaway expansion in the Big Rip scenario that leads to and *maintains* shrinking causal horizons. If that stops, and even reverses, then it would appear reasonable to assume that the causal horizon grows along with it.

Frampton and Takahashi's explanation seems akin to a personal causal horizon as opposed to some physical barrier that results in permanent causal disconnection. What is lacking is a discussion of what causes "turnaround" and maintains causal disconnection. Without those details, it seems reasonable to entertain Zhang's misgiving concerning the viability of the universe-fractionating mechanism.

We should note Frampton's emphasis on the limited time between deflation, turnaround, and contraction. That would make the relevant issue the behavior of phantom

---

[56] Frampton and Takahashi (2003, 2004).

[57] In a "Big Rip," the expansion rate of the universe becomes infinite and leads to a future singularity.

energy right at turnaround (given that there is no time for ordinary matter and radiation to reestablish contact, given the amount of contraction that occurs in the first fraction of a second after turnaround). The question seems to be: given the *homogeneity* of the phantom energy at turnaround, and its ability to interact with its surroundings (given the unbounded Hubble horizon right at turnaround), why would the universe split into separate domains rather than precipitate a single, global contraction?

Third, the presence of *any* matter or radiation (during contraction) will prevent cycling. This could be a problem, given that spontaneous structure can form as thermal fluctuations (even if the contraction stage begins without any matter or radiation). Bousso and Freivogel explain:

> In a long-lived vacuum with positive cosmological constant, structure can form in two ways. Structure can form in the conventional way (through a period of inflation followed by reheating), or it can form spontaneously as a rare thermal fluctuation. Because deSitter space is thermal, if the vacuum is sufficiently long-lived spontaneous structure formation will occur. (Bousso & Freivogel 2007, p. 4)

The Baum–Frampton space is not a de Sitter space, but it is also thermal. Hence, one would expect that matter would still fluctuate into existence spontaneously. If so, then (at a reasonable probability) the Baum–Frampton cyclicity would not work.

Cosmologist Thomas Banks contends that a contracting space filled with quantum fields will have an "ergodic" property as the space shrinks. Its fields become highly excited as one approaches the end of contraction and these fields will produce chaotic fluctuations. Spontaneously created matter with a different equation of state will dominate the energy density. That, and the inhomogeneity of the fluctuations, will prevent cycling. Banks and Fischler even suggest that the fields will spontaneously produce a dense "fluid" of black holes leading to a condition they call a "Black Crunch" (Banks & Fischler 2002) for arbitrary states approaching full contraction.[58] Hence, it appears that the Baum–Frampton cyclicity will not work.[59]

---

[58] Banks complains:

> I have a problem with ALL cyclic cosmologies. . . . The collapsing phase of these models always have a time dependent Hamiltonian for the quantum field fluctuations around the classical background. Furthermore the classical backgrounds are becoming singular. This means that the field theories will be excited to higher and higher energy states (define energy in some adiabatic fashion during the era when the cosmology is still fairly slowly varying, and use this to classify the states, even though it is not conserved). High energy states in field theory have the ergodic property–they thermalize rapidly, in the sense that the system explores all of its states. Willy Fischler and I proposed that in this situation you would again tend to maximize the entropy. We called this a Black Crunch and suggested the equation of state of matter would again tend toward $p = \rho$. It seems silly to imagine that, even if this is followed by a re-expansion, that one would start that expansion with a low entropy initial state, or that one had any control over the initial state at all. (pers. comm., October 12, 2007)

[59] We note that Xin Zhang has his own competing cyclic model, in which he admits that thermal fluctuations pose a serious problem for phantom bounce cosmologies:

> It is noteworthy that the cyclic universe discussed in this paper is an ideal case, and there are still several severe obstacles existing in the cyclic cosmology, *such as the density fluctuation growth in the contraction phase*, black hole formation, and entropy increase, which can obstruct the realization of a truly cyclic cosmology. (Zhang, Zhang, & Liu 2007; emphasis added)

While phantom bounce cosmologies (such as Baum–Frampton and Xin Zhang's own model) do represent a frontier worth exploring, there seem to be unanswered questions as to the viability of such an approach. The field is too young to pass full judgment. But some questions that *can* be answered (such as the ergodic/chaotic approach to a singular bounce) seem to indicate that problems native to cyclic cosmologies remain.

IIId. A fourth alternative?: time deconstruction

As Borde et al. point out in their seminal paper, one of their primary assumptions was the following:

> The intuitive reason why de Sitter inflation cannot be past-eternal is that, in the full de Sitter space, exponential expansion is preceded by exponential contraction. Such a contracting phase is not part of standard inflationary models, and does not appear to be consistent with the physics of inflation. If thermalized regions were able to form all the way to past infinity in the contracting spacetime, the whole universe would have been thermalized before inflationary expansion could begin. In our analysis we will exclude the possibility of such a contracting phase by considering spacetimes for which the past region obeys an averaged expansion condition, by which we mean that the average expansion rate in the past is greater than zero: $H_{av} > 0$. (Borde, Guth, & Vilenkin 2003, p. 1)

In his 2003 lecture at the Kavli Institute at UCSB, Guth acknowledges, "[Anthony] Aguirre and [Steve] Gratton have proposed a model that evades our theorem, in which the arrow of time reverses at the $t = -$infinity hypersurface, so the universe 'expands' in both halves of the full de Sitter space."[60] It is possible, then, to evade the BVG Theorem through a gross deconstruction of the notion of time. Suppose one asserts that in the past contracting phase the direction of time is reversed. Time then flows in both directions *away* from the singularity. Is this reasonable? We suggest *not*, for the Aguirre–Gratton scenario (Aguirre & Gratton 2002) denies the evolutionary continuity of the universe which is topologically prior to *t* and our universe. *The other side of the de Sitter space is not our past.* For the moments of that time are not earlier than *t* or any of the moments later than *t* in our universe. There is no connection or temporal relation whatsoever of our universe to that other reality. Efforts to deconstruct time thus fundamentally reject the evolutionary paradigm.

Section III summary

Primarily, attempts to overcome the new singularity theorem of Borde, Vilenkin, and Guth center on generating universe models that do not feature average positive expansion in their past. This can be done by having average negative expansion (i.e. contraction) or by having zero average expansion (an asymptotically static model). Both attempts seem to encounter insurmountable difficulties. The contraction model features acausal fine-tuning to its asymptotic past and BKL chaos as the contraction nears a singularity in its pre–Big

---

[60] Alan Guth, speech to the Kavli Institute for Theoretical Physics, October 2003. Available at: http://online.kitp.ucsb.edu/online/strings_c03/guth/.

Bang "bounce" into our present expanding reality. BKL chaos may, in fact, prove that it is impossible for a universe to pass through a singularity.

Zero average expansion models are usually constructed in two different ways. Either the expansion asymptotically approaches zero as time (looking backward) approaches negative infinity. One can also consider an infinite number of cycles where expansion and contraction exactly cancel for each cycle. The first case has the dilemma that it must begin static and then transition to an expansion. Hence, the static phase is metastable, which implies that it is finite in lifetime. The universe begins to exist.

Cyclic models usually fail due to the necessary buildup of entropy from cycle to cycle. If there were an infinite number of cycles, why is the universe not in a state of "heat death"? The entropy would also have the effect of making the amplitude of the cycles (the maximum size of the universe during a cycle) grow over time. Looking backward in time, then, this implies a first cycle to the finite past. One can attempt to overcome these problems, as Baum and Frampton do, by claiming that a single universe fractionates into a multiverse of contractions. Most of the contracting "children" from the parent universe will have shed the entropy developed from the preceding phase and hence permit cycling.

But as we have seen, their mechanism appears to fail to fractionate. Even if it did fractionate into separate entropy-free domains, even an initially empty daughter universe would develop BKL chaos when approaching a contraction and disrupt the scenario!

The last gambit, that of claiming that time reverses its arrow prior to the Big Bang, fails because the other side of the Big Bang is *not* the past of our universe.

Hence, these models either have a beginning or are not viable.

IV. Quantum gravity

A final, expected exception to the Hawking–Penrose theorems is a formulation known as "quantum gravity." The two great pillars of twentieth-century physics, Einstein's GR (the science of the very large) and QM (the science of the very small), both enjoy overwhelming observational support. However, if the standard Big Bang theory were correct in its prediction that the universe must approach a singularity in its distant past, then QM must eventually govern it. But GR is a classical theory, not a quantum field theory. For an extremely high-density early universe, where gravity acted both as the dominant force and as a quantum force, we must have a new theory – quantum gravity – to describe it. Further, in 2002,[61] the first observational evidence that gravity is indeed a quantum force was reported. One of the Hawking–Penrose assumptions for their singularity theorems, however, was that GR is a correct description of the universe. Since this is not true at the scale of singularities, perhaps singularities did not exist after all. Perhaps, extrapolating backward, the universe evolves smoothly *through* the Big Bang to an unknown past.

Here are three prominent candidates for a theory of quantum gravity (Figure 3.12). The main job of some quantum gravity models is to get one through the singularity to an (it is hoped) eternal past. Others will accept a "beginning" to the universe but will deconstruct the notions of time, nothingness, or causation.

---

[61] Nesvizhevsky et al. (2001), see http://www.newscientist.com/article.ns?id = dn1801, or http://physicsworld.com/cws/article/news/3525. The experiment was done by bouncing supercold neutrons and noticing that the height of the bounce was quantized.

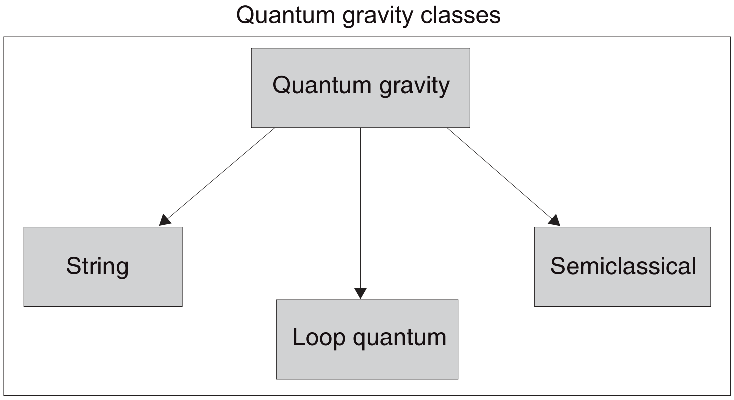


Figure 3.12 Families of quantum gravity cosmologies.

IVa. String models

String theory is by far the most popular method proposed so far to unify quantum theory with GR. Essentially, string theory proposes that the elementary entities of nature are not point particles (zero-dimensional objects) but are strings (one-dimensional objects). String theory eliminates many of the problems that occur in particle theories. Particle interactions can occur down to literally zero distance; where force laws blow up (recall, for example, that gravity's dependence on distance is $1/r^2$; that is, the force becomes infinite as range goes to zero) and predict infinite (i.e. nonsensical) answers. String theory calms this behavior by introducing a minimum distance (the "Planck" distance) to interactions. By "spreading out" string interactions, infinities are avoided.

Another advantage of string theory is that it can explain in a non–*ad hoc* manner the existence of different types of elementary "particles." Differing "particle" properties could be merely different types of vibrations that occur on a string (similar to musical notes). It was hoped (and still is by some) that the theory would naturally predict the characteristics of the elementary particles that are otherwise free parameters in the earlier theory (which is called the "Standard model").

The "minimum distance" feature of string theory is thought to be a desirable feature for cosmological models because, analogously to the case of particle interactions, the standard Big Bang model predicts that the scale factor of the universe shrinks to literally zero size. String theory could "calm" this feature of the model by suggesting a minimum size. This could even overcome the Hawking–Penrose theorems and suggest that there was a "before" to the singular condition. Perhaps the Big Bang was not an ultramundane event at which time itself came into being. If it was not, the door is open to a past eternal universe.

String theory has also given birth to "brane" cosmology, where the emphasis is on the background within which strings propagate rather than the strings themselves (Figure 3.13). These backgrounds, called *n*-branes, can be variable in the number of their dimensions. For example, it is normally proposed that our three-dimensional space is one of these "three-branes" which may or may not be floating around in a higher-dimensional space (usually called "the bulk").

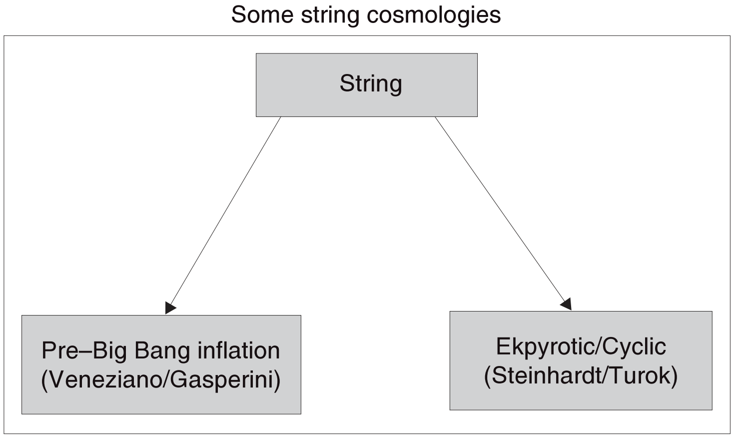


Figure 3.13 String cosmology models and proponents.

IVa(i). Pre–Big Bang inflation (PBBI)

There are two string models that can aspire to describe adequately a past infinite pre–Big Bang environment. These are the PBBI scenario of Gabriele Veneziano and Maurizio Gasperini and the Ekpyrotic/cyclic model of Paul Steinhardt and Neil Turok. The PBBI scenario is a classic version of an asymptotically static model. Here is how one of its authors describes it:

> According to the scenario, the pre-bang universe was almost a perfect mirror image of the post-bang one. If the universe is eternal into the future, its contents thinning to a meager gruel, it is also eternal into the past. Infinitely long ago it was nearly empty, filled only with a tenuous, widely dispersed, chaotic gas of radiation and matter. The forces of nature, controlled by the dilaton field, were so feeble that particles in this gas barely interacted. As time went on, the forces gained in strength and pulled matter together. (Veneziano 2004, p. 63)

Through gravitational contraction, regions of the pre–Big Bang universe turned into black holes. Due to quantum effects, once the density of a black hole reached a critical value, it underwent a "bounce" into a Big Bang. Our universe then persists within this otherwise closed surface (with respect to the 'outer' background space where everything began).

In his popularization in *Scientific American*, Veneziano seems to suggest that his beginning is an infinitely distant but never reachable (i.e. ideal) point. The article implies that the model is to be interpreted realistically.[62] But one must be wary in interpreting the infinity of the past for this model, as it is easy to misuse the concept of an infinite limit, as well to ignore the distinction between a realist and an instrumentalist interpretation of the model. The problems in interpreting this model are similar to those encountered when assessing the Emergent model class. It is worth noting that the coauthor of the PBBI model, Maurizio Gasperini, indicates that the entire asymptotic past (or future) should not be taken as real:

---

[62] Or is it *Scientific American* that has inserted the realist interpretation?

> . . . I find it misleading to talk of a 'future of the PBB scenario,' because the PBB scenario only (possibly) applies to describe some (more or less extended) portion of the past history of our cosmos, and is expected to smoothly join the standard cosmological scenario at an early enough epoch, so as to reproduce standard results on nucleosynthesis, baryogenesis, structure formation, and so on. In other words, the PBB scenario can be regarded as a model for explaining the initial conditions of our standard cosmological configuration, in a way which is string-theory consistent, but it cannot be extrapolated towards the future without further assumptions, which at present have not been fully worked out (with the exception of the dilaton model of dark energy proposed by Piazza, Veneziano and myself on PRD 65, 023508 (2002)). (pers. comm., January 10, 2006)

Can one build a realist interpretation of this model? It is interesting to contrast the depiction of PBBI as proving "the myth of the beginning of time" in a setting where sensational conclusions are encouraged (*Scientific American*) with the characterization of PBBI as a "toy model" in a setting where scientists are naturally conservative (peer-reviewed academic journals).[63]

As described in the academic literature, the model appears to have an initial phase. The relevant phases are:

(1) A static (Milne) universe, or string perturbative vacuum (SPV) phase. This means that the universe is empty (energy and energy density is zero) and is static, that is, neither expanding nor contracting globally or locally.
(2) A quasi-Milne phase, which constitutes a "perturbed" SPV. Here "H is (small and) positive in the String frame, (small in modulus and) negative in the Einstein frame, and tends to zero as t goes to minus infinity, and the Universe approaches the SPV configuration (where H is identically zero, since the spacetime is flat)" (M. Gasperini, pers. comm., January 16, 2007).
(3) An "inflationary" phase. In one set of coordinates (the Einstein frame), matter collapses into trapped surfaces, or black holes. In another set of coordinates (the string frame), this can be viewed as a spatial expansion. This happens regionally rather than globally.
(4) A post–Big Bang FRW phase that is typical of the standard Hot Big Bang model.

The authors begin building the model at a finite time in the past where a condition called "asymptotic past triviality" (APT) obtains (Veneziano & Gasperini 2002, p. 54). APT represents the boundary between phase (2) and phase (3). The period of contraction (inflation in the string frame) is itself finite.

The authors then project this state into the future, and asymptotically evolve it into the past. Similar to the Maartens version of the Emergent model, one takes the APT past-directed duration of phase (2) to be infinite (Figure 3.14).

Unlike the Emergent model, perturbed SPV lasts for a while and then individual patches that meet the Hawking–Penrose condition for a closed trapped surface begin *regional*

---

[63] "The so-called 'pre-big bang' scenario described in this report has to be seen in the above perspective as a possible example, even just as a toy model of what cosmology can look like if we assume that the sought for standard model of gravity and cosmology is based on (some particular version of) superstring theory" (Veneziano & Gasperini 2002, p. 4).

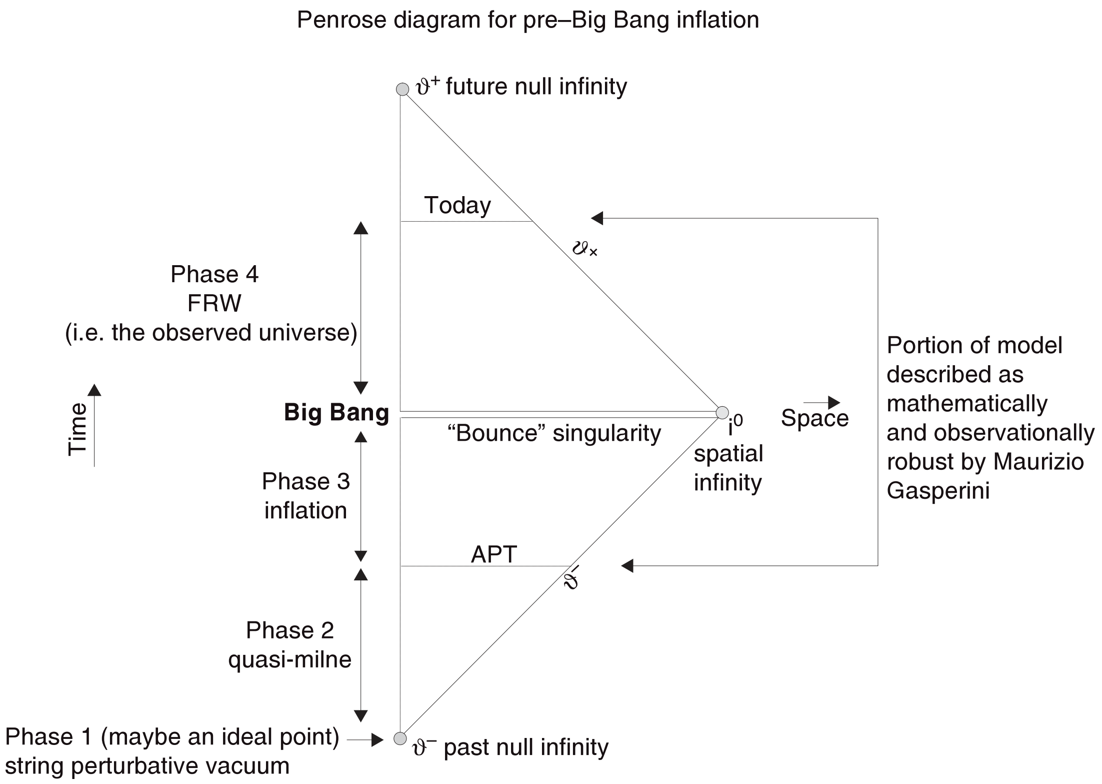


Figure 3.14 Construction of the pre–Big Bang inflation (PBBI) model is given at the "asymptotic past triviality" or APT point. From here, the model is projected forward in time to give the Big Bang universe. It is also projected backwards to past infinity. Note that the APT data is given at a finite time in the past in the diagram. The proposed string perturbative vacuum era would be at "past null infinity" or the lower point marked "ϑ^-." The past infinite is "null" rather than "timelike" because the gravidilaton waves that embed it are massless. FRW = Friedmann–Robertson–Walker.

contraction (inflation in string coordinates) (Figure 3.15). Should we treat SPV like ESS? Gasperini observes that a significant feature of SPV is that it is unstable:

> . . . the SPV is unstable already at the classical level . . . it [decay of the SPV] can be described as a quantum transition, but it is a process which is also classically allowed. (pers. comm., January 4, 2007)
>
> . . . the instability of the SPV is similar to the instability of a classical ball placed exactly at the top of a perfectly symmetric hill. In principle, if the system starts initially in the unique equilibrium configuration, and there are no external perturbations, the system might remain static forever. In practice, however, there are physical perturbations removing the system from equilibrium, sooner or later, with a probability which is chaotically (or randomly) distributed. In the case of the SPV the perturbations removing it from equilibrium are the quantum fluctuations of the background fields (in particular of the dilaton). In addition, the exact equilibrium configuration can only be achieved as an asymptotic extrapolation, in the limit in which the cosmic time goes to minus infinity: in practice, at any given finite physical time, the system is always displaced a bit from equilibrium. (pers. comm., January 9, 2007)

Figure 3.15 Rate of universe expansion versus time in the pre–Big Bang inflation model. (source: Maurizio Gasperini, available at: http://www.ba.infn.it/~gasperin/)

> I would say that the SPV is not a phase extended in time, it is only an asymptotic initial state which is approached, however, in an infinite time. In practice, the physical description never starts "exactly" from that state, but from a state which represents an arbitrarily small perturbation of the SPV. (pers. comm., February 27, 2007)
>
> . . . if I live in the initially collapsing portion of spacetime, then I have a chance to go through the bounce, at some epoch in the future. But this does not concern the entire spacetime. There are spacetime regions which are collapsing and eventually bouncing into a FRW like Universe, and others which do not. It is possible, in principle, to live in regions of spacetime never experiencing the bouncing nor the collapse, and staying for ever in a configuration well described by the string perturbative vacuum (or by a quantum perturbation of it). (pers. comm., January 9, 2006)

So if the SPV were real, it would be a state with a finite lifetime. Some time after decay, random portions will be sufficiently dense to form closed trapped surfaces and begin gravitational contraction. Other regions could remain indefinitely in the post-SPV state (i.e. already perturbed from an equilibrium condition).

Veneziano and Gasperini's language at times suggests treating SPV as an ultimately unrealized extrapolation similar to what is described in the *Scientific American* article. But here and elsewhere, they seem to suggest that the SPV is quite real (if the asymptotic past of the model were taken seriously):

> The whole process may be seen as the slow, but eventually explosive, decay of the string perturbative vacuum (the flat and interaction-free asymptotic initial state of the pre-big bang scenario), into a final, radiation-dominated state typical of standard cosmology. (Veneziano & Gasperini 2002, p. 21)

The PBBI model is described as a "decay" or a quantum "tunneling" event similar to the semiclassical models of Hawking–Hartle and Vilenkin (see section IVc) (Figure 3.16). But:

Publisher's Note:
Permission to reproduce this image online was not granted by the copyright holder. Readers are kindly requested to refer to the printed version of this chapter.

Figure 3.16 Pre–Big Bang inflation as a string quantum transition. The figure contrasts "creation from nothing" models such as the Hawking–Hartle no-boundary approach or the Vilenkin "tunneling from nothing" (see section IVc) with a string approach. Here, a prior state quantum tunnels into our current Friedmann–Robertson–Walker universe. WDW = Wheeler–DeWitt equation, which is the basis for the semiclassical cosmologies to be discussed in section IVc (source: Maurizio Gasperini, available at: http://www.ba.infn.it/~gasperin/)

> [PBBI] can also be interpreted as a tunneling process, not "from nothing", however [which is how the Hawking–Hartle and Vilenkin models are interpreted; see section IVc], but "from the string perturbative vacuum." (Veneziano & Gasperini 2002, p. 208)

It is reasonable to presume that if there is a physical transition from state "A" to state "B," and state "B" is ontologically real, then state "A" must be taken to be ontologically real.[64] Significant as well is that the state right after SPV decay (if SPV were a realized physical state) would obtain only a finite time ago to the past, given that Gasperini's description of the decay product is identical to the following:

> The generic regular solution thus approaches Milne as t [approaches negative infinity] but, *at any finite large* [negative time], *also contains small dilatonic (and gravitational-wave) perturbations* giving $0 < \Omega \ll 1$. As $t \to -\infty$, $\Omega \to 0$. As time goes forward, instead, Ω tends to grow until, at some critical time $-T_0$, Ω becomes $O(1)$, in some region of space. From that moment on, in that "lucky" patch, the metric starts to deviate from Milne and dilaton-driven inflation

---

[64] If state "A" is unreal, then no transition at all takes place but rather an absolute "coming into being" of state "B." On this interpretation, the universe began to exist.

> sets in, pushing Ω extremely close to 1 in that patch. (Veneziano 1998, p. 10; emphasis added)[65]

So if the SPV state were real, there could only be a finite timeline from it to the present. Only if the perturbation itself were unreal would the past eternal nature of the model survive. Is this realistic? As we look backward from the APT point, what would the significance be of a real quantum perturbation to a state which was arbitrarily close to SPV? Would not one have a situation that was virtually identical to simply starting with "real" SPV? This is essentially a "pseudobeginning" scenario.

Vaas classifies this cosmology as "soft bang/pseudobeginning" just as he does the Emergent models.[66] Although the behavior of perturbed SPV is ultimately different from ESS (or perturbed ESS), the similarity with respect to classical (and quantum) instability should be the controlling one. The two models are mathematically similar with respect to the asymptotic approach to an infinitely distant unstable state. Vaas argues that there is no obvious arrow of time and hence the asymptotic past is likely a mathematical artifact.

Perhaps the past infinite timeline is only *technical* in nature: the duration from the APT point extrapolated *backward* is infinite, but that is to be understood *instrumentally*. In response to the question, "if the SPV is unstable, and within a finite time quantum fluctuations of the dilaton will disturb its equilibrium, does that not imply that the past must be finite?" Gasperini answers:

> From a physical point of view, a spacetime manifold has an infinite (past) extension if its timelike or null geodesics can be (past) extended for infinite values of their affine parameter, without ending into a singularity. This property is satisfied by the past extension of the pre-big bang solutions, so, in this sense, they are past eternal. (pers. comm., January 17, 2007)

As Gasperini indicates, it is true that, *in a sense*, the solutions are past eternal. In pure GR, it is true that backward-directed geodesics will trace all the way to past infinity. In fact, in pure GR, even if one assumed a *real* SPV state, it would still be the case that, technically speaking, backward-directed geodesics are past eternal for the same reason that a real ESS is technically past eternal.

But the relevant question seems to be this: do all past-directed geodesics from a quasi-Milne state (or a quasi-ESS) intersect with a significant system perturbation, or would

---

[65] For clarification, a Milne universe is a special state where the critical density parameter Ω is exactly equal to zero; the universe is empty. Recall that the critical density parameter Ω determines if an FRW cosmology will have a closed, flat, or open geometry. It is closed if $\Omega < 1$, flat if $\Omega = 1$, and open if $\Omega > 1$. If the cosmological constant is zero, then a closed universe will recollapse, and a flat or open universe will be ever-expanding.

[66] Vaas elaborates:

> A related problem refers to the pre-big bang model (Veneziano & Gasperini 2003). Here the string vacuum – where a local collapse in the Einstein frame (which corresponds to a dilaton-driven inflation in the string frame) before the big bang occurs – is quite simple, homogeneous, almost empty and does not have an overall arrow of time. But, mathematically, the origin of the pre-big bang – or, to be more precise, any pre-big bang, for the model does also imply a multiverse scenario – traces back to a maximally simple, static condition only in the infinite past (principle of asymptotic past triviality). But this can also be interpreted just as a local predecessor of a big bang and not a feature characterizing the infinite string vacuum as a whole. (Vaas 2004, pp. 18–9)

at least one geodesic trace, undisturbed, to past infinity? Consider the following comparison:

A. Suppose that the SPV (or ESS) were viewed as ontologically real rather than just idealized points at past infinity. Within a finite time, a quantum perturbation would disturb the state and the resultant timeline to the present would also be finite. The model would not be past infinite.
B. Suppose, instead, that ESS or SPV are taken to be asymptotic ideal points. For analysis' sake, start at the present and look backward. Within a finite time, one is arbitrarily close to the ideal condition. Now consider any (new) quantum fluctuation that occurs to the quasi-SPV (or quasi-ESS) state while one is tracing the backward timeline. The probability of this is essentially 1.

What meaningful difference is there between the universe in case A and case B (at the "new" fluctuation point)? Looking backward, there will be an unbounded number of all types of fluctuations of all sizes, any one of which will arrest the supposed asymptotic development of the SPV (or ESS). The proposed past infinite extrapolation seems to be a mathematical artifact (or at least explanatorily vacuous). Even taken seriously (i.e. not as a "toy" model), the model does not predict that the past is infinite.

IVa(ii). Ekpyrotic/cyclic

The Ekpyrotic model is a cyclic model, but not in the old sense of a universe that undergoes an eternal periodic sequence of expansion and collapse. The Ekpyrotic model makes use of the extradimensional nature of string theory to propose that cycling occurs, but in a higher dimension. Authors Paul Steinhardt and Neil Turok view the new model as preferable to the post–Big Bang inflationary model because it has fewer *ad hoc* features. For example, (present-day) dark energy is an add-on to the earlier model but is a natural and necessary feature of their new cycling model. String theory allows entities called "branes" that could be representative of what we would otherwise call our three (spatial)-dimensional universe. String theory demands six extra dimensions of space in order to be self-consistent. It is thought that these extra dimensions are (usually) tightly curled up around the three macroscopic dimensions and hence usually unobservable. (Think of a soda straw versus a one-dimensional line. The circle formed by looking at the straw edge-on could be thought of as a second dimension curled around the first.)

The Ekpyrotic model proposes that one of these extra dimensions (the "bulk" dimension) is of macroscopic size. Within this extra dimension lie two three-branes, either of which could represent our universe. These three-branes periodically collide, just as if they were connected by a spring. When they do so, the energy of the collision is transferred to the branes (Figure 3.17).

This energy is converted into the matter (and radiation) that ultimately gravitates into galaxies. The rest of the normal Big Bang sequence follows (stars, planets, and so on). As the branes separate from each other, the branes themselves are always expanding. (There are versions of the model where the branes undergo limited contraction, but always expand, on net, with each cycle.) Eventually, stars burn out, the galaxies recede beyond each galaxy's visual horizon, and the universe enters a period of cold, burned-out cinders. Meanwhile, the branes cycle toward another collision.

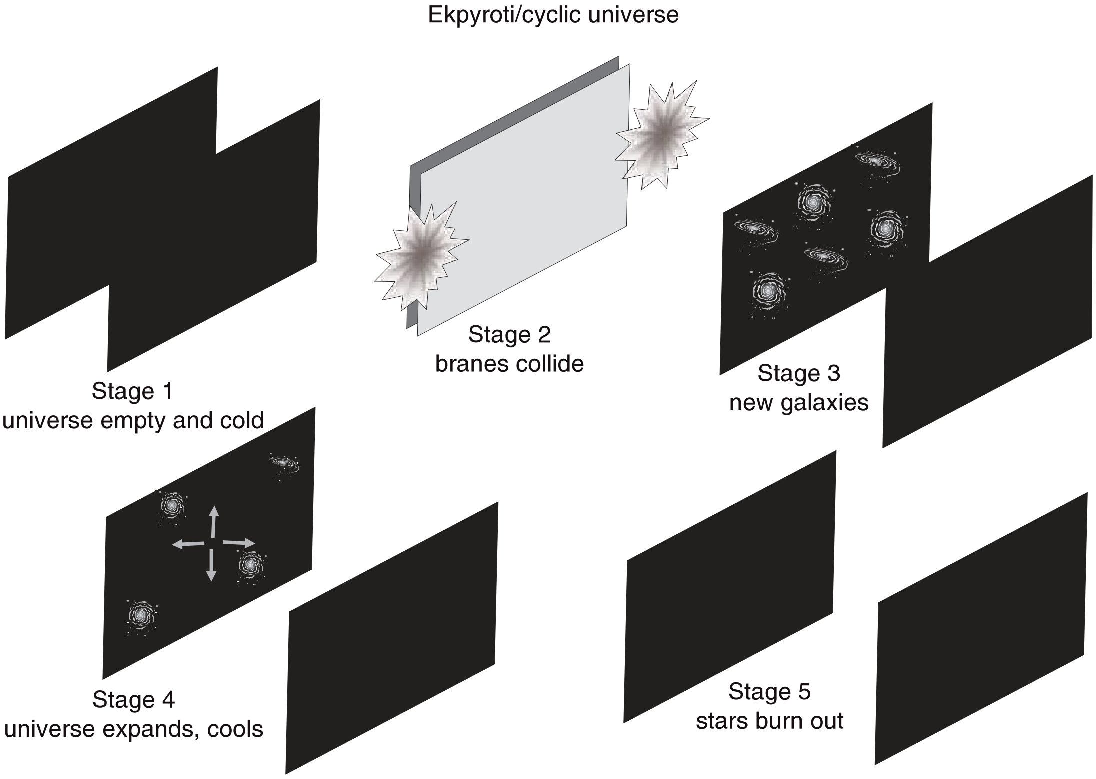


Figure 3.17 Pictorial description of the Ekpyrotic cycle.

The universe is then renewed with a new cycle. The energy that is released into the branes during each collision is replenished by an inexhaustible supply – gravitational potential energy. In this way, Ekpyrosis is an "open" system. There is always a limitless supply of free energy. This feature of cosmology as an open system (as opposed to the old "heat death" scenario) is new but widespread in contemporary cosmology.

Despite avoiding the "heat death" scenario, however, there is now a new problem to solve. If new energy is dumped onto the brane for infinite time, then either the entropy density or the energy density at each point must be infinite. This would obviously be contrary to observation. The continual expansion of the infinite-sized branes, however, keeps the entropy density constant. So (net) expansion is a critical model feature.

Steinhardt recognizes that his model of the universe is not truly beginningless. Here are the relevant comments from his Web site:

- Has the cyclic model been cycling forever?
- In principle, it is possible that the universe has undergone a semi-infinite number of cycles in its past during which the volume increases from cycle to cycle. Even though this would take an infinite time according to ordinary clocks, this cannot be the full story. This cycling regime would not cover all space-time. Something must have preceded the cycles.
- A similar issue arises in inflationary cosmology. In both cases, this is an open question. The issue is referred to as the problem of geodesic incompleteness referring to the fact

> that a purely expanding phase does not span the entire space-time and one has to consider what happened before.[67]

Steinhardt's website FAQ page indicates that the model is past geodesically incomplete. Here the authors comment in their published work:

> The most likely story is that cycling was preceded by some singular beginning. Consider a universe that settles into cycling beginning from some flat slice in the distant past many bounces ago. Any particles produced before cycling must travel through an exponentially large number of bounces, each of which is a caustic surface with a high density of matter and radiation at rest with respect to the flat spatial slices. Any particle attempting this trip will be scattered or annihilated and its information will be thermalized before reaching a present-day observer. Consequently, the observer is effectively insulated from what preceded the cycling phase, and there are no measurements that can be made to determine how many cycles have taken place. Even though the space is formally geodesically incomplete, it is as if, for all practical purposes, the universe has been cycling forever. (Steinhardt & Turok 2005, p. 5)

Steinhardt and Turok suggest the universe began in a singularity but that "for all practical purposes" it has been cycling forever. This claim is based on the fact that virtually no information as to the initial conditions of the universe could have survived to the present. Steinhardt explains that photons carrying this information would be "semi-infinitely redshifted" (pers. comm., January 27 and 30, 2004). There is a "semi-infinite" number of cycles between that boundary and the present. How are we to understand this? The description of the model as "de Sitter like" and the associated "semi-infinite" past timeline is nearly the same situation seen earlier with the discussion of the Misner "Mixmaster" universe (section IIIa). The key difference between Mixmaster and the Ekpyrotic model is that Mixmaster is intrinsically chaotic at its singularities (so not necessarily a good physical model of our universe), while the Ekpyrotic model avoids BKL chaos by having a positive equation of state (which is a unique feature of this cosmology) (pers. comm., January 17, 2006).[68]

Within the Ekpyrotic model an observer would see an infinite number of bounces with roughly a trillion years per bounce. So on ordinary clocks, past time is infinite (pers. comm., January 27 and 30, 2004). Yet there is clearly a past boundary (an infinitely distant beginning point?) preceding this behavior. One should note that the reason this model has a beginning is precisely because it falls under the previously mentioned BVG Theorem. Borde et al. explicitly apply their theorem to the Ekpyrotic model of Steinhardt and Turok (Borde, Guth, & Vilenkin 2003).

That implies, among other things, that the "past boundary" must be reached in a finite amount of time. So Steinhardt seems to be mistaken in his prediction of a semi-infinite number of cycles, unlike the Emergent and PBBI models, which evade the BVG Theorem (in proper time; i.e. if "an ordinary clock" is the most appropriate measure of physical time).

---

[67] Cosmologist Paul Steinhardt's internet site; FAQ section for the Ekpyrotic model http://wwwphy.princeton.edu/~steinh/cyclicFAQS/index.html#eternal.

[68] Recall that the equation of state is negative (for a cosmological constant pressure = minus density) for a typical inflationary model. For Ekpyrosis, the collision singularity occurs in the "bulk" dimension between the two branes; not within the brane that we ourselves would live on.

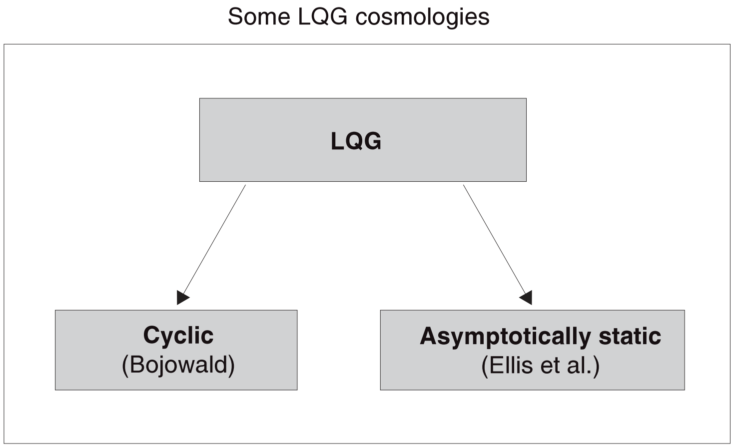


Figure 3.18 Candidate loop quantum gravity (LQG) cosmologies.

Steinhardt indicated to us that measuring the distance between the present and the beginning is ambiguous. What is not ambiguous, however, is that there is a boundary. And are the infinite cycles a necessary feature of the theory? In personal communication, he recognized that the theory does not require an infinite number of cycles.

Steinhardt, similar to Maurizio Gasperini, suggested to us a pragmatic view of his model. The model's description of the approach to the past boundary (and the boundary) is not (yet) rigorous. The boundary referenced by the BVG Theorem could be the mark of something fundamental, such as a singularity, or instead be the marker for a transition to different physics. If one asserts that the boundary merely marks the transition to different physics, then the issue as to the age of the precycling configuration asserts itself. Significantly, Steinhardt argues the cyclic solution is a dynamic attractor (this *is* rigorous) (pers. comm., January 27 and 30, 2004). What this means is that if one sets up the system with generic initial conditions within the twin-brane configuration, it will automatically, within a finite time, converge to the cyclic solution. Thus, the precycling configuration would have a finite lifetime. Thus, there is still an unexplained origin. So the Ekpyrotic universe (whether the boundary represents a genuine singularity of not) can safely be said to begin to exist.[69]

IVb. Loop Quantum Gravity (LQG)

Another theory of quantum gravity is the loop quantum approach (Figure 3.18). LQG takes the view that space-time itself is quantized; that is to say, space-time is divided into discrete constituent parts. It is a theory that aims to fill in the gap in standard GR by answering the question, what really happens at a singularity?

According to LQG, singularities do not really exist. As in string theory, there is a minimum size to nature that prevents microscopic mathematical infinities. Hence, time and space do

---

[69] Rüdiger Vaas has suggested that, while the cycling in the Ekpyrotic model had an origin, the brane components themselves could be past eternal. As we have seen, however, an explanation is necessary as to how they got into their initial noncycling state a finite time to the past. Vaas does not provide one (Vaas 2004, p. 17).

not come to an end as one comes to a "singularity." There will be a past timeline. This leads to the conclusion of an asymptotically static past, or a true cyclic past.

Martin Bojowald is the foremost exponent of this approach. We may think of Bojowald's model as a variation on the old Tolman cyclic model. There is only one universe. There are only three spatial dimensions. There is no "free" energy injected into the situation (such as there is in inflationary or Ekpyrotic scenarios). The Tolman model had two problems which prevented its wide acceptance: (1) there is no known physical mechanism for producing a cyclic "bounce" and (2) thermodynamic considerations show that the universe of the present day should have achieved thermodynamic equilibrium ("heat death"). This suggests that the past is finite. Bojowald recognizes both problems and believes that he can solve them. His basic approach is the same to both issues; the problems will turn out to vanish upon the generalization of current "classical" theory into LQG.

With regard to the first problem, the major difficulty has been resolving a type of chaos predicted to occur near classical singularities[70]. This chaos, named "BKL" after its discoverers (Belinsky, Khalatnikov, and Lifshitz) has been shown to be "calmed" by a loop quantum approach. As of 2007, some loop quantum theorists have been able to show that, for certain idealized models, a transition through a Big Bang condition is feasible (Ashtekar, Pawlowski, & Singh 2006). So, while a generalized proof is still lacking, the project seems promising.

The second condition is more daunting. How can there be truly cyclic behavior (one cycle looks pretty much like the last one, although there is not an event-by-event recurrence) when the second law of thermodynamics predicts that entropy must increase from cycle to cycle? Using a semiclassical approach to calculate entropy, Penrose finds that the end of our current cycle (the "Big Crunch") should differ in entropy from the Big Bang singularity by the stupendous factor of 10^{22} (Penrose 2005, p. 730).[71] Given no energy input from outside (and Bojowald argues that the system is truly closed), how can this outcome be avoided? There seem to be three possibilities:

1. The problem is epistemic only. In a 2004 paper, Bojowald and his colleagues appear to favor this solution:

> While the effective dynamics is consistent with our expectations for both the beginning and the end of the universe, the apparent time reversal asymmetry remains. This is explained by the fact that the situation is, in fact, time asymmetric due to our own position in the universe. We can see only some part of it, not the whole space-time, and in particular we see only a small part of the beginning. With the current understanding, the observable part of our universe can well be part of a classical space-time with a very inhomogeneous initial singularity.

---

[70] This is the same type of chaos that Ellis mentioned as an obstacle to infinite contraction models in section IIIa and is related to the problem mentioned by Banks for oscillating models (section IIIb).

[71] Penrose considers that there are 10^{80} baryons in the observable universe. He then suggests that the maximum entropy for the universe is equivalent to a black hole with this mass. Should the fate of the universe be to ultimately collapse in a "Big Crunch," this would be the entropy contributed by these 10^{80} baryons. Penrose uses the Hawking–Bekenstein formula for the entropy of a black hole, and in natural units (where constants of nature such as the speed of light are set to unity), finds that this entropy is approximately 10^{123}. The entropy of our universe in the current day is far lower than this by about 22 orders of magnitude.

> Since most of the initial singularity is unobservable, however, it is not discussed further. The final singularity, on the other hand, is completely unobservable until it is reached. If we compare only observable properties from within the universe, we simply cannot possibly know enough to tell whether past and future singularities are similar. If we compare the theoretical structure of a space-time from outside, then we conclude that in fact there is no conceptual difference between the beginning and the end of a generic spacetime.
>
> Only if we compare the observable part of the initial singularity with the theoretical expectation for a final singularity does the time asymmetry appear. (Bojowald & Hossain 2004, p. 38)[72]

2. Our current "classical" understanding of entropy is misleading. Bojowald suggested this possibility in personal communication:

> The interpretation of entropy is as a measure for the lack of information an outside observer can obtain if he just knows macroscopic parameters such as the total mass or angular momentum.
>
> The situation in non-equilibrium thermodynamics is more complicated, which would be relevant, *e.g.,* for colliding black holes or violent stages of cosmology. For cosmology, it is also not so clear what the total entropy should be associated with mathematically, so a counting is difficult. For black holes, on the other hand, entropy refers to degrees of freedom in the black hole region, which can be identified and counted within the theory.
>
> Black hole entropy then describes the lack of information in classical stages of black holes (how many quantum states there are for given mass and other parameters). This is an absolute lack of knowledge in classical gravity because the black hole region is concealed from outside observers.
>
> With quantum theory, however, black holes evaporate and thus reveal information at later stages (although it is still disputed to what degree this is realized). *The lack of information is then only temporary and apparent because an outside observer is simply not patient enough to wait until he can recover all information. In other words, entropy in this context is observer dependent and not an absolute quantity.* Since it includes only the black hole but not the observer or anything outside, it is also not the entropy relevant for cosmology.
>
> The usual intuitive picture in cosmology is as follows: When there are many black holes, this apparent entropy is very high for outside observers. *But if all degrees of freedom are considered, including those in the black holes which will re-emerge after evaporation, or one waits until after the black holes have evaporated one would obtain a smaller amount. This does not mean that entropy decreases; the accessibility of information by observers just changes.* (pers. comm., February 28, 2006; emphasis added)

3. Cycle by cycle, the entropy state is genuinely reversible. Again, this alternative emerges in personal communication:

> **Sinclair**: 'What it sounds to me you are saying in this last communication is that (with regard to the cyclic model) the same energy is endlessly recycled. This is a closed system, unlike some other cosmological proposals out there. Hence you are talking about a system that is fully reversible. This isn't a case where there are dissipative, irreversible processes that build up over time and produce the "heat death" scenario.'

---

[72] Note that the terms "initial" singularity and "final" singularity refer to the states that begin and end our current cosmological cycle. Bojowald *et al*. are not referring to the beginning and end of time.

> **Bojowald**: 'At least in the cyclic version. If there is no recollapse at large volume, the universe would just have gone through a single bounce and will keep expanding. The end may be such a heat death, but since we don't know the field content of our universe (as evidenced by the dark matter and dark energy puzzles) the far future may be quite different from what it appears to be now'. (pers. comm., February 28, 2006)

It is important to note that Bojowald (and his colleagues) are not committed to a model with a past-infinite number of cycles, as his last response shows. He is open to the possibility of an irreversible rise in entropy as a function of time. So the fact that entropy rises, cycle by cycle, and would trip up a proposed past infinite cyclic model is not, *per se*, a test of the viability of the loop quantum approach as a candidate for quantum gravity. Instead our interest (in this section) is limited to beginningless cyclic models.

Considering the first solution, Bojowald's reply to Penrose would be that there is a large, unobservable part of the initial singularity that is a genuine generic manifold, that is, a state of maximum entropy featuring random inhomogeneity and anisotropy. Hence, the entropy of the initial and final singularities would be similar. An inflation mechanism (of a small patch of this manifold) would then produce the requisite homogeneity and isotropy of the current FRW universe. Penrose, however, had anticipated this objection. Using an anthropic observer selection argument, he argues that the size of the inflationary patch we should expect to see should be much smaller based on thermodynamic criteria (by the factor $10^{(10\wedge 123)}$) (Penrose 2005, p. 763).

Penrose suggests that life might need a universe only 1/10 the size of our current (visible) universe. He obtains the probability of an appropriately sized initial patch of a generic manifold using the Hawking–Bekenstein equation for the entropy of a black hole. The exponent "123" in Penrose's formula is based on the square of the mass within the observable universe. So multiplying the radius of the universe by a tenth would have the following effect: the mass within this smaller sized universe would be reduced by a factor of 10^{-3} (since volume is proportional to r^3), and mass is squared in the entropy formula. Hence, the exponent is reduced by 6; so the overall entropy is reduced from (10^{123}) to (10^{117}). The probability of finding ourselves in either state would be approximately 10 raised to the appropriate power.[73]

How many more inflationary events in a multiverse, then, would produce a life amenable but smaller universe? This is obtained from dividing out the probabilities:

$$E = 10^{-(10\wedge 117)}/10^{-(10\wedge 123)}.$$

Take the logarithm of both sides and simplify:

$$\mathrm{LOG}(E) = -10^{117} + 10^{123} = 10^{123}.$$

Here -10^{117} is negligible compared with the larger 10^{123}. Hence, $E = 10^{(10\wedge 123)}$. The reciprocal of this represents the likelihood of finding ourselves *in a big universe* versus a small one.

---

[73] The entropy is related to the number of possible configurations of a system of particles. The number of configurations is approximately equal to the exponential of the entropy. Given the size of the numbers involved, there is essentially no difference between e^x and 10^x. So Penrose uses base 10 for convenience.

So it is exceedingly improbable to find ourselves as the product of an inflationary event of a generic manifold, as Bojowald et al. originally proposed. Hence, Penrose argues, the entropy of the initial manifold must be exceedingly low.

Aside from this, the epistemic argument takes no account of entropy generation during the cycle. Over infinite time, this would have to be a factor, although it may be negligible for a single cycle. Eggs break, people get old, stars burn out. This entropy may be negligible compared to black hole formation. But over infinite cycles, it would add up.

The second solution – that the "classical" understanding of entropy is misleading – mitigates the problem of entropy growth but does not resolve it. While entropy as classically calculated may be too high (i.e. Penrose's estimation overestimates entropy due to contextual lack of information), the quantum approach still recognizes entropy (and entropy growth) as a genuine physical quantity. Black holes are still highly entropic. So their formation during a cycle, especially if one lands in a Big Crunch, would still cause a final manifold to have more entropy than an initial manifold. One would still expect that this situation would imply a beginning, since it implies a heat death given infinite cycles. Hence, Bojowald is only being realistic in opting for the third solution, that the cyclic LQG model does, indeed, need to be fully reversible.

But opposing the second law of thermodynamics is a formidable task; one recalls the words of early twentieth-century cosmologist Sir Arthur Eddington:

> If someone points out to you that your pet theory of the universe is in disagreement with Maxwell's equations – then so much the worse for Maxwell's equations. If it is found to be contradicted by observation – well, these experimentalists do bungle things sometimes. But if your theory is found to be against the second law of thermodynamics I can give you no hope; there is nothing for it but to collapse in deepest humiliation. (Eddington 1948, p. 74)

So the question is, can LQG really prove reversibility? We must await further developments in the field. It is fair to say that the prevailing view in cosmological community at large disagrees with Bojowald. As Bojowald himself admits, the jury is still out:

> **Sinclair**: 'Is the assumption of entropy reversal: a) an initial assumption around which a self-consistent LQG model is built, or b) a natural fallout of LQG models?'
>
> **Bojowald**: 'It is definitely b), as far as we can see it currently. Many details have to be filled in, but we do not make any a priori assumptions about entropy. Since entropy is not a fundamental object but a measure for our ignorance of what is happening microscopically, it is not even possible to make such an assumption in a theory like loop quantum gravity. We can only make assumptions on microscopic objects, and then see what this implies for more common quantities. What we don't know yet is how entropy changes balance out exactly. *So we are not sure if entropy does not increase from cycle to cycle. We can only say that the usual black hole arguments are not as strong as usually assumed*'. (pers. comm., March 29, 2006; emphasis added)

To his credit, Bojowald has not simply assumed zero net entropy in his model. He is using the right approach; allowing the physics to predict entropy accrual over time. So far, however, the only reliable conclusion is that LQG may show that Penrose's entropy arguments must be modified from their semiclassical orientation (what we have called option 2). This has failed to show, however, that universe entropy does not increase cumulatively cycle-by-cycle. The force of Penrose's argument remains intact even if his quantitative assessment of entropy must change.

Aside from the entropy issue, there remains the issue of dark energy, which may have the potential to stop cycling and induce an open-ended expansion. The current empirically observed dark energy effect, for example, appears adequate to produce an open-ended accelerated expansion. This result would be definitive if the dark energy were of the form of a cosmological constant (i.e. its value were independent of space and time; see Barrow & Dabrowski 1995).[74] As related earlier (Overbye 2006), this does appear to be the fate of the present day universe. But if an entropy gain (cycle-to-cycle) is denied, one can never have more than one "cycle." The cosmological constant would have led to open-ended expansion *the first time*. Hence, the initial singularity (our Big Bang) represents an absolute beginning.

If the dark energy were of the form of "quintessence" (i.e. had a value that is dependent on space and/or time), however, then it would be possible that its value could reverse and be consistent with a collapse phase, even given the current observational evidence. But then a new problem could intrude. Bojowald recognizes that after the bounce and the following energy transfer, different modes of the matter fields will become excited such that the next bounce will differ from the preceding one. But if the quintessence term changes, then perhaps the most generic model of LQG would be a hybrid between the cyclic and single-bounce models. On some particular cycle a value for the quintessence term would be such that it would lead to an open-ended expansion. Bojowald responds,

> If there is just a cosmological constant, it would be fixed for all cycles and not change. But if there is some kind of quintessence, you are right that its initial conditions for the classical phase would be affected by the bounce transition. So your scenario can be realized in suitable quintessence models. However, what people usually prefer are quintessence models which have an attractor behavior at late times, or the so-called tracking solutions. This allows one to avoid too much fine-tuning, and it makes the dynamics less sensitive to changes in initial values. For such models one thus does not generically expect big changes between cycles. On the other hand, since the effect would be quite dramatic if open-ended expansion can be realized, even a non-generic possibility can be important. (pers. comm., March 9, 2006)

Given an infinite number of rolls of the dice, any nonzero probability that quintessence could produce an open-ended expansion would be sufficient to do so. An open-ended expansion implies that that the overall number of cycles has been finite, and hence, the model would not be beginningless.[75]

In general, LQG looks like a promising alternative to string theory as a candidate for quantum gravity. But building a genuinely beginningless cyclic LQG model seems to be a far more difficult challenge.[76]

---

[74] According to the National Aeronautics and Space Administration (NASA) (http://map.gsfc.nasa.gov/m_mm/mr_limits.html), their current data tends to favor the cosmological constant theory for dark energy as opposed to quintessence, although the latter is not ruled out.

[75] Barrow and Dabrowski do indicate that if the dark energy were of the type to decay away into matter and radiation; that is, if it is impermanent, then cycling would recommence after the decay.

[76] We note that there are other LQG models that feature Vaas-type pseudobeginnings and seem to us to be viable.

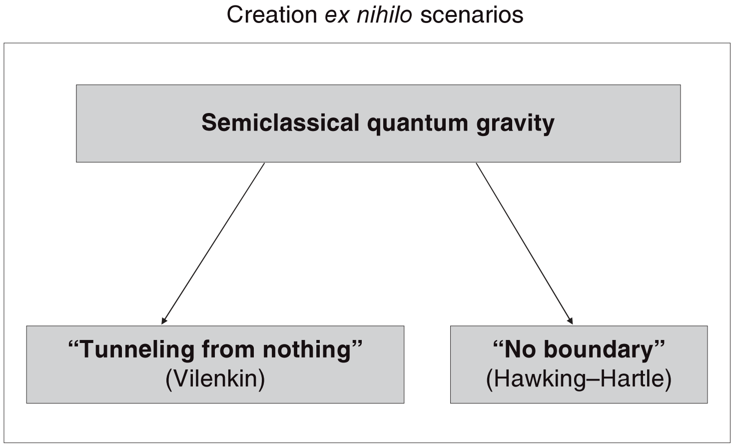


Figure 3.19 Quantum models with an explicit beginning to the finite past.

IVc. Semiclassical creation ex nihilo *models*

When inflation was first explored as a concept, it was understood to be a phase in the history of our universe that occurred sometime after the origin event and ended long ago. This naturally led to the question: "How could inflation itself have started?" (Figure 3.19).

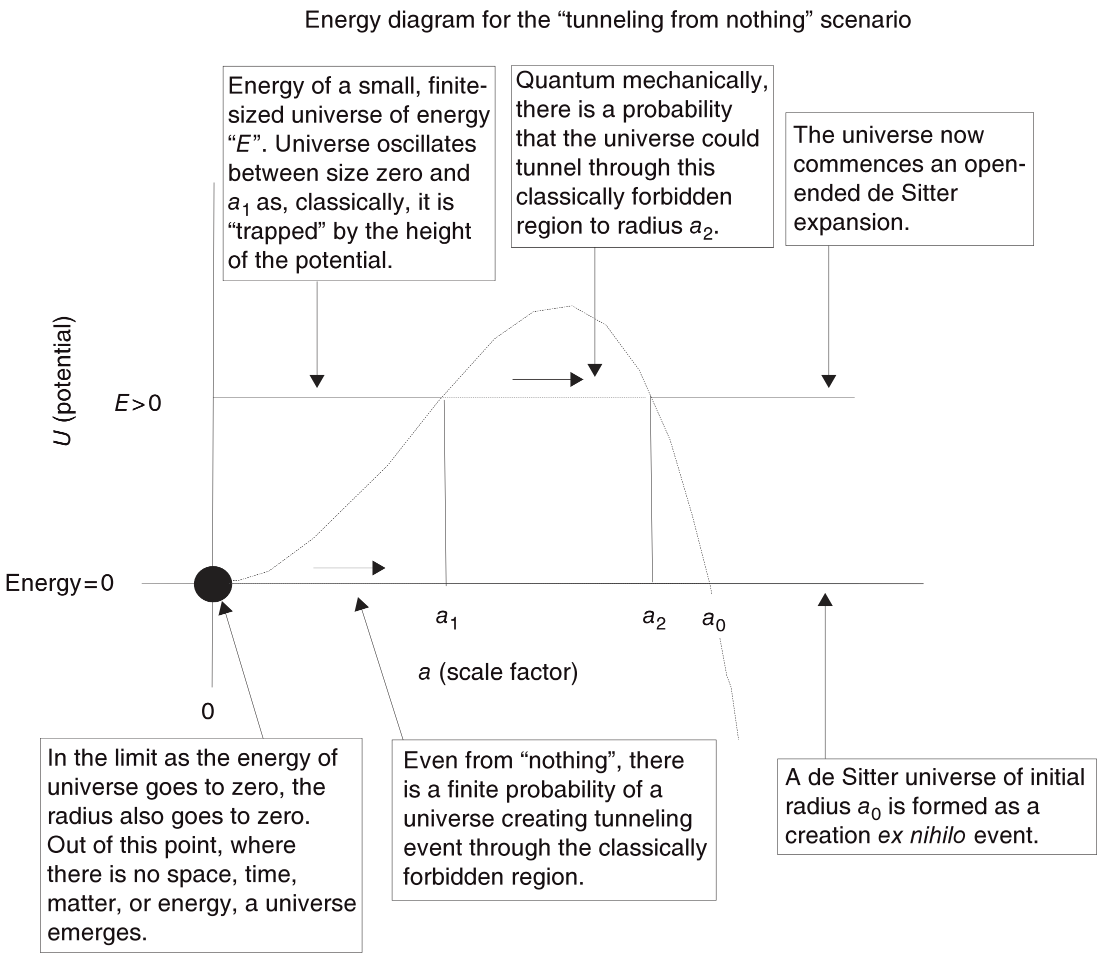
That understanding has changed with the work of cosmologists such as Andrei Linde (chaotic inflation) and the recent work on the String Landscape by pioneers such as Susskind, Bousso, and Polchinski (and many others). Here, inflation is not viewed just as a phase in development but instead as the dominant feature of a larger multiverse, within which our "universe" is just a regional phenomenon. Nevertheless, given the singularity theorems developed by Borde, Vilenkin, and Guth, inflation itself is viewed as not past eternal.[77] Hence, the old question still persists: How did inflation get started? Vilenkin (1982) explains an approach he initiated to address this question:

> Many people suspected that in order to understand what actually happened in the beginning, we should treat the universe quantum-mechanically and describe it by a wave function rather than by a classical spacetime. This quantum approach to cosmology was initiated by DeWitt and Misner, and after a somewhat slow start received wide recognition in the last two decades or so. *The picture that has emerged from this line of development is that a small closed universe can spontaneously nucleate out of nothing, where by 'nothing' I mean a state with no classical space and time.* The cosmological wave function can be used to calculate the probability distribution for the initial configurations of the nucleating universes. Once the universe nucleates, it is expected to go through a period of inflation, driven by the energy of a false vacuum. The vacuum energy is eventually thermalized, inflation ends, and from then on the universe follows the standard hot cosmological scenario. (Vilenkin 2002, p. 2; emphasis added)

Vilenkin uses quantum tunneling of a particle through a potential well as an analogy for the whole universe. In Figure 3.20, the '$E > 0$' line represents an ordinary, closed FRW

---

[77] One remembers, of course, exceptions such as asymptotically static models and Linde's objection that the behavior of the whole may not be the same as its parts.



Energy of a small, finite-sized universe of energy “E”. Universe oscillates between size zero and a_1 as, classically, it is “trapped” by the height of the potential.

Quantum mechanically, there is a probability that the universe could tunnel through this classically forbidden region to radius a_2.

The universe now commences an open-ended de Sitter expansion.

In the limit as the energy of universe goes to zero, the radius also goes to zero. Out of this point, where there is no space, time, matter, or energy, a universe emerges.

Even from “nothing”, there is a finite probability of a universe creating tunneling event through the classically forbidden region.

A de Sitter universe of initial radius a_0 is formed as a creation *ex nihilo* event.

Figure 3.20 Creation *ex nihilo* of a universe.

universe, which (classically) does not have enough energy to escape into an open-ended expansion. Hence, it goes through continued Big Bang/Big Crunch cycles. In classical physics (GR), this state of affairs would persist forever.[78] But quantum gravity provides a way out. There is a finite probability that, instead of a recollapse, the universe will “tunnel” through the energy barrier and commence an inflationary expansion instead.

This approach still does not solve the problem of creation; rather it has moved the question back one step: to the initial, tiny, closed, and metastable universe. This universe state can have existed for only a finite time. Where did it come from?

Vilenkin’s solution was to consider what happens in the limit as the energy of this initial closed universe becomes zero. The scale factor of the universe becomes zero as well. This is the genesis of the claim that the universe is created from “nothing.” There is no space, time, matter, or energy. This constitutes a topological transformation: “Creation of a universe from nothing . . . is a transition from the null topological sector containing no universes at all to the sector with one universe of topology S^3” (Vilenkin 1994, p. 23). Vilenkin grants that nothingness so conceived is not the same as the *absence* of being:

---

[78] Assuming, of course, that the problems of cyclic cosmologies (entropy buildup, bounce physics, etc.) were solvable.

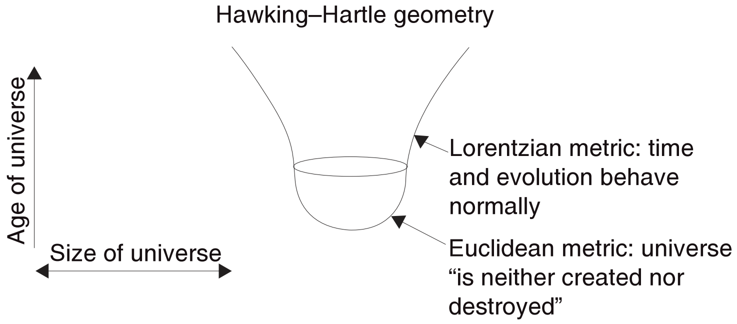


Figure 3.21 Transition to "normal" time in a Hartle–Hawking approach. This is really just a mathematical technique called *analytic continuation*, which is useful for problems with "badly behaved" functions, but which does not imply ontological commitment.

> I understand that a universe of zero radius is not necessarily the same thing as no universe at all. But mathematically my quantum tunneling from nothing is described by the same "shuttlecock" geometry as Hartle and Hawking [NB Figure 3.21]. (In fact, the shuttlecock first appeared in my 1982 paper.) This geometry is certainly not past-eternal, and there is no point on the Euclidean sphere which you can identify as the initial universe of zero size. So, if the Hartle-Hawking approach avoids the "paradoxes of creation", I don't see why my mine doesn't. (pers. comm., October 23, 2006)

> The mathematical description of a tunneling universe suggests that time, space and matter came into being a finite time ago. "Nothing" refers to the "prior" state. (pers. comm., October 30, 2006)

The universe clearly has a beginning in this approach. But the claim seems to go beyond this: the universe "came into being" from a "prior" state of "nothing." This latter claim (that the universe comes into being) is significant because the notion of time becomes ill defined in these models near the beginning.

A similar approach to Vilenkin's is the "no-boundary proposal" of James Hartle and Stephen Hawking. Hartle and Hawking make use of Richard Feynman's approach to QM. Feynman's (1985) approach is to find the probability for a certain final quantum state by a path integral, a "sum over histories." In quantum terms, this is a superposition of states. Every possible universe history is part of the wave function; each possible end state has an associated probability of realization (alternatively, one can view all the possible states as existing "somewhere" in a realized multiverse). The path integral of the universe, as one might imagine, presents an intractable problem; hence, Hartle and Hawking make an educated guess as to a subset of possible universes that are expected to dominate the calculation and assert compact spaces only in order to solve the problem that path (or contour) integrals often are badly behaved. They tend to diverge and give nonsense answers (such as "infinity" for a probability).

To rectify the divergence problem, Hartle and Hawking make use of Euclidean metrics to describe the earliest phases of the universe's existence. This is a mathematical technique called analytic continuation (or more specifically, a Wick rotation), which allows one to analyze a better behaved function in the domain of interest. A "normal" or "Lorentzian" metric has signature (–,+,+,+), indicating a special character to the time dimension. A Euclidean metric (+,+,+,+) treats time the same as it does the space dimensions. The Wick rotation takes the real time variable "*T*" and replaces it with the imaginary quantity "*I* × *T*". Here "*i*" is the square root of negative one. Thus, Hartle and Hawking are said to employ "imaginary time" in their model.

The use of a Euclidean metric has the effect of removing the initial singularity predicted by "pure" GR – hence the term "no boundary proposal." There is, in general, no ontological commitment associated with this change in metric signature. But Hawking, in his popular writings, appears to assert the physical reality of the Euclidean signature. This is the genesis of Hawking's claim that "[the universe] would neither be created nor destroyed. It would just BE" (Hawking 1988, p. 141). Hawking goes on to ask "What place, then, for a creator?" Yet, curiously, Hartle and Hawking also claim their universe can be interpreted to have come into being out of "nothing":

> One can interpret the functional integral over all compact four-geometries bounded by a given three-geometry as giving the amplitude for that three-geometry to arise from a zero three-geometry; that is, a single point. In other words, the ground state is the probability for the Universe to appear from nothing. (Hawking & Hartle, 1983, p. 2961) [They then refer to Vilenkin's "tunneling from nothing" paper.][79]

A third interpretation of these results exists as well. Recall that Gott and Li have criticized the creation *ex nihilo* approach on two grounds. (1) Transitions in QM are always between allowed classical states (Vilenkin and Hartle–Hawking's approach has a transition from a classically forbidden region to a classically allowed region). (2) The Vilenkin and Hartle–Hawking approaches should contain realistic energy fields (something closer to what we actually see in nature). If they did, then Heisenberg's uncertainty principle would require that the initial state of their models have a finite and nonzero energy. It that is the case, then semiclassical quantum models actually start in a classically allowed metastable state, rather than "nothing." Gott and Li elaborate:

> The problem with this model [Vilenkin and Hawking–Hartle] is that it ignores the "zero-point energy." If there is a conformal scalar field ϕ, then the "energy" levels should be $E_n = n + 1/2$. Even for $n = 0$ there is a "zero-point-energy." The potential makes the system behave like a harmonic oscillator in the potential well near $a = 0$. A harmonic oscillator cannot sit at the bottom of the potential well – the uncertainty principle would not allow it. There must be some zero-point-energy and the particle must have some momentum, as it oscillates within the potential well when the field ϕ is included. Thus, when the "zero point-energy" is considered, we see that the initial state is not a point but a tiny oscillating ($0 < a < a_1$) big bang universe, that oscillates between big bangs and big crunches (though the singularities at the big bangs and big crunches might be smeared by quantum effects). This is the initial *classical* state from which the tunneling occurs. *It is metastable, so this oscillating universe could not have existed forever: after a finite half-life, it is likely to decay.* It reaches maximum radius a_1, and then tunnels to a classical de Sitter state at minimum radius a_2 where $a_2 < a_0$. (Gott & Li 1998, p. 38; emphasis added)

---

[79] Note that in Vilenkin's comment (above), he indicates that the geometry of his model can be understood in terms of an initial Euclidean metric (which would represent the "forbidden" region of the energy diagram (Figure 3.20).

The relevant question for the interpretation of these models, then, is: Is the universe (i) created from nothing, (ii) neither created not destroyed but, in effect, timelessly subsistent, or (iii) left ultimately unexplained, since the existence of the initial, metastable, closed, Planck-sized universe out of which our universe was born is not itself accounted for? Option (i), if adopted, clearly implies that the universe began to exist. Nonetheless, as even Vilenkin himself recognizes, "nothing" as he describes it is not the same as absence of being. Whatever reality underlies the laws of QM must exist at least, so as to transform the null topological sector into an FRW universe. The laws themselves, if they exist, are mere abstract objects of a propositional nature and so do not stand in causal relations to anything. As such they are irrelevant to what happens in the world; it is the reality that corresponds to them that distinguishes merely logical from real possibility. As Heinz Pagels once remarked: "This unthinkable void converts itself into the plenum of existence–a necessary consequence of physical laws. Where are these laws written into that void? What "tells" the void that it is pregnant with a possible universe? It would seem that even the void is subject to law, a logic that exists prior to space and time" (Pagels 1985, p. 347). Option (i), then, is mistaken, being based upon an idiosyncratic use of the word "nothing." That initial state is clearly not nothing, but something.

As for option (ii), Hawking himself seems to give good grounds for treating their proposal as an instrumental approach only. In his collaboration with Roger Penrose, *The Nature of Space and Time* (Hawking & Penrose 1996), he demonstrates the same mathematical approach (analytic continuation) to describe pair production of electron–positron pairs in a strong electric field. This is a standard mathematical technique sometimes used when complex analytic functions are better behaved in a certain domain than their real counterparts. It does not imply ontological commitment to the alternative description, however. It seems to us that given the unintelligibility of the "imaginary time" region in these models, it is most reasonable to treat this approach as nonrealist in character.

As for option (iii), we seem to have the same sort of situation that we encountered with the Emergent and PBBI models with their associated metastable ESS and SPV states. The universe cannot be past eternal because the initial metastable state can have had only a finite lifetime. This seems to us to be the most reasonable option to take for a realist interpretation of these models. It employs known, meaningful interpretations of physical phenomena from "classical" quantum theory and extends them to the quantum gravity models. One avoids the problems associated with the novelty of asserting a zero-energy condition for the initial state (denied by the Heisenberg uncertainty principle), the novelty of asserting a quantum transition from a forbidden to a classically allowed state (normal quantum theory only includes transitions over or through forbidden regions from one allowed state to another), and it is consistent with more realistic energy fields. Option (iii) is also consistent with the second premise of the *kalam* cosmological argument.

2.34. Summary

Taking as a springboard challenges to the Hawking–Penrose singularity theorems, we have surveyed the historical development of three research programs each pursuing known exceptions to the theorems:[80] (1) CTCs; (2) violation of strong energy condition (eternal

---

[80] The fourth and fifth conditions, viz., satisfaction of a generic energy condition and the existence of a closed trapped surface in our past are easily met.

inflation); and (3) falsity of GR (quantum gravity). Major theoretical developments concerning options (2) and (3) were inflationary theory (repulsive gravity) and semiclassical quantum gravity.

CTCs, while interesting, seem to fail given the CPC. Counterexamples to the conjecture can be found, but they seem to be unphysical and/or infinitely fine-tuned toy models.

As for inflation, an exotic type of energy field possesses the bizarre property of *negative* pressure. This energy wants to collapse in on itself due to pressure, yet according to Einstein's equations, pressure also produces a gravitational force. If the pressure is negative, the gravitational force is *repulsive*. It turns out that the repulsion is the stronger of the two tendencies; *greatly* so. The universe can expand many orders of magnitude in a fraction of a second; from invisibly small to bigger than the entire observable sky. This development permitted a reappraisal of the question of origins that lasted two decades. Finally, a new singularity theorem, developed by Arvind Borde, Alexander Vilenkin, and Alan Guth showed that this model – the inflationary universe – still had a beginning in the finite past.

It is fascinating to note that the recent history of cosmology can be mapped by attempts to overcome these singularity theorems. Following the BVG Theorem in 2003, attempts to build models have been based on exceptions to *that* theorem. These were (1) average past expansion of universe is negative (contraction-bounce), (2) average past expansion is zero (asymptotically static universe), (3) average past expansion is zero (cyclic universe), and (4) exotic space-time.

The first, somewhat akin to the de Sitter universe, featured an infinite contraction into a bounce at the singularity, followed by our current expansion. But it featured a Hobson's choice between an acausally fine-tuned contraction or an infinitely distant beginning point. In either case, the bounce was predicted to be chaotic (due to BKL oscillations), and hence its "Big Bang" would look nothing like the one we actually measure.

The second case, exemplified by the Emergent model class, features an unstable or metastable initial state followed by an inflationary expansion. But an unstable state (ESS) or a metastable state (the LQG addition to the Emergent model) has a finite lifetime. So the beginning of the universe reasserts itself.

The third case has long been known, since the original models of Richard Tolman, to be problematic on entropy grounds. Baum and Frampton have sought to solve this problem through an approach dubbed "the phantom bounce." Here, the universe undergoes a superexpansion aided by the effects of "phantom" energy, which accelerates the expansion with effects similar to inflation. The universe would then fractionate into a multiverse, with almost all of the daughter universes having jettisoned their entropy given the initial expansion. Could this work? It does not appear so, the chief difficulty being that causal reconnection of universe likely occurs at the "turnaround" point (when expansion goes over to contraction) leading to *one* contraction, rather than to many. A more certain conclusion is that, even given an "empty" universe[81] undergoing contraction, chaotic fluctuations as the contraction nears a singularity would *create* matter, thereby leading to a chaotic crunch, which would prevent cycling.

The fourth case features a deconstruction of time itself. It postulates two mirror-image, inflationary expansions, where the arrows of time lead *away* from a past boundary. Thus,

---

[81] That is, except for the zero entropy phantom energy.

the mirror universe is not our past. This is just a case of a double Big Bang. Hence, the universe *still* has an origin.

What about the attempts at quantum gravity models (many of which overlap with those previously discussed)? They include (1) string models, (3) LQG models, and (3) semiclassical quantum gravity models.[82]

The most popular new field is the class of string models. Two prominent approaches are the Ekpyrotic/cyclic and PBBI models. But the first is subject to the BVG singularity theorem and hence has a beginning. The second is probably to be interpreted instrumentally only and seems in any case to have the characteristics of the Emergent class with respect to the initial state. Hence, it has a metastable beginning phase, called the SPV, which cannot be eternal in the past. The most popular string model is the generalization of inflationary theory known as the String Landscape. This scenario, however, is known to have a beginning to the finite past due to the same BVG Theorem.

LQG is a competitor to string theory. We saw one such application when LQG was incorporated into the Emergent model class. Another LQG approach is to try to build a viable cyclic model. LQG seems a promising approach to address the issue of BKL chaos, hence, perhaps, providing a justification for a bouncing model. But the cyclic LQG still seems to fail to account for the entropy effects that usually doom infinite cyclicity. Even if it did, current observations show that our universe is in an open-ended expansion rather than a Big Bang/Big Crunch cycle. So current LQG attempts do not appear to support a past eternal universe.

The semiclassical quantum gravity models have, in their very approach, a beginning to the finite past. The beginning has been described in three possible (and not consistent) ways. Either:

1. The universe came into being from a prior state of null topology (but somehow containing the laws of physics themselves) to a Lorentzian metric (the normal universe). Hence, the universe "tunneled from nothing" into existence.
2. The initial state of the universe is uncreated. This is due to the nature of time in a Euclidean metric. It is equivalent to a spatial dimension.
3. Because the initial state of the geometry must have a zero point energy, it is in a classical state with a Lorentzian metric. It is a metastable closed universe. Hence, this state could not have existed forever and, in a manner unexplained by the model, began to exist.

The second description seems to be purely instrumental in character. The first and third imply that the universe began to exist. Hence, semiclassical models are supportive of the universe's having a beginning.

Our survey shows that contemporary cosmology is quite supportive of the second premise of the *kalam* cosmological argument. Further, this conclusion is not reached through ferreting out elaborate and unique failure conditions for scores of individual

---

[82] In interests of economy, we have not discussed earlier attempts to postulate a regional, immanent status for the Big Bang which leave the origin of the initial space in question. It would seem that this space cannot contract, expand, or be static without violating criteria mentioned earlier (singularity theorems, chaotic bounces, metastable beginnings, etc.). Thus, these models proved to be untenable.

Table 3.1 Effective principles in ruling out a beginningless model

Model average expansion history	*Condition requiring a beginning*
1) Expanding models	Singularity theorems
2) Asymptotically static models	Metastability
3) Cyclic models	Second law of thermodynamics
4) Contracting models	Acausal fine-tuning

models. Rather, the repeated application of simple principles seems effective in ruling out a beginningless model.[83] They are found in Table 3.1.

It seems that the field of cosmology, therefore, yields good evidence that the universe began to exist.

1.0. Everything That Begins to Exist Has a Cause

To return, then, at length to the first premise of the *kalam* cosmological argument

1.0. Everything that begins to exist has a cause,

we take (1.0) to be obviously true – at the least, more plausibly true than its negation. Three reasons can be given in its support.

1.1. *Ex nihilo nihil fit*

First and foremost, the principle is rooted in the metaphysical intuition that something cannot come into being from nothing. For to come into existence without a cause of any sort is to come into being from nothing.[84] To suggest that things could just pop into being uncaused out of nothing is to quit doing serious metaphysics and to resort to magic. Nobody *sincerely* believes that things, say, a horse or an Eskimo village, can just pop into being without a cause. But if we make the universe an exception to (1.0), we have got to think that the whole universe just *appeared* at some point in the past for no reason whatsoever.

Sometimes it is said that quantum physics furnishes an exception to the claim that something cannot come into being uncaused out of nothing, since on the subatomic level,

---

[83] This does not exhaust the list of possible model formulations. Different types of mosaic models are likely to be the next thing to come down the pike. Nor did we survey every model that exists in the present (nor could we). There is also the complaint of philosopher James Brian Pitts: "What rights do unborn theories possess?"

[84] See Oppy's suggestion that we construe the first premise to assert that everything that begins to exist has a cause of some sort, whether efficient or material (Oppy 2006b, p. 153). His protest that the cause of the universe could then be a material cause is not troubling because the incorporeality of the First Cause plausibly follows from a conceptual analysis of what it is to be a cause of the space-time universe, on which see further discussion below.

so-called "virtual particles" come into being from nothing. In the same way, certain cosmogonic theories are interpreted as showing that the universe could have sprung into being out of the quantum vacuum or even out of nothingness. Thus, the universe is said to be the proverbial "free lunch."

This objection, however, is based on misunderstanding. In the first place, wholly apart from the disputed question of whether virtual particles really exist at all,[85] not all physicists agree that subatomic events are uncaused. A great many physicists today are quite dissatisfied with the traditional Copenhagen interpretation of quantum physics and are exploring deterministic theories like that of David Bohm.[86] Indeed, most of the available interpretations of the mathematical formalism of QM are fully deterministic. Quantum cosmologists are especially averse to Copenhagen, since that interpretation in a cosmological context will require an ultramundane observer to collapse the wave function of the universe. Thus, quantum physics hardly furnishes a proven exception to (1.0). Second, even on the indeterministic interpretation, particles do not come into being out of nothing. They arise as spontaneous fluctuations of the energy contained in the subatomic vacuum, which constitutes an indeterministic cause of their origination. Third, the same point can be made about theories of the origin of the universe out of a primordial vacuum. Popularizers touting such theories as getting "something from nothing" apparently do not understand that the vacuum is not nothing but is a sea of fluctuating energy endowed with a rich structure and subject to physical laws. Such models do not, therefore, involve a true origination *ex nihilo*.[87]

Neither do theories such as Vilenkin's quantum creation model. Vilenkin invites us to envision a small, closed, spherical universe filled with a so-called false vacuum and containing some ordinary matter. If the radius of such a universe is small, classical physics predicts that it will collapse to a point; but quantum physics permits it to "tunnel" into a state of inflationary expansion. If we allow the radius to shrink all the way to zero, there still remains some positive probability of the universe's tunneling to inflation. As we have seen, Vilenkin equates the initial state of the universe explanatorily prior to tunneling with nothingness. But postulating such an equivalence is grossly misleading. As Vilenkin's own diagram illustrates (Vilenkin 2006, p. 180), quantum tunneling is at every point a function from something to something (Figure 3.22). For quantum tunneling to be truly from nothing, the function would have to have only one term, the posterior term. Another way of seeing the point is to reflect on the fact that to have no radius (as is the case with nothingness) is not to have a radius, whose measure is zero. Thus, there is no basis for the claim that quantum physics proves that things can begin to exist without a cause, much less that universe could have sprung into being uncaused from literally nothing.

A more pertinent objection to the justification of (1.0) on the basis of the metaphysical principle that something cannot come from nothing issues from the partisans of the B-Theory of time (Grünbaum 1967, p. 153; 2000, p.16). For B-Theorists deny that in beginning to exist the universe *came into being* or *became actual.* They thereby focus attention on the theory of time underlying the *kalam* cosmological argument. From start to finish, the *kalam* cosmological argument is predicated upon the A-Theory of time. On a B-Theory

---

[85] See Weingard (1982) for doubts.

[86] See Cushing (1994) and Cushing (1998).

[87] See remarks by Kanitscheider (1990).

Publisher's Note:
Permission to reproduce this image
online was not granted by the
copyright holder. Readers are kindly
requested to refer to the printed version
of this chapter.

Figure 3.22 Quantum tunneling of the universe to an inflationary condition. Note that the tunneling is at every point from something to something; the origin of the initial point remains unexplained. (source: Vilenkin 2006)

of time, the universe does not in fact come into being or become actual at the Big Bang; it just exists tenselessly as a four-dimensional space-time block that is finitely extended in the *earlier than* direction. If time is tenseless, then the universe never really comes into being, and, therefore, the quest for a cause of its coming into being is misconceived. Although G. W. F. Leibniz's question, *Why is there* (tenselessly) *something rather than nothing?*, should still rightly be asked, there would be no reason to look for a cause of the universe's beginning to exist, since on tenseless theories of time the universe did not begin to exist in virtue of its having a first event anymore than a meter stick begins to exist in virtue of having a first centimeter. In affirming that things which begin to exist need a cause, the *mutakallim* assumes the following understanding of that notion, where "*x*" ranges over any entity and "*t*" ranges over times, whether instants or moments of nonzero finite duration:

A. *x* begins to exist at *t* iff *x* comes into being at *t*.
B. *x* comes into being at *t* iff (i) *x* exists at *t*, and the actual world includes no state of affairs in which *x* exists timelessly, (ii) *t* is either the first time at which *x* exists or is separated from any $t' < t$ at which *x* existed by an interval during which *x* does not exist, and (iii) *x*'s existing at *t* is a tensed fact.

By presupposing an A-Theory of time, according to which temporal becoming is real, the *mutakallim* justifiably assumes that the universe's existing at a first moment of time represents the moment at which the universe came into being. Thus, the real issue separating the proponent of the *kalam* cosmological argument and critics of the first premise is the objectivity of tense and temporal becoming.

It is worth recalling that having a beginning does not entail having a beginning point, lest someone should essay to avert the beginning of the universe by asserting that the series of past states of the universe might converge toward $t = 0$ in the metrically finite past as a merely ideal limit.[88] On such a proposal, it is alleged, the universe lacks a beginning point and can therefore be said to have no earliest time of its existence. Hence, it does not, despite its past temporal finitude, begin to exist, and so no cause of the universe's origin need be postulated.

It is not clear that such a proposal is even physically possible;[89] but let that pass. The fundamental philosophical shortcoming of this proposal is its gratuitous assumption that having a beginning entails having a beginning point. This is not, in fact, how the locution "begins to exist" is typically understood.[90] Contemporary cosmologists frequently "cut out" the initial cosmological singularity as a merely ideal point on the boundary of space-time, so that the universe has no first instant of its existence; but they do not therefore think that the universe no longer begins to exist or that the mystery of the origin of the universe has thereby been solved.

Granted a tensed theory of time, the key criterion for determining if something has a beginning is its past metrical finitude.[91] Something has a beginning just in case the time during which it has existed is finite. Time itself may be said to begin to exist just in case for any nonzero, finite interval of time that one picks, there are only a finite number of congruent intervals earlier than it. Or, alternatively, time begins to exist just in case for some specified nonzero, finite interval of time, there are no congruent intervals earlier than it. In either case beginning to exist does not entail having a beginning point.

So understood, deleting the beginning point of a thing's existence does not imply that that thing no longer begins to exist and therefore came into being uncaused. As mentioned, cosmologists continue to puzzle over the origin of the universe out of nothing, whether or not it had an initial instant of its existence. One is reminded in this connection of the ancient sorites-style problems of starting and stopping.[92] If there is a last instant at which some object is at rest, then when does it begin to move? The answer can only be

---

[88] Quentin Smith suggested as much on the occasion of the forum "Science and Religion," University of California, Santa Barbara, January 30, 2004, featuring William Lane Craig, Richard Gale, Alvin Plantinga, and Quentin Smith, available in DVD format through www.veritas-ucsb.org. See also Grünbaum (2000, pp. 16–7).

[89] It will depend crucially on one's theory of quantum gravity. A typical approach to marrying quantum theory to GR involves describing the evolution of space-time as a path integral (a sum over all possible paths) in superspace, which is a space of points representing three-dimensional configurations of the universe. The points of this configuration space can be regarded as instantaneous states or even as instants, but the fact that in the quantum theory one has a path integral, rather than a single path, makes it impossible to "stack" these instants into a unique history constituting a space-time. So, eventually, the dividing process becomes ill-defined in the theory. The fact that the first split second of the universe's existence as measured in cosmic time is not resolvable into a unique sequence of ever briefer states is not inconsistent with there being a first second of its existence.

[90] We might mention as well that the proposal commits us to the reality of points, which surely overloads the expression "begins to exist" with unintended ontological commitments. Such metaphysical issues should be decided by argument, not definition.

[91] Oppy seems ready to concede that the second premise of the *kalam* cosmological argument is true if we accept that something begins to exist iff it is finite in the past (Oppy 2006b, p. 147). Given a tensed theory of time, we think that that biconditional holds for wholly temporal entities. So the remaining question, as Oppy says, is whether everything that begins to exist in this sense has a cause. While not denying the causal principle, Oppy denies that that premise is either obviously true or supported by experience (Oppy 2006b, pp. 150–3).

[92] See Sorabji (1983, pp. 403–21).

that there is no first instant of its motion. Nonetheless, the object does begin to move and plausibly requires a cause to set it in motion. Similarly, if a thing has not always existed, then it is plausible that that thing began to exist and requires a cause to bring it into being, whether or not there was a first instant at which it existed.

In view of the seemingly evident metaphysical absurdity of something's coming into being without a cause, one would not anticipate that naturalists would deny the first premise of the *kalam* cosmological argument and assert that the universe sprang into existence uncaused out of nothing. It is therefore surprising how many nontheists, confronted with the evidence of the universe's beginning, have taken this route. For example, Quentin Smith, commenting that philosophers are too often adversely affected by Heidegger's dread of "the nothing," concludes that the most reasonable belief is that we came "from nothing, by nothing, and for nothing"(Craig & Smith 1993, p. 135)[93] – a nice ending to a sort of Gettysburg address of atheism, perhaps. Such a confession merely expresses the faith of an atheist. For it is, we repeat, literally worse than magic.

1.2. Why only universes?

Second, if things really could come into being uncaused out of nothing, then it becomes inexplicable why just anything or everything does not come into existence uncaused from nothing. Why do bicycles and Beethoven and root beer not pop into being from nothing? Why is it only universes that can come into being from nothing? What makes nothingness so discriminatory? There cannot be anything about nothingness that favors universes, for nothingness does not have any properties. Nothingness is the absence of anything whatsoever. As such, nothingness can have no properties, since there literally is not anything to have any properties. Nor can anything constrain nothingness, for there is not anything to be constrained. How silly, then, when popularizers say things such as "Nothingness is unstable to quantum fluctuations" or "The universe tunneled into being out of nothing"!

Some critics have responded to this problem by asserting that while (1.0) is true with respect to things *in* the universe, it is not true *of* the universe itself. But this proposed differentiation misconstrues the nature of the Causal Principle. Premise (1.0) does not state a merely physical law such as Boyle's law or the second law of thermodynamics, which are contingent laws of nature. Premise (1.0) is not a physical principle. Rather, it is a metaphysical principle: being cannot come from nonbeing; something cannot come into existence uncaused from nothing. Such claims are not contingent upon the properties, causal powers, and dispositions of the natural kinds of substances which happen to exist. Critics have given no good reason for construing such claims as merely physical rather

---

[93] Smith's most recent criticism of the *kalam* cosmological argument is also a denial of the first premise, despite Smith's avowal that he now accepts the conclusion that the universe has a cause of its existence (Smith 2007, pp. 182–98). Smith's most recent published position is that the initial singular point of the universe is not real and that, therefore, the sequence of instantaneous states of the universe is a beginningless series converging toward zero as a limit. Each state is caused by its prior states and there is no first state. Any nonzero interval or state, such as the first second of the universe's existence, "is not caused by any or all of its instantaneous states and is not caused by any external cause" (Smith 2007, p. 189). Smith takes "the beginning of the universe" to refer to the Planck era, that state which lasts until 10^{-43} seconds after the singularity. As a state of nonzero duration, the beginning of the universe therefore has no cause of any sort. The universe therefore comes into being uncaused out of nothing.

than as metaphysical claims. The Causal Principle plausibly applies to all of reality, and it is thus metaphysically absurd that the universe should pop into being uncaused out of nothing.

Second, why think in any case that the universe is an exception to the rule? As Arthur Schopenhauer once remarked, the Causal Principle is not something we can dismiss like a cab once we have arrived at our desired destination. Sometimes critics will say that while it is impossible for things to come into being uncaused *in* time, things can come into being uncaused *with* time, that is, at a first moment of time. But until the premise's detractors are able to explain the relevant difference between embedded moments of time and a first moment of time, there seems to be no reason to think it more plausible that things can come into being uncaused at a first moment than at a later moment of time. If something cannot come into existence uncaused at *t*, where *t* is preceded by earlier moments of time, why think that if we were to annihilate all moments earlier than *t*, then that thing could come into existence uncaused at *t*? How could the existence of moments earlier than an uncaused event be of any possible relevance to the occurrence of that event?

Indeed, given a pure A-Theory of time, according to which only the present exists, every moment of time is a fresh beginning, qualitatively indistinguishable from a first moment of time, for when any moment is present, earlier moments have passed away and do not exist. Thus, if the universe could exist uncaused at a first moment of time, it could exist uncaused at any moment of time. There just does not seem to be any relevant difference. It follows that if the latter is metaphysically impossible, so is the former.

Third, the objection stifles scientific exploration of cosmological questions. The absolute beginning of time predicted by the standard Friedmann–Lemaître model was the crucial factor in provoking not only the formulation of the Steady-State model of continuous creation, but a whole series of subsequent models all aimed at avoiding the origin *ex nihilo* of our universe predicted by the standard model. Both philosophers and physicists have been deeply disturbed at the prospect of a beginning of time and an absolute origination of the universe and so have felt constrained to posit the existence of causally prior entities such as quantum vacuum states, inflationary domains, imaginary time regimes, and even timelike causal loops. The history of twentieth-century astrophysical cosmology would be considerably different if there were thought to be no need of a causal explanation of the origin of time and the universe.

By contrast, the *mutakallim* cannot be similarly accused of stifling science because it is only with the conceptual analysis of the argument's conclusion that he is able to identify the cause of the universe as a being of religious significance, and the *mutakallim* will in any case welcome attempts to falsify his theistic hypothesis in hopes of corroboration of his preferred hypothesis by the failure of such naturalistic explanations.

1.3. Experiential confirmation

Finally, (1.0) is constantly confirmed in our experience. Scientific naturalists thus have the strongest of motivations to accept it. It is precisely on the strength of the role played by the Causal Principle in science that the naturalist philosopher of science Bernulf Kanitscheider warns, "If taken seriously, the initial singularity is in head-on collision with the most successful ontological commitment that was a guiding line of research since Epicurus and Lucretius," namely, *out of nothing nothing comes*, which Kanitscheider calls "a metaphysical hypothesis which has proved so fruitful in every corner of science that we are surely well-advised to try as hard as we can to eschew processes of absolute origin" (Kanitscheider 1990, p. 344). Doubtless, as mentioned earlier, this same conviction inspired many of the efforts to craft cosmological models aimed at averting the absolute beginning of the universe. If we do have good reasons for accepting the fact of the universe's absolute origin, that fact in no way entails that the Causal Principle is false. *Qua* physicists, we may refuse to draw the inference implied by the two premises on the quite reasonable grounds that one thereby crosses the threshold of science into the realm of metaphysics, but *qua* philosophers, we are free to draw whatever inferences the premises logically imply.

Wesley Morriston opposes two other empirical generalizations to the Causal Principle which he thinks enjoy comparable support but are allegedly incompatible with the *kalam* argument, to wit, (i) *everything that begins to exist has a material cause* and (ii) *causes always stand in temporal relations to their effects* (Morriston 2002, p. 162). Notice, however, that neither of these principles is incompatible with the Causal Principle enunciated in (1.0). They form a consistent triad. Morriston, in truth, offers no defeater at all for the argument's causal premise, taken as an empirical generalization.

What is true is that the conjunction of (1.0) and (2.0) will imply the falsity of at least (i), if not (ii). As putative defeaters of the conclusion (3.0) of the *kalam* argument, however, (i) and (ii) are not compelling. The evidence for (i) is, indeed, impressive. But it is not unequivocal or universal.[94] More importantly, (i) may be simply overridden by the arguments for the finitude of the past. For if it is impossible that there be an infinite regress of past events, it is impossible that the First Cause be a material object, since matter/energy is never quiescent.[95] If (2.0) commends itself to us, then, we must reject (i), and it would

---

[94] Morriston himself takes our own power to control our actions to be the paradigm of causality. But, as J. P. Moreland argues, such control plausibly requires some sort of dualism (Moreland 1998b, pp. 68–91), in which case we have a clear counterexample to the claim that every effect has a material cause. For not only do I cause effects in my physical body, but my mental states are causally connected.

Again, many philosophers believe that immaterial, abstract objects such as numbers, sets, propositions, and so on exist necessarily and eternally. But there are also many abstract objects which seem to exist contingently and noneternally, for example, the equator, the center of mass of the solar system, Beethoven's *Fifth Symphony*, Leo Tolstoy's *Anna Karenina*, and so forth. None of these is a material object. Tolstoy's novel, for example, is not identical to any of its printed exemplars, for these could all be destroyed and replaced by new books. Nor can Beethoven's *Fifth* be identified with any particular series of ink marks or any performance of the symphony. Now these things all began to exist: the equator, for example, did not exist before the earth did. But if they began to exist, did they have a cause or did they come into being out of just nothing? (Notice that it makes sense to ask this question even though these entities are immaterial and so have no material cause.) Many philosophers would say that these objects did indeed have causes: it was Tolstoy, for example, who created *Anna Karenina*. So in cases such as these (and they are legion), we do, indeed, have instances of efficient causation without material causation. We may not agree that such abstract objects really exist; but charity demands that we say that the view defended by our philosophical colleagues is a coherent one.

Moreover, in the realm of physics itself, we have already alluded to the creation of space in the expanding universe as a case of the efficient causation of something real in the absence of any material cause. Moreover, some physicists have taken background fluctuation models to be a counterexample to (i), since if the total energy of the universe is on balance zero, the universe did not borrow any energy from the vacuum and so has no material cause, even though it was spawned by the vacuum as its efficient cause. (See Craig 1998, pp. 332–59).

[95] See Craig (1991, pp. 104–8).

be the height of folly then to go on to reject (1.0) as well. For if coming into being without a material cause seems impossible, coming into being with neither a material nor an efficient cause is doubly absurd.

As for (ii), it appears to be merely an accidental generalization, akin to *Human beings have always lived on the Earth*, which was true until 1968. There does not seem to be anything inherently temporal about a causal relationship. More importantly, however, (ii) is not at all incompatible with the *kalam* argument's conclusion, since its defender may hold that God exists timelessly sans creation and temporally at and subsequent to the moment of creation, so that His act of causing the beginning of the universe is simultaneous with the universe's beginning to exist.[96]

1.4. Objections

J. L. Mackie, in response to the *kalam* cosmological argument, reserved his chief criticism for its first premise. He complains, "there is *a priori* no good reason why a sheer origination of things, not determined by anything, should be unacceptable, whereas the existence of a god [*sic*] with the power to create something out of nothing is acceptable" (Mackie 1982, p. 94). Indeed, he believes *creatio ex nihilo* raises problems: (i) If God began to exist at a point in time, then this is as great a puzzle as the beginning of the universe. (ii) Or if God existed for infinite time, then the same arguments would apply to his existence as would apply to the infinite duration of the universe. (iii) If it be said that God is timeless, then this, says Mackie, is a complete mystery.

Now it is noteworthy that Mackie never *refutes* the principle that whatever begins to exist has a cause. Rather, he simply demands what good reason there is *a priori* to accept it. He writes, "As Hume pointed out, we can certainly conceive an uncaused beginning-to-be of an object; if what we can thus conceive is nevertheless in some way impossible, this still requires to be shown" (Mackie 1982, p. 89; cf. Oppy 2006b, p. 151). But, as many philosophers have pointed out, Hume's argument in no way makes it plausible to think that something could really come into being without a cause. Just because I can imagine an object, say a horse, coming into existence from nothing, that in no way proves that a horse really could come into existence that way. The *mutakallim* plausibly claims that it is ontologically impossible for something to come uncaused from nothing. Does anyone really believe that, however vivid his imagination of such an event, a raging tiger, say, could suddenly come into existence uncaused, out of nothing, in the room right now? The same applies to the universe: if there was absolutely nothing prior to the existence of the universe – no God, no space, no time – how could the universe possibly have come to exist?[97]

---

[96] Craig (2001, pp. 275–80).

[97] Elsewhere, Mackie reveals his true sentiments: "I myself find it hard to accept the notion of self-creation *from nothing*, even *given* unrestricted chance. And how *can* this be given, if there really is nothing?" (Mackie 1982, p. 126). It seems inconceivable that the space-time universe could have come into being without a cause, for in that case there was not even the potentiality of the universe's existence prior to the beginning (since there was no "prior"). But how could the universe become actual if there was not even the potentiality of its existence? It seems more plausible to hold that its potentiality lay in the power of its cause to create it.

In fact, Mackie's appeal to Hume at this point is counterproductive. For Hume himself clearly believed in the Causal Principle. In 1754, he wrote to John Stewart, "But allow me to tell you that I never asserted so absurd a Proposition as *that anything might arise without a cause*: I only maintain'd, that our Certainty of the Falsehood of that Proposition proceeded neither from Intuition nor Demonstration, but from another source" (Grieg 1932, 1: p. 187). Even Mackie confesses, "Still this [causal] principle has some plausibility, in that it is constantly confirmed in our experience (and also used, reasonably, in interpreting our experience)" (Mackie 1982, p. 89). So why not accept the truth of the Causal Principle as plausible and reasonable – at the very least more so than its denial?

Because, Mackie thinks, in this particular case the theism implied by affirming the principle is even more unintelligible than the denial of the principle. It makes more sense to believe that the universe came into being uncaused out of nothing than to believe that God created the universe out of nothing.

But is this really the case? Consider the three problems Mackie raises with *creatio ex nihilo*. Certainly, the *mutakallim* will not hold (i) that God began to exist or (ii) that God has existed for an infinite number of, say, hours, or any other unit of time prior to creation. But what is wrong with (iii), that God is, without creation, timeless? This may be "mysterious" in the sense of "wonderful" or "awe inspiring," but it is not, so far as we can see, unintelligible; and Mackie gives us no reason to think that it is.

Moreover, there is also an alternative that Mackie failed to consider, namely, (iv) prior to creation God existed in an undifferentiated time in which hours, seconds, days, and so forth simply do not exist.[98] Because this time is undifferentiated, it is not incompatible with the *kalam* arguments against an infinite regress of events. This alternative would require us to distinguish between time as it plays a role in physics and time as a metaphysical reality, a distinction famously defended by Isaac Newton on the basis of God's eternal duration independent of any physical measures thereof (Newton 1966, I: p. 6). Even apart from theism, the distinction between time and our physical measures thereof is one that is quite plausible and intuitive, and, given theism, it becomes nearly incumbent, since God could experience a temporal succession of contents of consciousness in the absence of any physical world at all. Mackie, then, is quite unjustified in rejecting the first premise of the argument as not being intuitively obvious, plausible, and reasonable.

So it seems that the Causal Principle enunciated in (1.0) has considerable warrant for us and that putative defeaters of the principle can be undercut or rebutted.

3.0. The Cause of the Universe

From the two premises, it follows logically that the universe has a cause. This is a staggering conclusion, for it implies that the universe was brought into existence by a transcendent reality.

Or does it? As we have seen, Gott and Li have sought to avoid this conclusion by defending the extraordinary hypothesis that *the universe created itself*. Noting that GR allows for the possibility of CTCs, they hypothesize that as we trace the history of the universe back through an original inflationary state, we encounter a region of CTCs prior to inflation (Figure 3.4). According to one possible scenario, a metastable vacuum inflates, producing

---

[98] Such a view is defended by Padgett (1992).

an infinite number of (Big Bang type) bubble universes. In many of these a number of bubbles of metastable vacuum are created at late times by high-energy events. These bubbles usually collapse and form black holes, but occasionally one will tunnel to create an expanding, metastable vacuum or baby universe. One of these expanding, metastable vacuum baby universes "turns out to be the original inflating metastable vacuum we began with." Gott and Li conclude that "the laws of physics may allow the universe to be its own mother" (Gott & Li 1998, p. 023501–1).

We have seen that even on the physical level Gott and Li's model is probably not a plausible account of the universe's origin. But the Gott–Li hypothesis raises even more fundamental metaphysical issues about the nature of time which render their hypothesis either metaphysically impossible or else superfluous. For all instances of causal influence over the past which have been suggested – whether we are talking about CTCs, time travel, tachyonic antitelephones, or whatever – presuppose the truth of the B-Theory of time. For on the A-Theory, at the time at which the effect is present, the cause is future and therefore literally nonexistent. Thus, the effect just comes into being from nothing. Not only is this scenario metaphysically absurd, but it also defeats the claim that the universe is self-caused. Rather the universe just came uncaused from nothing.

Thus, the Gott–Li hypothesis presupposes the B-Theory of time. But if one presupposes such a view of time, then Gott and Li's hypothesis becomes superfluous. For, as we have seen, on a B-Theory of time the universe never truly comes into being at all.[99] The whole four-dimensional space-time manifold just exists tenselessly, and the universe has a beginning only in the sense that a meter stick has a beginning prior to the first centimeter. Although the space-time manifold is intrinsically temporal in that one of its four dimensions is time, nonetheless it is extrinsically timeless, in that it does not exist in an embedding hypertime but exists tenselessly, neither coming into nor going out of being. The four-dimensional space-time manifold is in this latter sense eternal. Thus, there is no need for the device of causal loops or CTCs at the beginning to explain how it came into being.

Given the truth of the A-Theory of time, the idea that the universe is self-created, that is to say, that it brought itself into being via CTCs, is metaphysically impossible because it reduces to the notion that the universe sprang into existence uncaused out of nothing. The universe, then, must have an ultramundane cause.

Properties of the First Cause

Conceptual analysis of what it is to be a cause of the universe enables us to recover a number of striking properties which this ultramundane cause must possess and which are of theological significance. For example, the cause must be uncaused, since, as we have seen, an

---

[99] This is the salient point of Grünbaum's most recent salvo against the inference to a First Cause of the origin of the universe (Grünbaum 2000). As a B-Theorist, Grünbaum does not believe that the universe ever came into being, even if it had a first temporal interval. As he elsewhere writes, "coming *into* being (or 'becoming') is *not* a property of *physical* events themselves but only of human or conscious awareness of these events" (Grünbaum 1967, p. 153). What Grünbaum fails to see, however, is that the claim that an absolute beginning of the universe entails that the universe came into being is rooted, not in the presupposition of the so-called Spontaneity of Nothingness, but in an A-Theory of time.

infinite regress of causes is impossible. One could, of course, arbitrarily posit a plurality of causes in some sense prior to the origin of the universe, but ultimately, if the philosophical *kalam* arguments are sound, this causal chain must terminate in a cause which is absolutely first and uncaused. There being no reason to perpetuate the series of events beyond the origin of the universe, Ockham's Razor, which enjoins us not to posit causes beyond necessity, strikes such further causes in favor of an immediate First Cause of the origin of the universe. The same principle dictates that we are warranted in ignoring the possibility of a plurality of uncaused causes in favor of assuming the unicity of the First Cause.

This First Cause must also be beginningless, since by contraposition of Premise (1.0), whatever is uncaused does not begin to exist. Moreover, this cause must be changeless, since, once more, an infinite temporal regress of changes cannot exist. We should not be warranted, however, in inferring the immutability of the First Cause, since immutability is a modal property, and from the Cause's changelessness we cannot infer that it is incapable of change. But we can know that the First Cause is changeless, at least insofar as it exists sans the universe. From the changelessness of the First Cause, its immateriality follows. For whatever is material involves incessant change on at least the molecular and atomic levels, but the uncaused First Cause exists in a state of absolute changelessness. Given some relational theory of time, the Uncaused Cause must therefore also be timeless, at least sans the universe, since in the utter absence of events time would not exist. It is true that some philosophers have argued persuasively that time could continue to exist even if all events were to cease (Shoemaker 1969; Forbes 1993), but such arguments are inapplicable in the case at hand, where we are envisioning, not the cessation of events, but the utter absence of any events whatsoever. In any case, the timelessness of the First Cause sans the universe can be more directly inferred from the finitude of the past. Given that time had a beginning, the cause of the beginning of time must be timeless.[100] It follows that this Cause must also be spaceless, since it is both immaterial and timeless, and no spatial entity can be both immaterial and timeless. If an entity is immaterial, it could exist in space only in virtue of being related to material things in space; but then it could not be timeless, since it undergoes extrinsic change in its relations to material things. Hence, the uncaused First Cause must transcend both time and space and be the cause of their origination. Such a being must be, moreover, enormously powerful, since it brought the entirety of physical reality, including all matter and energy and space-time itself, into being without any material cause.

Finally, and most remarkably, such a transcendent cause is plausibly taken to be personal. Three reasons can be given for this conclusion. First, as Richard Swinburne (1991, pp. 32–48) points out, there are two types of causal explanation: scientific explanations in terms of laws and initial conditions and personal explanations in terms of agents and their volitions. For example, in answer to the question, "Why is the kettle boiling?" we might be told, "The heat of the flame is being conducted via the copper bottom of the kettle to the water, increasing the kinetic energy of the water molecules, such that they vibrate so violently that they break the surface tension of the water and are thrown off in the form of steam." Or alternatively, we might be told, "I put it on to make a cup of tea. Would you like some?" The first provides a scientific explanation, the second a personal explanation. Each

---

[100] This needs some qualification, since the *kalam* argument strictly demonstrates only that metric time had a beginning. Perhaps the cause exists changelessly in an undifferentiated time in which temporal intervals cannot be distinguished. On this view, God existed literally before creation, but there was no moment, say, 1 hour or 1 million years before creation.

is a perfectly legitimate form of explanation; indeed, in certain contexts it would be wholly inappropriate to give one rather than the other. Now a first state of the universe *cannot* have a scientific explanation, since there is nothing before it, and therefore, it cannot be accounted for in terms of laws operating on initial conditions. It can only be accounted for in terms of an agent and his volitions, a personal explanation.

Second, the personhood of the First Cause is already powerfully suggested by the properties which have been deduced by means of our conceptual analysis. For there appear to be only two candidates which can be described as immaterial, beginningless, uncaused, timeless, and spaceless beings: either abstract objects or an unembodied mind. Abstract objects such as numbers, sets, propositions, and properties are very typically construed by philosophers who include such things in their ontology as being precisely the sort of entities which exist necessarily, timelessly, and spacelessly. Similarly, philosophers who hold to the possibility of disembodied mind would describe such mental substances as immaterial and spaceless, and there seems no reason to think that a Cosmic Mind might not also be beginningless and uncaused. No other candidates which could be suitably described as immaterial, beginningless, uncaused, timeless, and spaceless beings come to mind. Nor has anyone else, to our knowledge, suggested any other such candidates. But no sort of abstract object can be the cause of the origin of the universe, for abstract objects are not involved in causal relations. Even if they were, since they are not agents, they cannot volitionally exercise a causal power to do anything. If they were causes, they would be so, not as agents, but as mindless events or states. But they cannot be event-causes, since they do not exist in time and space. Even if we allow that some abstract objects exist in time (e.g. propositions which change their truth-value in virtue of the tense of the sentences which express them), still, in view of their abstract nature, it remains utterly mysterious how they could be causally related to concrete objects so as to bring about events, including the origin of the universe. Nor can they be state-causes of states involving concrete objects, for the same reason, not to mention the fact that in the case at hand we are not talking about state-state causation (i.e. the causal dependence of one state on another), but what would amount to state-event causation (namely, the universe's coming into being because of the state of some abstract object(s)), which seems impossible. Thus, the cause of the universe must be an unembodied mind.

Third, this same conclusion is also implied by the fact that only personal, free agency can account for the origin of a first temporal effect from a changeless cause. We have concluded that the beginning of the universe was the effect of a First Cause. By the nature of the case, that cause cannot have any beginning of its existence nor any prior cause. Nor can there have been any changes in this cause, either in its nature or operations, prior to the beginning of the universe. It just exists changelessly without beginning, and a finite time ago it brought the universe into existence. Now this is exceedingly odd. The cause is in some sense eternal, and yet the effect which it produced is not eternal but began to exist a finite time ago. How can this be? If the necessary and sufficient conditions for the production of the effect are eternal, then why is not the effect eternal? How can all the causal conditions sufficient for the production of the effect be changelessly existent and yet the effect not also be existent along with the cause? How can the cause exist without the effect?

One might say that the cause came to exist or changed in some way just prior to the first event. But then the cause's beginning or changing would be the first event, and we must ask all over again for its cause. And this cannot go on forever, for we know that a beginningless series of events cannot exist. There must be an absolutely first event, before which there was no change, no previous event. We know that this first event must have been caused. The question is: How can a first event come to exist if the cause of that event exists changelessly and eternally? Why is the effect not coeternal with its cause?

The best way out of this dilemma is agent causation, whereby the agent freely brings about some event in the absence of prior determining conditions. Because the agent is free, he can initiate new effects by freely bringing about conditions which were not previously present. For example, a man sitting changelessly from eternity could freely will to stand up; thus, a temporal effect arises from an eternally existing agent. Similarly, a finite time ago a Creator endowed with free will could have freely brought the world into being at that moment. In this way, the Creator could exist changelessly and eternally but choose to create the world in time. By "choose" one need not mean that the Creator changes his mind about the decision to create but that he freely and eternally intends to create a world with a beginning. By exercising his causal power, he therefore brings it about that a world with a beginning comes to exist.[101] So the cause is eternal, but the effect is not. In this way, then, it is possible for the temporal universe to have come to exist from an eternal cause: through the free will of a personal Creator.

A conceptual analysis of what properties must be possessed by an ultramundane First Cause thus enables us to recover a striking number of the traditional divine attributes. An analysis of what it is to be cause of the universe reveals that:

4.0. If the universe has a cause, then an uncaused, personal Creator of the universe exists, who sans the universe is beginningless, changeless, immaterial, timeless, spaceless, and enormously powerful.

From (3.0) and (4.0), it follows that:

5.0. Therefore, an uncaused, personal Creator of the universe exists, who sans the universe is beginningless, changeless, immaterial, timeless, spaceless, and enormously powerful.

This, as Thomas Aquinas was wont to remark, is what everybody means by "God."

---

[101] Such an exercise of causal power plausibly brings God into time, if He was not temporal already. As Moreland explains, in the case of personal causal explanations, the salient factors are the existence of an agent with his relevant properties and powers, the agent's intention to bring about some result, an exercise of the agent's causal powers, and in some cases a description of the relevant action plan. So "a personal explanation (divine or otherwise) of some basic result R brought about intentionally by person P where this bringing about of R is a basic action A will cite the intention I of P that R occur and the basic power B that P exercised to bring about R" (Moreland 1998b, p. 75; see further Moreland 1998a). Notice that it is insufficient for P to have merely the intention and power to bring about R. There must also be a basic action on the part of P, an undertaking or endeavoring or exercise of P's causal powers. Thus, it is insufficient to account for the origin of the universe by citing simply God, His timeless intention to create a world with a beginning, and His power to produce such a result. There must be an exercise of His causal power in order for the universe to be created. That entails, of course, an intrinsic change on God's part which brings Him into time at the moment of creation. For that reason He must be temporal since creation, even if He is timeless sans creation.

We should not say that in agent causation the agent causes his causing of some effect. Partisans of agent causation typically say that the agent's causing some effect is not an event requiring a cause, either because it is not itself an event, but just a way of describing an agent's causing an event, or if it is an event, then it is not further caused (O'Connor 2000, chap. 3). Neither alternative requires revision of Premise (1.0), which concerns, not events, but substances which come into being.

Objections

Certain thinkers have objected to the intelligibility of this conclusion. For example, Adolf Grünbaum has marshaled a whole troop of objections against inferring God as the Creator of the universe (Grünbaum 1990b). As these are very typical, a brief review of his objections should be quite helpful.

Grünbaum's objections fall into three groups. Group I seeks to cast doubt upon the concept of "cause" in the argument: (1) When we say that everything has a cause, we use the word "cause" to mean something that transforms previously existing materials from one state to another. But when we infer that the universe has a cause, we must mean by "cause" something that creates its effect out of nothing. Since these two meanings of "cause" are not the same, the argument is guilty of equivocation and is thus invalid. (2) It does not follow from the necessity of there being a cause that the cause of the universe is a conscious agent. (3) It is logically fallacious to infer that there is a *single* conscious agent who created the universe.

But these objections do not seem to present any insuperable difficulties: (1) The univocal concept of "cause" employed throughout the argument is the concept of something which brings about or produces its effects.[102] Whether this production involves transformation of already existing materials or creation out of nothing is an incidental question. Thus, the charge of equivocation is groundless. (2) The personhood of the cause does not follow from the two premises of the cosmological argument proper, but rather from a conceptual analysis of the notion of a First Cause of the beginning of the universe, as we have seen. (3) The inference to a single cause of the origin of the universe seems justified in light of the principle, commonly accepted in science, that one should not multiply causes beyond necessity. One is justified in inferring only causes such as are necessary to explain the effect in question; positing any more would be gratuitous.

The objections of Group II relate the notion of causality to the temporal series of events: (1) Causality is logically compatible with an infinite, beginningless series of events. (2) If everything has a cause of its existence, then the cause of the universe must also have a cause of its existence.

Both of these objections, however, seem to be based on misunderstandings. (1) It is not the concept of causality which is incompatible with an infinite series of past events. Rather the incompatibility, as we have seen, is between the notion of an actually infinite number of things and the series of past events. The fact that causality has nothing to do with it may be seen by reflecting on the fact that the philosophical arguments for the beginning of the universe would work even if the events were all spontaneous, causally unconnected events. (2) The argument does not presuppose that everything has a cause. Rather the operative Causal Principle is that *whatever begins to exist has a cause*. Something that exists eternally and, hence, without a beginning would not need to have a cause. This is not special pleading for God, since the atheist has always maintained the same thing about the universe: it is beginningless and uncaused.

Group III objections are aimed at the alleged claim that creation from nothing surpasses all understanding: (1) If creation out of nothing is incomprehensible, then it is irrational to believe in such a doctrine. (2) An incomprehensible doctrine cannot explain anything.

---

[102] That is to say, an efficient cause. Alternatively, we could leave the question of material causation open by taking "cause" to mean either an efficient or a material cause. Then our conceptual analysis of what it is to be a cause of the universe will eliminate the alternative that the cause is a material cause, leaving us with an ultramundane efficient cause.

But with regard to (1), creation from nothing is not incomprehensible in Grünbaum's sense. By "incomprehensible," Grünbaum appears to mean "unintelligible" or "meaningless." But the statement that a finite time ago a transcendent cause brought the universe into being out of nothing is clearly a meaningful statement, not mere gibberish, as is evident from the very fact that we are debating it. We may not understand *how* the cause brought the universe into being out of nothing, but such efficient causation without material causation is not unprecedented, as we have seen, and it is even more incomprehensible, in this sense, how the universe could have popped into being out of nothing without *any* cause, material or efficient. One cannot avert the necessity of a cause by positing an absurdity. (2) The doctrine, being an intelligible statement, obviously does constitute a purported explanation of the origin of the universe. It may be a metaphysical rather than a scientific explanation, but it is no less an explanation for that.

Grünbaum has one final objection against inferring a cause of the origin of the universe: the cause of the Big Bang can be neither *after* the Big Bang (since backward causation is impossible) nor *before* the Big Bang (since time begins at or after the Big Bang). Therefore, the universe's beginning to exist cannot have a cause (Grünbaum 1990a, 1991; cf. Craig 1994a). But this argument pretty clearly confronts us with a false dilemma. For why could not God's creating the universe be *simultaneous* (or coincident) with the Big Bang? God may be conceived to exist timelessly (or in an undifferentiated time) without the universe and in time from the moment of creation. Perhaps an analogy from physical cosmology will be illuminating. The initial Big Bang singularity is not considered to be part of physical time, but to constitute a boundary to time. Nevertheless, it is causally connected to the universe. In an analogous way, we could say that God's timeless eternity is, as it were, a boundary of time which is causally, but not temporally, prior to the origin of the universe. It seems, therefore, that it is not only coherent but also plausible in light of the *kalam* cosmological argument that God existing changelessly alone without creation is timeless and that He enters time at the moment of creation in virtue of His causal relation to the temporal universe. The time of the first event would be not only the first time at which the universe exists but also, technically, the first time at which God exists, since sans the universe God exists timelessly.[103] The moment of creation is, as it were, the moment at which God enters time. His act of creation is thus simultaneous with the origination of the universe.

Conclusion

The first premise of the *kalam* cosmological argument is obviously more plausibly true than its contradictory. Similarly, in light of both philosophical argument and scientific evidence, its second premise, although more controversial, is again more plausibly true than its negation. The conclusion of the argument involves no demonstrable incoherence and, when subjected to conceptual analysis, is rich in theological implications. On the basis of the *kalam* cosmological argument, it is therefore plausible that an uncaused, personal Creator of the universe exists, who sans the universe is beginningless, changeless, immaterial, timeless, spaceless, and enormously powerful.

---

[103] Leftow (1991, p. 269, cf. p. 201) captures this idea nicely. Senor has dubbed such a model of divine eternity "accidental temporalism" (Senor 1993, p. 88). See further Craig (2001, pp. 256–80).

References

Aguirre, A. and Gratton, S. (2002) Steady state eternal inflation. *Physical Review D* 65, 083507, preprint http://arxiv.org/abs/astro-ph/0111191 (accessed July 11, 2008).

Al-Ghazali (1962) *Kitab al-Iqtisad fi'l-Iqtiqad.* Ankara, Turkey: University of Ankara Press.

Al-Ghazali (1963) *Tahafut al-Falasifah*. Trans. S. A. Kamali. Lahore: Pakistan Philosophical Congress.

Al-Idji (1971) *Mawakif*. Cited in *Encyclopaedia of Islam*, s.v. "Ilm al-Kalam" (by L. Gardet).

Ashtekar, A., Pawlowski, T., and Singh, P. (2006) Quantum nature of the Big Bang. *Physical Review D* 74, 084003, preprint: http://arxiv.org/abs/gr-qc/0602086v2.

Balaguer, M. (1998) *Platonism and Anti-Platonism in Mathematics.* New York: Oxford University Press.

Balaguer, M. (2001) A theory of mathematical correctness and mathematical truth. *Pacific Philosophical Quarterly* 82, 87–114.

Banks, T. and Fischler, W. (2002) Black crunch. http://arxiv.org/abs/hep-th/0212113v1 (accessed July 11, 2008).

Barrow, J. and Dabrowski, M. (1995) Oscillating universes. *Monthly Notices of the Royal Astronomical Society* 275, 850–62.

Barrow, J. D. (2005) *The Infinite Book*. New York: Pantheon Books.

Barrow, J. D. and Tipler, F. (1986) *The Anthropic Cosmological Principle*. Oxford: Clarendon Press.

Baum, L. and Frampton, P. H. (2006) Deflation at turnaround for oscillatory cosmology. http://arxiv.org/abs/astro-ph/0608138 (accessed July 11, 2008).

Baum, L. and Frampton, P. H. (2007) Turnaround in cyclic cosmology. *Physical Review Letters* 98, 071301, p. 1, preprint: http://arxiv.org/abs/hep-th/0610213.

Belinsky, V. A., Khalatnikov, I. M., and Lifshitz, E. M. (1970) Oscillatory approach to a singular point in the relativistic cosmology. *Advances in Physics* 19, 525–73.

Benardete, J. A. (1964) *Infinity: An Essay in Metaphysics*. Oxford: Clarendon Press.

Bojowald, M. (2005) Original questions. *Nature* 436, 920–1.

Bojowald, M., Date, G., and Hossain, G. M. (2004) The Bianchi IX model in loop quantum cosmology. *Classical and Quantum Gravity* 21, 3541, preprint: http://arxiv.org/abs/gr-qc/0404039.

Bolzano, B. (1950) *Paradoxes of the Infinite*. Trans. Fr. Prihonsky. London: Routledge & Kegan Paul.

Boolos, G. (1986–7) Saving Frege from contradiction. *Proceedings of the Aristotelian Society* 87, 137–51.

Borde, A., Guth, A., and Vilenkin, A. (2003) Inflationary spacetimes are not past-complete. *Physical Review Letters* 90, 151301, preprint: http://arxiv.org/abs/gr-qc/0110012.

Bousso, R. and Freivogel, B. (2007) A paradox in the global description of the multiverse. *Journal of High Energy Physics* 6, preprint: http://arxiv.org/PS_cache/hep-th/pdf/0610/0610132v2.pdf.

Cantor, G. (1915) *Contributions to the Founding of the Theory of Transfinite Numbers.* Trans. and Intro. P. E. B. Jourdain. New York: Dover.

Chihara, C. S. (1990). *Constructibility and Mathematical Existence.* Oxford: Clarendon Press.

Chihara, C. S. (2004) *A Structural Account of Mathematics.* Oxford: Clarendon Press.

Chihara, C. S. (2005) Nominalism. In S. Shapiro (ed.), *The Oxford Handbook of Philosophy of Mathematics and Logic*, 483–514. Oxford: Oxford University Press.

Conway, D. A. (1984) "It would have happened already": on one argument for a first cause. *Analysis* 44, 159–66.

Coule, D. H. (2005) Quantum cosmological models. *Classical and Quantum Gravity* 22, R125–2308, p. 31, preprint: http://arxiv.org/abs/gr-qc/0412026v3.

Craig, W. L. (1979) *The Kalam Cosmological Argument.* London: Macmillan.

Craig, W. L. (1980) *The Cosmological Argument from Plato to Leibniz*. London: Macmillan & Co.

Craig, W. L. (1985) Critical notice of *Time, Creation and the Continuum*, by Richard Sorabji. *International Philosophical Quarterly* 25, 319–26.

Craig, W. L. (1991) The *Kalam* Cosmological argument and the hypothesis of a quiescent universe. *Faith and Philosophy* 8, 104–8.

Craig, W. L. (1992) The origin and creation of the universe: a reply to Adolf Grünbaum. *British Journal for the Philosophy of Science* 43, 233–40.

Craig, W. L. (1994a) Prof. Grünbaum on creation. *Erkenntnis* 40, 325–41.

Craig, W. L. (1994b) A response to Grünbaum on creation and Big Bang cosmology. *Philosophia Naturalis* 31, 237–49.

Craig, W. L. (1998) Design and the cosmological argument. In W. Dembski (ed.), *Mere Creation*, 332–59. Downers Grove, Ill.: Inter-Varsity.

Craig, W. L. (2000a) *The Tensed Theory of Time: A Critical Examination.* Synthese Library 293. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Craig, W. L. (2000b) *The Tenseless Theory of Time: A Critical Examination*. Synthese Library 294. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Craig, W. L. (2000c) The extent of the present. *International Studies in the Philosophy of Science* 14, 165–85.

Craig, W. L. (2001) *God, Time and Eternity*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Craig, W. L. (2008) Critical notice of *Philosophical Perspectives on Infinity*, by G. Oppy. *Philosophia Christi* 10, 435–52.

Craig, W. L. and Moreland, J. P. (2003) *Philosophical Foundations of a Christian Worldview.* Downer's Grove, Ill.: Inter-Varsity Press.

Craig, W. L. and Sinnott-Armstrong, W. (2003) *God?: A Debate between a Christian and an Atheist.* Oxford: Oxford University Press.

Craig, W. L., and Smith, Q. (1993) *Theism, Atheism, and Big Bang Cosmology*. Oxford: Clarendon Press.

Dales, R. C. (1990) *Medieval Discussions of the Eternity of the World*. Brill's Studies in Intellectual History 18. Leiden, the Netherlands: E. J. Brill.

Damour, T. and Henneaux, M. (2000) Chaos in superstring cosmology. *Physical Review Letters* 85, 920–3, preprint: http://aps.arxiv.org/abs/hep-th/0003139.

Davidson, H. A. (1987) *Proofs for Eternity, Creation, and the Existence of God in Medieval Islamic and Jewish Philosophy*. New York: Oxford University Press.

Davies, P. (1978) Spacetime singularities in cosmology. In J. T. Fraser (ed.), *The Study of Time III*, 78–9. Berlin: Springer Verlag.

Davies, P. (1983) *God and the New Physics.* New York: Simon & Schuster.

Dedekind, R. (1963) The nature and meaning of numbers. In Richard Dedekind (ed.), W. W. Beman (trans.), *Essays on the Theory of Numbers*, 29–115. New York: Dover.

Dretske, F. (1965) Counting to infinity. *Analysis* 25, 99–101.

Earman, J. and Mosterin, J. (1999) A critical look at inflationary cosmology. *Philosophy of Science* 66, 1–49.

Eddington, A. S. (1948) *The Nature of the Physical World*. New York: Macmillan.

Einstein, A. (1917) Cosmological considerations on the general theory of relativity. With notes by A. Sommerfeld, translated by W. Perrett and J. B. Jefferey, *The Principle of Relativity*, 177–88. Reprint, New York: Dover Publications, 1952.

Ellis, G. F. R. and Maartens, R. (2004) The emergent universe: inflationary cosmology with no singularity. *Classical and Quantum Grav*ity 21, 223, preprint: http://arxiv.org/abs/gr-qc/0211082.

Ellis, G. F. R., Kirchner, U., and Stoeger, W. R. (2003) Multiverses and physical cosmology. http://arXiv.org/abs/astro-ph/0305292 (accessed July 11, 2008).

Ellis, G. F. R., Murugan, J., and Tsagas, C. G. (2004) The emergent universe: an explicit construction. *Classical and Quantum Gravity* 27, 233, preprint: http://arxiv.org/abs/gr-qc/0307112.

Feynman, R. (1985) *QED, The Strange Theory of Light and Matter*. Princeton, NJ: Princeton University Press.

Forbes, G. (1993) Time, events, and modality. In R. Le Poidevin and M. MacBeath (eds.), *The Philosophy of Time*, 80–95. Oxford: Oxford University Press.

Fraenkel, A. A. (1961) *Abstract Set Theory.* Studies in Logic and the Foundations of Mathematics. 2nd rev. edn. Amsterdam, the Netherlands: North-Holland.

Frampton, P. (2007a) Cyclic universe and infinite past, preprint http://arxiv.org/abs/0705.2730 (accessed July 11, 2008).

Frampton, P. (2007b) Comment on "Can black holes be torn up by a phantom in cyclic cosmology?" by X. Zhang, preprint http://arxiv.org/abs/0709.1630 (accessed July 11, 2008).

Frampton, P. H. and Takahashi, T. (2003) The fate of dark energy. *Physical Letters B* 557, 135, preprint: http://arxiv.org/abs/astro-ph/0211544.

Frampton, P. H. and Takahashi, T. (2004) Bigger rip with no dark energy. *Astropart. Phys.* 22, 307, preprint: http://arxiv.org/abs/astro-ph/0405333.

Friedmann, A. (1922) Über die Krümmung des Raumes. *Zeitschrift für Physik* 10, 377–86.

Gale, R. (2007) The failure of classic theistic arguments. In M. Martin (ed.), *The Cambridge Companion to Atheism*, Cambridge Companions to Philosophy, 86–101. Cambridge: Cambridge University Press.

Gamow, G. (1946) *One, Two, Three, Infinity*. London: Macmillan.

Gott, J. R. III, Gunn, J. E., Schramm, D. N., and Tinsley, B. M. (1976) Will the universe expand forever? *Scientific American*, March, 65.

Gott, J. R. III and Li, L.-X. (1998) Can the universe create itself? *Physical Review D* 58: 2, 023501–1.

Greene, B. (2004) *The Fabric of the Cosmos*. New York: Alfred A. Knopf.

Grieg, J. (ed.) (1932) *The Letters of David Hume*, 2 vols. Oxford: Clarendon Press.

Grünbaum, A. (1950–1) Relativity and the atomicity of becoming. *Review of Metaphysics* 4, 143–86.

Grünbaum, A. (1967) The anisotropy of time. In T. Gold (ed.), *The Nature of Time*, 49–186. Ithaca, NY: Cornell University Press.

Grünbaum, A. (1973) *Philosophical Problems of Space and Time*, 2nd ed. Boston Studies for the Philosophy of Science 12. Dordrecht, the Netherlands: D. Reidel.

Grünbaum, A. (1990a) Pseudo-creation of the Big Bang. *Nature* 344, 821–2.

Grünbaum, A. (1990b) The pseudo-problem of creation in physical cosmology. In J. Leslie (ed.), *Physical Cosmology and Philosophy*, Philosophical Topics, 92–112. New York: Macmillan.

Grünbaum, A. (1991) Creation as a pseudo-explanation in current physical cosmology. *Erkenntnis* 35, 233–54.

Grünbaum, A. (2000) A new critique of theological interpretations of physical cosmology. *British Journal for the Philosophy of Science* 51, 1–43.

Guth, A. H. (1981) Inflationary universe: a possible solution to the horizon and flatness problems. *Physical Review D* 23: 2, 347–56.

Hackett, S. (1957) *The Resurrection of Theism*. Chicago: Moody Press.

Hawking, S. (1988) *A Brief History of Time*. New York: Bantam Books.

Hawking, S. (1992) The chronology protection conjecture. *Physical Review D* 46, 603–11.

Hawking, S. and Hartle, J. (1983) The wave function of the universe. *Physical Review D* 28: 12, 2960–75.

Hawking, S. and Penrose, R. (1970) The singularities of gravitational collapse and cosmology. *Proceedings of the Royal Society of London A* 314, 529–48.

Hawking, S. and Penrose, R. (1996) *The Nature of Space and Time*. Princeton, NJ: Princeton University Press.

Hawthorne, J. (2000) Before-effect and Zeno causality. *Noûs* 34, 622–33.

Hazen, A. (1993) Slicing it thin. *Analysis* 53, 189–92.

Hellman, G. (1989) *Mathematics without Numbers: Towards a Modal-Structural Interpretation*. Oxford: Oxford University Press.

Hellman, G. (2001) Three varieties of mathematical structuralism. *Philosophia Mathermatica* 3, 129–57.

Hellman, G. (2005) Structuralism. In S. Shapiro (ed.), *The Oxford Handbook of Philosophy of Mathematics and Logic*, 536–62. Oxford: Oxford University Press.

Hilbert, D. (1964) On the infinite. In P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics*, 183–201. Englewood Cliffs, N J: Prentice-Hall.

Hiscock, W. A. (2000) Quantized fields and chronology protection, preprint: http://arxiv.org/abs/gr-qc/0009061v2 (accessed July 11, 2008).

Hubble, E. (1929) A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences* 15, 168–73.

Hume, D. (1978) *A Treatise of Human Nature*, 2nd edn. Edited and with notes by P. H. Nidditch. Oxford: Clarendon Press.

Kanitscheider, B. (1990) Does physical cosmology transcend the limits of naturalistic reasoning? In P. Weingartner and G. Doen (eds.), *Studies on Marco Bunge's "Treatise,"* 337–50. Amsterdam: Rodopi.

Kasner, E. and Newman, J. (1940) *Mathematics and the Imagination*. New York: Simon & Schuster.

Leftow, B. (1991) *Time and Eternity.* Cornell Studies in Philosophy of Religion. Ithaca, NY: Cornell University Press.

Lemaître, G. (1927) Un univers homogène de masse constante et de rayon croissant, rendant compte de la vitesse radiale des nébuleuses extragalactiques. *Annales de la Société scientifique de Bruxelles* 47, 49–59.

Linde, A. (1998) The self-reproducing inflationary universe. *Scientific American*, May, 103.

Linde, A. (2005) Inflation and string cosmology. *Journal of Physics: Conference Series 24* 151–60, preprint: http://arxiv.org/PS_cache/hep-th/pdf/0503/0503195v1.pdf.

Mackie, J. L. (1982) *The Miracle of Theism*. Oxford: Clarendon Press.

Moore, A. W. (1990) *The Infinite*. London: Routledge.

Moreland, J. P. (1998a) Libertarian agency and the Craig/Grünbaum debate about theistic explanation of the initial singularity. *American Catholic Philosophical Quarterly* 71, 539–54.

Moreland, J. P. (1998b) Searle's biological naturalism and the argument from consciousness. *Faith and Philosophy* 15, 68–91.

Morriston, W. (2002) Craig on the actual infinite. *Religious Studies* 38, 147–66.

Naber, G. L. (1988) *Spacetime and Singularities: An Introduction*. Cambridge: Cambridge University Press.

Nesvizhevsky, V., Börner, H. G., Petukhorv, A. K., et al. (2001) Quantum states of neutrons in the Earth's gravitational field. *Nature* 415, 297–99.

Newton, I. (1966) *Sir Isaac Newton's "Mathematical Principles of Natural Philosophy" and his "System of the World."* 2 vols. Trans. A. Motte. Rev. Florian Cajori. Los Angeles: University of California Press.

O'Connor, T. (2000) *Persons and Causes*: *The Metaphysics of Free Will.* Oxford: Oxford University Press.

Oppy, G. (2006a) *Philosophical Perspectives on Infinity.* Cambridge: Cambridge University Press.

Oppy, G. (2006b) *Arguing about Gods*. Cambridge: Cambridge University Press.

Overbye, D. (2006) 9 billion-year-old "dark energy" reported. *The New York Times*, January 15, 2008.

Padgett, A. (1992) *God*, *Eternity*, *and the Nature of Time*. New York: St. Martin's.

Pagels, H. (1985) *Perfect Symmetry: The Search for the Beginning of Time*. New York: Simon & Schuster.

Penrose, R. (2005) *The Road to Reality*. New York: Alfred A. Knopf.

Philoponus, J. (1987) *Against Aristotle, on the Eternity of the World*. Trans. C. Wildberg. London: Duckworth.

Philoponus, J. and Simplicius (1991) *Place, Void, and Eternity*. Trans. D. Furley and C. Wildberg. Ithaca, NY: Cornell University Press.

Routledge Encyclopedia of Philosophy (1998) s.v. "God, Arguments for the Existence of," (by A. Plantinga).

Rucker, R. v. B. (1980) The actual infinite. *Speculations in Science and Technology* 3, 63–76.

Russell, B. (1937) *The Principles of Mathematics*, 2nd edn. London: George Allen & Unwin.

Senor, T. D. (1993) Divine temporality and creation *ex nihilo*. *Faith and Philosophy* 10, 86–92.

Shoemaker, S. (1969) Time without change. *Journal of Philosophy*. 66, 363–81.

Small, R. (1986) Tristram Shandy's last page. *British Journal for the Philosophy of Science* 37, 213–6.

Smith, Q. (2007) Kalam cosmological arguments for atheism. In M. Martin (ed.), *The Cambridge Companion to Atheism*, Cambridge Companions to Philosophy, 182–98. Cambridge: Cambridge University Press.

Sobel, J. H. (2004) *Logic and Theism: Arguments for and against Beliefs in God*. Cambridge: Cambridge University Press.

Sobel, J. H. (2007) Online revisions of chapter V of *Logic and Theism*. http://www.scar.utoronto.ca/~sobel/OnL_T/ (accessed March 7, 2008).

Sorabji, R. (1983) *Time, Creation and the Continuum*. Ithaca, NY: Cornell University Press.

Stanford Encyclopedia of Philosophy (2004a) s.v. "Cosmological Argument" (by B. Reichenbach).

Stanford Encyclopedia of Philosophy (2004b) s.v. "Platonism in Metaphysics" (by M. Balaguer), http://plato.stanford.edu/archives/sum2004/entries/platonism/ (accessed March 7, 2008).

Steinhardt, P. and Turok, N. (2005) The cyclic model simplified. *New Astronomy Reviews* 49, 43–57, preprint: http://arxiv.org/abs/astro-ph/0404480.

Swinburne, R. (1991) *The Existence of God*, rev. edn. Oxford: Clarendon Press.

Thorne, K. (1994) *Black Holes and Time Warps*. New York: W.W. Norton & Co.

Vaas, R. (2004) Time before time: classifications of universes in contemporary cosmology, and how to avoid the antinomy of the beginning and eternity of the world. In W. Löffler and P. Weingartner (eds.), *Knowledge and Belief. Papers of the 26th International Wittgenstein Symposium*, 351–3. Kirchberg am Wechsel: Austrian Ludwig Wittgenstein Society, preprint: http://arXiv.org/abs/physics/0408111.

Veneziano, G. (1998) A simple/short introduction to pre-big-bang physics/cosmology, preprint: http://arxiv.org/abs/hep-th/9802057 (accessed July 11, 2008).

Veneziano, G. (2004) The myth of the beginning of time. *Scientific American*, May, 63.

Veneziano G. and Gasperini, M. (2002) The pre big bang scenario in string cosmology. *Physics Reports* 373, 1, preprint: http://arxiv.org/abs/hep-th/0207130.

Vilenkin, A. (1982) Creation of universes from nothing. *Physics Letters B* 117, 25–28.

Vilenkin, A. (1994) Approaches to quantum cosmology. *Physical Review D* 50, 2581–94, preprint: http://lanl.arxiv.org/abs/gr-qc/9403010v1.

Vilenkin, A. (2002) Quantum cosmology and eternal inflation. *The Future of Theoretical Physics and Cosmology*, Proceedings of the Conference in Honor of Stephen Hawking's 60th birthday, preprint: http://arxiv.org/abs/gr-qc/0204061 (accessed July 11, 2008).

Vilenkin, A. (2006) *Many Worlds in One*. New York: Hill & Wang.

Wald, R. M. (1984) *General Relativity*. Chicago, Ill.: The University of Chicago Press.

Walton, K. L. (1990) *Mimesis as Make–Believe*. Cambridge, Mass.: Harvard University Press.

Wheeler, J. A. (1980) Beyond the hole. In H. Woolf (ed.), *Some Strangeness in the Proportion*, 354–75. Reading, Mass.: Addison-Wesley.

Whitrow, G. J. (1980) *The Natural Philosophy of Time*, 2nd edn. Oxford: Clarendon Press.

Wittgenstein, L. (1976) *Lectures on the Foundations of Mathematics*. Ed. C. Diamond. Sussex, England: Harvester.

Wolfson, H. A. (1966) Patristic arguments against the eternity of the world. *Harvard Theological Review* 59, 354–67.

Wolfson, H. A. (1976) *The Philosophy of the Kalam*. Cambridge, Mass.: Harvard University Press.

Yablo, S. (2000) A paradox of existence. In A. Everett and T. Hofweber (eds.), *Empty Names, Fiction, and the Puzzles of Non-Existence*, 275–312. Stanford, Cal.: Center for Study of Language and Information.

Yablo, S. (2001) Go figure: a path through fictionalism. In P. A. French and H. K. Wettstein (eds.), *Figurative Language*, Midwest Studies in Philosophy 25, 72–102. Oxford: Blackwell.

Yablo, S. (2005) The myth of the seven. In M. E. Kalderon (ed.), *Fictionalism in Metaphysics*, 88–115. Oxford: Clarendon Press.

Yandell, D. (2003) Infinity and explanation, damnation and empiricism. In S. Wallace (ed.), *The Existence of God*, 131–40. Aldershot, Hampshire: Ashgate.

Zhang, J., Zhang, X., and Liu, H. (2007) Holographic dark energy in a cyclic universe. *European Physical Journal C* 52, 693–9, preprint: http://arxiv.org/PS_cache/arxiv/pdf/0708/0708.3121v2.pdf.

Zhang, X. (2007a) Can black holes be torn up by phantom in cyclic cosmology?, preprint: http://arxiv.org/PS_cache/arxiv/pdf/0708/0708.1408v1.pdf.

Zhang, X. (2007b) Comment on "Turnaround in Cyclic Cosmology" by Paul Frampton and Lauris Baum, preprint: http://arxiv.org/abs/0711.0667v1.