

1 Must the universe have a cause?

Nothing will come of nothing.

William Shakespeare, *King Lear*

THE MYSTERIES OF EXISTENCE

Why does the universe exist? Why do living things exist? Why do intelligent beings capable of suffering exist? These are among the most fundamental questions we can ask, and one of the most appealing reasons to believe in the existence of a benevolent creator is that it seems to answer them, whereas atheism seems unable to do so.

For the atheist, the universe is all that there is. There is nothing outside it. Consequently, there is nothing to point to as a cause of the universe's existence. So why there is something, rather than nothing, remains a mystery. Similarly, the evolution of life, from the atheist's perspective, serves no wider purpose. Life simply exists, it seems, for no other reason than its own perpetuation through reproduction. And the fact that the constitution of the universe should happen to have been such as to permit the evolution of life, and that exactly the right conditions for the evolution of life were realised, is similarly mysterious. Once we suppose there to be a creator, however, who has intentions and the limitless power to act on those intentions, these mysteries disappear.

Or do they? *Does* theism provide answers to the mysteries of existence? Is the atheist unable to produce rival, and equally satisfying, answers? Are the mysteries themselves genuine mysteries at all, or merely symptoms of fundamental intellectual confusion? Whether theism does indeed provide answers where atheism does not is the main theme of the first part of this book. We shall begin by looking at one influential and compelling argument for a creator of the universe which exploits our puzzlement over the existence and nature of that universe: the cosmological argument.

A FIRST CAUSE?

There are, in fact, a number of cosmological arguments. What they have in common is an observation about some very general feature of the universe, and the assertion that something must be the ultimate cause, or at least the ultimate explanation, of that feature. The arguments we shall examine conclude that the existence of the universe itself must have a cause. This cause cannot be part of the universe itself, for otherwise there would be something which caused itself to exist, and this, we intuitively think, is impossible. For example, suppose we believe, on the authority of a number of physicists, that the universe originated in the so-called ‘Big Bang’: an explosion from an almost infinitesimally small region of enormous density. We might say that everything that occurred *after* the Big Bang was caused by the Big Bang. But since the Big Bang is part of the universe’s history, we must include the Big Bang as part of what we are referring to by ‘the universe’. It would then be quite mistaken to say that ‘The Big Bang was the cause of the universe’, for this would mean ‘The Big Bang was the cause of the Big Bang and everything that came afterwards’. So, if the universe as a whole has a cause, this cause is not the Big Bang.

In this chapter we shall look at three versions of the cosmological argument. The first I shall call the *basic* cosmological argument, because the other two are modifications of it. It goes as follows:

The basic cosmological argument

- 1 Anything that exists has a cause of its existence.
- 2 Nothing can be the cause of its own existence.
- 3 The universe exists.

Therefore: The universe has a cause of its existence which lies outside the universe.

Although no-one has defended a cosmological argument of precisely this form, it provides a useful stepping-stone to the other, more sophisticated, versions. Before discussing it, we might note that the view that the cause of the universe’s existence should be an intelligent, benevolent creator who has an interest in his creation clearly requires more than this very brief argument. An argument for God, as he is conceived of by the theist, must surely involve a series of interconnected arguments, each contributing some further aspect to our understanding of God. Nevertheless, being persuaded by an argument for a cause of the universe is to take a large step towards theism.

Most proponents of cosmological arguments insist that the universe has not merely a cause but *a first* cause: something which is not caused by anything else. Now the first two premises of the basic argument,

- 1 Anything that exists has a cause of its existence.
- 2 Nothing can be the cause of its own existence.

are actually incompatible with the existence of a first cause. For if *everything* has a cause outside itself, then we are inevitably led to an infinite regression of causes: A was caused by B, which was caused by C, which was caused...etc. So, if we want to allow the possibility of a first cause, we must modify either (1) or (2). We could restrict either or both of them just to the parts of the universe, being careful, however, to include the universe itself as something which has a cause. Premise (1) could thus become:

Anything which exists *and is not outside the universe* has a cause of its existence.

We are, presumably, safe in assuming that the universe itself is not outside the universe. The problem with this amendment of the first premise, however, is that it seems rather arbitrary. We need to specify *what* it is about the universe which requires both it and anything within it to have a cause. This takes us to the two influential variants of the basic argument.

THE TEMPORAL AND MODAL COSMOLOGICAL ARGUMENTS

How else, then, may we amend the first premise, that everything has a cause? It is certainly true that everything that we can directly observe seems to have a cause of its existence. At least, this is true of clouds, houses, mountains, rivers, and so on. But what is also true is that these things all *began* to exist at a certain time, and the fact that they began to exist when they did, and not earlier or later, calls for causal explanation. Now, arguably, it is *only* those things which began to exist at a certain time whose existence calls for causal explanation. If something began to exist at some time, we can point to a time before it existed and say that that was when the cause of the thing's existence occurred. But if something has always existed, then we cannot point to a time before it existed. This suggests that things which have always existed have no cause. If this is so, then the proponent of the cosmological argument should offer a more restricted first premise:

- 1a Everything that begins to exist has a cause of its existence.

6 *The limits of theistic explanation*

But what of the universe? Did it begin to exist, or has it always existed? According to the Big Bang theory, the universe *did* have a beginning. If we are confident of this, then we can offer a more restricted form of the argument, which I shall call the *temporal* cosmological argument, as follows:

The temporal cosmological argument

- 1a Everything that begins to exist has a cause of its existence.
- 2 Nothing can be the cause of its own existence.
- 3a The universe began to exist.

Therefore: The universe has a cause of its existence which lies outside the universe.

So, to the question ‘What is it about the universe which requires it to have a cause?’, the proponent of the temporal argument can answer: the fact that the universe has a beginning. What is special about the first cause is that it has no beginning, and that is why it does not require a cause. So a regress of causes may be avoided.

Can we be confident about premise (3a), however? Suppose that the Big Bang theory is false—not an unreasonable supposition since, after all, cosmological theories are highly controversial, and even if there were universal agreement among physicists on this question—which is not the case—such agreement would not make the theory true. For all we know, the universe may not have had a beginning. This suggests two possibilities: (i) The universe extends infinitely far into the past; (ii) The universe is temporally closed: i.e., it is finite yet has neither a beginning nor an end. The first of these is perhaps easier to contemplate than the second, though both make considerable demands on our imagination. On the first view, we can represent the history of the universe as a series of events laid out along a line and which has no first member. Let us call this the ‘infinite past’ model.

$$\dots S_{t-4} \Rightarrow S_{t-3} \Rightarrow S_{t-2} \Rightarrow S_{t-1} \Rightarrow S_t$$

Figure 1.1 The infinite past model

‘ S_t ’ denotes all the events occurring in the universe at a particular time, t . ‘ S_{t-1} ’ denotes all the events occurring at an earlier time, $t-1$, etc. Some, perhaps all, of the events occurring at t will be caused by events occurring at $t-1$. In this sense, every member of the series has an antecedent cause. Since the series has no first member, no member is without a cause. On the second view, in contrast, we should represent the history of the universe as a series of events laid out around a circle. Let us call this the ‘closed time’ model.

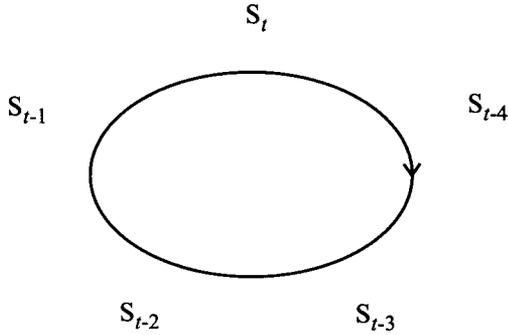


Figure 1.2 The closed time model

Here, again, there is no first member of the series: every event is preceded by some other event. But, unlike the infinite past model, the closed time model represents the history of the universe as only finitely extended: the past does not stretch indefinitely far back. However, although the past is only finite, it does not have a beginning, for all the events which occur before, e.g. S_b , also occur *after* S_t . A simple analogy for this view of the universe is provided by the surface of the earth: if you set out from some point on the equator, remain on the equator, and do not go through the same place more than once, your journey will only be finitely long; not because you will eventually reach a barrier through which you cannot pass, but because you will end up at your starting point.

It is tempting to be misled by this analogy with the earth's surface and suppose that the situation represented by Fig. 1.2 is that of history repeating itself. Just as we can go round and round the earth's surface, so we may imagine that, having come back to S_b , the universe will go round again and repeat the past sequence of events in the same order. But the situation represented by Fig. 1.2 is not that of history repeating itself. The events occur once and once only, but no event is the first. For example, take the event of my birth. The closed time model entails, not that I will be born again, but (and this will no doubt seem rather puzzling) that my birth is both in the relatively recent past *and* in the future—though the very distant future, if the circle is a large one.

Both on the infinite past model and on the closed time model, the universe does not have a beginning. The temporal cosmological argument does not therefore apply in these cases, because premise (3a), that the universe began to exist, would be false. Precisely because the temporal version of the argument seems to give hostages to empirical

fortune in this way, some defenders of the cosmological argument might prefer not to restrict the first premise just to things which have a beginning. And perhaps they would be right not to do so, for, if *only* things which have a beginning have a cause for their existence, then the discovery of conclusive evidence that the universe did not have a beginning would be a serious threat to belief in a creator.

To recapitulate the discussion so far, the cosmological argument concludes that there is a cause of the universe (or some feature of the universe), namely God. The more general version of the argument starts from the premise that everything that exists has a cause. It was then suggested that only things that *begin* to exist need to be explained in causal terms, and thus that the universe only has a cause if it has a beginning. Can this assumption be questioned? Is it possible that, even if the universe is as portrayed in Fig. 1.1 or Fig. 1.2, it may still have a cause? Here is a reason for thinking so. Although it is true, both on the infinite past model and on the closed time model, that each event in the history of the universe has a cause, we do not thereby have a causal explanation of the existence of *the universe as a whole*. We can answer the question, ‘Why did this or that particular event occur when it did?’ But we cannot answer the question, ‘Why does the universe exist at all?’ This question remains, whichever view of the universe we adopt. So we should leave the door open for a causal explanation of both an infinite past world and a closed time world.

But, then, what exactly is it, if not the fact that it has a beginning, that makes the existence of the universe mysterious, and that motivates us to look for a cause? One answer is that the existence of a universe is a purely *contingent* matter. That is, although there is in fact a universe, things might have been otherwise: there might have been no universe at all. It is not impossible for there to have been absolutely nothing. And this is a feature of things which have causes, that their existence is a purely contingent matter. This reflection suggests another way of restricting the first premise of the basic argument, providing us with a third version, which I shall call the *modal* cosmological argument. (In this context, the word ‘modal’ refers to matters of necessity and possibility, ideas which we will look at more closely in the next chapter.) It goes as follows:

The modal cosmological argument

- 1b Everything whose existence is contingent has a cause of its existence.
- 2 Nothing can be the cause of its own existence.
- 3b The existence of the universe is contingent.

Therefore: The universe has a cause of its existence which lies outside the universe.

This, or something like it, is sometimes called ‘the argument from contingency’. Like the temporal argument, the modal argument allows for the existence of a first cause. In this case, however, the first cause would have to be something whose existence was not contingent, but necessary. That is, it would have been impossible for it not to exist. Only so could it lack a cause.

The plausibility of (3b), unlike that of (3a), does not depend in any way upon the outcome of scientific investigation. Because of this, the modal cosmological argument may seem more defensible than the temporal version. However, as we shall see in Chapter 3, there are problems with the notion of a necessary being as a cause of the universe.

It seems, then, as if there are ways to avoid a regress of causes. Let us now look at the first premises of the temporal and modal cosmological arguments.

PROBLEMS WITH THE FIRST PREMISE

‘Everything that begins to exist has a cause of its existence.’ How secure is this premise? A toadstool appears overnight in my garden. Seeing it the next morning, I am led to wonder both why it appeared at all and why it appeared last night and not sooner. With my elementary grasp of biology, I reason that there must have been spores in the soil. I reason further that conditions favoured the appearance of toadstools last night: there was sufficient moisture, there had been no hard frost, and any number of other important factors were present. Were I to study the spores in detail I should no doubt discover some internal physiological mechanism which, in conjunction with external conditions, was responsible for the appearance of the toadstool at just that time. To generalise: things come into existence because of the conditions that obtained just prior to their appearance.

Some will object that this generalisation is simply unjustified, for modern physics has discovered both that there are some phenomena at the sub-atomic level which occur quite randomly, and that, at or near the time of the Big Bang, the laws of physics break down, and so what emerges from the Big Bang is unpredictable. Therefore, the suggestion goes, there are things which begin to exist, and whose existence is purely contingent, which yet are uncaused. I think we should be very cautious about these grounds for rejecting the first premise. Physics itself is in a state of rapid development, and whatever theories are on offer at a particular time are not only the subject of controversies among physicists

but are also liable to be replaced at some later date. We should, in any case, be wary of the move from 'unpredictable' to 'uncaused'. We may simply be unable to discern, for reasons to do with the laws themselves, what laws are operating both at the sub-atomic level and at the Big Bang. So, rather than trying to attack directly the premise that everything that begins to exist has a cause, let us instead ask what authority it has. Is it simply a deeply held conviction? A guess? Or something more than that?

Let us look at three suggestions, each of them attempting to explain how the premise could count as something we *know* to be true. The first suggestion is that we know it to be true *a priori*, and this is because it is *analytically* true. The second suggestion is that we know it to be true *a priori*, but it is not analytically true. The third suggestion is that we do not know it to be true *a priori*, but rather we infer it inductively from our observations. I shall assume that these three answers exhaust the possible explanations of how the premise could count as a piece of knowledge. What do they mean?

Let us begin with the first suggestion. We know something to be true *a priori* if we can verify it without having to rely directly on observation or experience. For example, we know that twelve plus six equals eighteen without having to observe a group of twelve objects being added to a group of six objects and then counting the resulting group. Provided that we understand the number system, we can work out such a simple sum in our heads. Of course, in order to gain an understanding of the number system, we needed to have the requisite experiences, perhaps by manipulating counters, but once having acquired this understanding, we no longer need to appeal to experience in order to perform mathematical calculations. Now we know some propositions to be true *a priori* because they are also *analytically* true. So what is it for something to be analytically true? There is some disagreement amongst philosophers on this. On one account, analytic truths are those which are true by virtue of the meanings of the words. On another, analytic truths are those whose negations are self-contradictory. (The negation of a sentence is simply the result of putting 'It is not the case that' before that sentence.) An example of a sentence which both accounts would judge to be analytic is 'Anaesthetics reduce sensitivity to pain.' Suppose I sincerely asserted that I had just invented an anaesthetic that heightened people's sensitivity to pain. I could surely not have grasped the meaning of the term 'anaesthetic'. You would know that my assertion was false because 'anaesthetic' *means* something which reduces one's sensitivity to pain. You do not need to step into my laboratory to see

whether my assertion is true or not. The statement that 'I have invented an anaesthetic which heightens people's sensitivity to pain' implies that there is an anaesthetic which does *not* reduce sensitivity to pain, and this is self-contradictory.

For a large number of examples, the two accounts of analyticity agree on whether a proposition should be classified as analytic or not. Either will do for our purposes, but for simplicity I shall use 'analytic' to mean 'has a self-contradictory negation'.

So, to return to our premise, can it plausibly be regarded as analytically true? No. Someone who sincerely asserted that there were, or might be, some things which began to exist and yet were not caused would not obviously be contradicting themselves. It is true that we might be highly puzzled by the thought of something's coming into existence without a cause, since there would apparently be no explanation of why it came into existence when it did, nor indeed of why it came into existence at all. But such puzzlement is not the same as discovering a contradiction in the idea, and may arise simply because a thing without a cause is contrary to our experience.

As we said above, if something is analytically true, then we can know it to be true *a priori*. However, there may be some things that we know *a priori* but which cannot be captured by either of our definitions of 'analytic'. Kant thought that geometrical propositions were of this kind (though his characterisation of the analytic does not coincide precisely with either of our two definitions), and he labelled them 'synthetic *a priori*' truths. 'Synthetic' here simply means 'non-analytic'. One possible example of a synthetic *a priori* truth is 'Nothing is both red all over and green all over'. We do not need to verify this by appeal to experience, so we know it to be true *a priori*. But it is far from clear that its truth is guaranteed simply by the meaning of the words, or that 'Something is both red all over and green all over' is self-contradictory. In other words, it appears to be synthetic. Now if there are such things as synthetic *a priori* truths, then the possibility remains that the first premise is one such truth. This takes us to our second suggestion. Here it might be objected that we can at least conceive of the idea of something's not having a cause, whereas we cannot conceive of the falsity of an *a priori* truth. (Try conceiving of something's being red all over and green all over at the same time.) This will not convince defenders of the cosmological argument, however. We may conceive of an event without conceiving of its cause, they will say, but this is not to conceive of an event which has no cause. There is, however, another objection to this suggestion, which I shall present later.

What, finally, of the third suggestion? This was that we know that things that begin to exist have causes because we *inductively infer* it from observation. Here is an example of a—not very safe—inductive inference. I observe that the 12.20 train from Oxenholme to Windermere has been late four days running, and infer from this that this service is always late. Another example of such an inference is the inference from the fact that the British Conservative Party has won the last four general elections that it will win the next one. Yet another is the inference from the observation that a crow is black to the conclusion that all crows are black. Clearly, some inductive inferences are safer than others, but they all have a common form, which we can characterise as follows: an inductive inference is one which moves from a premise about some members of a certain class to a conclusion either about some of the other members of that class or about all the members of that class. The third suggestion, then, is that we infer from our observation of things and their causes, that everything that begins to exist has a cause.

Now, if the first premise is to support the conclusion of the cosmological argument, ‘everything’ must include the universe itself. So, if the third suggestion is correct, our experience justifies us in positing a cause for the universe. But the causes which we have experience of take place in time and space, and this is not an accidental connection. We suppose things to have causes because we want to explain why those things came into existence *at the times and places they did*. We therefore look for the causes of those things in the conditions which obtained just before, and in the vicinity of, the thing in question. Conditions which obtained elsewhere or at other times cannot provide the relevant explanation. Causation, then, is a *temporal* concept. (It is perhaps also a spatial concept, but I do not want to insist on that here.) It is this aspect of causation which threatens the inference from what we experience to a conclusion about everything which begins to exist.

Suppose the universe has a beginning in time, as the temporal cosmological argument requires. Three possibilities present themselves. The first is that time itself has a beginning, one which coincides with the beginning of the universe. The second is that there is a finite period of time before the beginning of the universe. The third is that there is an infinite period of time before the beginning of the universe. If the first of these possibilities obtains, then the universe cannot have a cause, at least not in the ordinary sense, for in the ordinary sense the cause of the existence of a thing is something which occurs just before the thing begins to exist. But, if the beginning of time coincides with

the beginning of the universe, then nothing could have occurred before the universe started to exist. If the second of the possibilities obtains, if it is true that everything that begins to exist has a cause, then the universe has a cause. But, since time itself has a beginning, it too must have a cause. But, by definition, nothing can occur before time itself. Time cannot have a cause for its existence, and so it provides a counterexample to the premise that everything that begins to exist has a cause. If the third possibility obtains, then, again, the universe can have a cause, but it would simply be the last member of an infinite chain of causes:

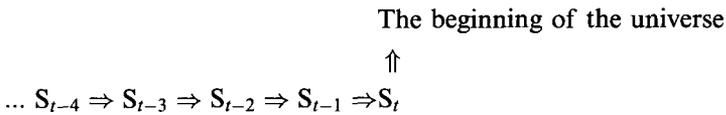


Figure 1.3 The beginning of the universe on the infinite past model

Why is this? Why could the cause of the universe not be something like an eternal, immutable God, who needs no cause for his existence? Well, the mere existence of God, or of any other object, could not causally explain why the universe came into existence. It must be something *about* God which does the explaining, such as his willing the universe to exist. But has he, for all time, willed the universe to exist? Why, then, did it not come into existence sooner? If there is a cause of the universe's coming into existence at precisely the moment it did, then it is something which obtained just before that event. We are then led to ask why that cause obtained when it did, and so, by similar reasoning, we are led back to the regress of causes which the temporal argument was supposed to avoid. So either the first premise is false, because there *is* something which begins to exist yet has no cause, and so cannot be either a piece of *a priori* knowledge or the result of a sound inductive inference, or there is no *first* cause.

So far in this section, we have concentrated entirely on the first premise of the temporal cosmological argument. What of the first premise of the modal argument, that everything whose existence is contingent has a cause of its existence? This is not obviously something we know *a priori*, nor is it obviously something we infer from experience, but whatever the supposed basis of its authority, it faces the objection we have just been discussing: causation is essentially a temporal concept. So, if the universe is supposed to have a beginning, the problems we

encountered above will still occur. The difference between the modal and temporal arguments is that the modal argument allows for the possibility of the universe's not having a beginning. But if it does not have a beginning, then it cannot have a cause in the ordinary sense, for nothing could then have occurred *before* the universe existed.

There is a further problem for the modal argument. According to the first premise, everything whose existence is contingent, i.e. everything which might not have existed, has a cause. But, arguably, time itself might not have existed: it too exists only contingently. So the first premise of the modal argument directs us to the conclusion that time itself has a cause. But since, as we noted above, nothing can occur before time itself, time cannot be said to have a cause.

We can conclude that, if the idea that the universe has a cause of its existence is to be defended, it must be on a very different understanding of 'cause' than the one with which we ordinarily operate. But, to justify the name, the 'cause' of the universe must at least play something like the role which ordinary causes play in our view of things. It must at least provide an explanation of why the universe exists. Whether such an explanation is possible is the subject of Chapter 3.

In this chapter, we have focused on the notion of causation. But the modal cosmological argument also introduces another important concept, that of necessity. It is now time to look at this notion.

SUMMARY

An ancient and influential argument for the existence of a creator is the cosmological argument. We examined three versions, all of which exploit the notion of causality. The first, the basic argument, begins with the premise that everything that exists has a cause. The problem with this argument is that it implies an infinite regress of causes, whereas God is supposed to be a *first* cause: something not caused by anything else. The difficulty can be overcome, however, if we restrict the first premise in some way. This led us to two other versions of the argument. The second version, the temporal argument, begins with the premise that everything which *begins* to exist has a cause. This argument only establishes that the universe has a cause if it can be established that the universe has a beginning, and there is some doubt as to whether this could be established. This difficulty is avoided by the third version, the modal cosmological argument, which begins with the premise that everything whose existence is merely contingent, i.e. which might not have existed, has a cause.

The problem with both the temporal and the modal arguments is that they necessarily represent the first cause as being something utterly unlike ordinary causes. Our ordinary notion of causation is bound up intimately with the notion of time. Causes take place at particular moments of time, and before their effects. A first cause, however, would have a completely different relationship to time. So different, in fact, that we have to admit that the universe cannot be said to have a cause in the ordinary sense of the word.

FURTHER READING

For an introductory discussion of cosmological arguments, see Chapter 5 of Brian Davies, *An Introduction to the Philosophy of Religion*, 2nd Edition, Oxford: Oxford University Press, 1993.

The classic statement of a number of cosmological arguments is the ‘five ways’ of St Thomas Aquinas. See his *Summa Theologiae*, Part I, Question 2, Article 3. There are a number of editions of Aquinas’s works: for example, Anton C. Pegis (ed.), *The Basic Writings of Saint Thomas Aquinas*, New York: Random House, 1945.

The temporal argument, under a different name, is discussed and defended at length in William Lane Craig’s *The Kalam Cosmological Argument*, London: Macmillan, 1979. ‘Kalam’ is an Arabic word, and Craig traces the development of the argument in the writings of Arabic philosophers, to whom we owe both versions of the cosmological argument. Craig spends a large part of the book defending the second premise, that the universe had a beginning in time, which he believes can be established on *a priori* grounds. Parts of the book are reprinted in William Lane Craig and Quentin Smith, *Theism, Atheism and Big Bang Cosmology*, Oxford: Clarendon Press, 1993. Craig argues that Big Bang cosmology supports theism, whereas Smith argues that it supports atheism in so far as it implies that the universe could not have had a cause. Smith also criticises Craig’s *a priori* arguments for the beginning of the universe.

See also Craig’s *The Cosmological Argument from Plato to Leibniz*, London: Macmillan, 1980, from which one can get a clear idea of the enormous variety of cosmological arguments.

For a discussion of the relation between the cosmological argument and the topic of the next chapter, the notion of a necessary being, see Chapter 4 of William L. Rowe’s *The Cosmological Argument*, Princeton: Princeton University Press, 1975.

The implications of the closed time model are explored in W.H. Newton-Smith’s *The Structure of Time*, Routledge & Kegan Paul, 1980. For what now strikes me, for reasons given above, as a misguided attempt

to reconcile the idea of a first cause with the closed time model, see Robin Le Poidevin, 'Creation in a Closed Universe or, Have Physicists Disproved the Existence of God?', *Religious Studies*, vol. 27(1991), pp. 39–48.

A famous statement of the modal argument was given by Leibniz in his essay 'On the Ultimate Origination of Things', in *G.W. Leibniz, Philosophical Writings*, ed. G.H.R. Parkinson, London: J.M. Dent, 1973, pp. 136–44. It is given a sympathetic treatment in Richard Swinburne's *The Existence of God*, Oxford: Clarendon Press, 1979, Chapter 7.