1. **Introduction**

The so-called “cosmological argument” (as it was first labeled by Immanuel Kant) has played a central role within Aristotelian philosophy from the very beginning. In fact, we find first-cause arguments before Aristotle: most importantly, in Plato’s *The Laws*, Book X. The inference to the existence of a god or gods is central to Aristotle’s two most important works in metaphysics, *Physics* and *Metaphysics*. The Neo-Platonists (including Plotinus and Proclus), who saw themselves as building on Aristotle’s foundations, made this argument fundamental to their understanding of reality.

Many of the ancient critics of Aristotle, such as John Philoponus, the Kalām tradition (al-Kindi, al-Ghazzālī), and Bonaventure, drew heavily from Aristotle in their philosophizing, and here too the cosmological argument took central stage. Finally, this argument continued to influence modern philosophy, in both its rationalist (Descartes, Spinoza, Leibniz, the pre-critical Kant) and empiricist (Locke, Samuel Clarke) varieties. This continuity demonstrates that modern philosophy retained more Aristotelian elements than is often recognized.

No revival of the Aristotelian tradition in metaphysics can afford to ignore the question of the viability of first-cause arguments. In fact, there has been a flourishing of first-cause argumentation within analytic metaphysics in the last fifteen years, a development with parallels in other branches of the field.

In the section 2 below, I deal with some preliminary issues about the form of the argument. I take on the most important question, that of justifying some form of a
causal principle or ‘principle of sufficient reason’, in section 3. Next, I consider (in section 4) the dual problems of avoiding an infinite regress and providing a principled ground for stopping that regress with God and not before. This section comprises four different approaches: Aquinas’s appeal to per se causation, al-Fārābī’s aggregation argument, a version formulated in terms of George Boolos’s plural logic (Boolos 1984), and Leibniz’s use of infinitary conjunction. I turn in section 5 to two arguments for a supernatural cause of the universe that do not rely on a global causal principle to reach their conclusion but instead appeal to inference to the best explanation. These arguments include an interpretation of Aquinas’s First Way (seeking an explanation of the perpetuation of time), and a Neo-Platonic argument from the spatial or spatiotemporal unity of the world. Finally, section 6 addresses the problem of moving from a first cause to God (as classically conceived).

2. PRELIMINARY ISSUES

Here is the general form of the cosmological or first-cause argument:

1. There are things of a certain kind F such that everything of that kind has a cause or causal explanation.
2. There must be an ultimate cause or explanation of each of these things.
3. Such an ultimate cause must be God.
4. Therefore, God exists.

The argument has just three premises. A defense of the argument must therefore tackle three problems:
1. **The Justification Problem.**
   What justifies the causal or explanatory principle appealed to in premise 1?

2. **The Regress Problem.**
   Why must the chain of explanations terminate?

3. **The Gap Problem.**
   What justifies identifying a first cause with God?

I will take up the Justification Problem in the next section (12), the Regress Problem in section 3, and the Gap Problem in section 5, after considering some variations that do not fit the general template in section 4.

Before proceeding with this three-part defense, there are a number of preliminary questions to ask about this argument. First, we can ask what sort of causal or explanatory relation is involved. This question is inseparable from the question of what kinds of ‘things’ are in question here: what are the relata of the causal or explanatory relation?

Second, we can ask about the relation between causation and time: are the causes and explanations supposed to precede their effects in time, or are we looking for some kind of synchronic or timeless explanation?

Finally, what sort of kind is F? In particular, should we identify F with the property of existing contingently? Can necessary beings be caused or explained?

### 1.1 The Explanatory Relation and its Relata

Throughout the latter part of the twentieth century, philosophers have been arguing about whether the fundamental causal relation is between things (events, conditions, existing entities) or truths (propositions, facts). In the 1960’s and 70’s, the consensus favored the primacy of truths. Causal explanation was supposed to be some kind of relation between true propositions, analogous to logical connectives like ‘or’ or ‘not’. The fundamental causal facts were supposed to take the form: \( q, \)
because \( p \). This propositional-connective model fits best with cosmological arguments that appeal to Leibniz’s “Principle of Sufficient Reason”: the principle that every contingent truth has some adequate explanation in terms of other, more fundamental truths, bottoming out in a foundation of necessary truths.

However, the latter part of the twentieth century has seen a resurgence of the competing model: causation as fundamentally a connection or tie between things (like events, states, or other spatiotemporally located entities). This model corresponds to cosmological arguments that appeal to a causal principle of the form: contingent (or wholly contingent) things must be caused by other, separate things, with some necessarily existing things providing the ultimate causes.

The two models are not completely unrelated. If we start with the real-connection model, we can say what it is to explain certain kinds of contingent truth. If \( C \) was the cause of \( E \), then the truth that \( E \) existed or occurred is explained by the truth that \( C \) existed or occurred. Alternatively, if we start with the propositional-connective model, then we can use that model to build an account of causal connections, so long was we suppose that each contingent truth is ‘made true’ by some concrete entity or arrangement of entities. If it is the case that \( q \) because \( p \), and if \( e(p) \) and \( e(q) \) are the ‘truthmakers’ of \( p \) and \( q \), respectively, then it should follow that \( e(p) \) is a cause of \( e(q) \). In the case in which \( p \) and \( q \) are each of the form: ‘\( x \) exists’, for some \( x \), this truthmaker account is quite plausible. A truth of the form ‘\( x \) exists’ is always made true by \( x \) itself.

The principal difference between the two models will emerge when we consider how to block the regress. For those adopting the propositional-connective model and Leibniz’s Principle of Sufficient Reason, the most attractive approach will be that followed by Richard Gale and Alexander Pruss (Gale and Pruss 1999): using the logical operation of conjunction to form the One Big Conjunctive Truth, and then apply the PSR to that, resulting in a necessary truth as its explanation. Those adopting the real-
connection model and a global Causal Principle, have a number of alternative ways of achieving the same result, including aggregation, plural quantification, and more direct arguments against the possibility of infinite regresses.

The Aristotelian tradition contributes several crucial insights to this problem, without finally deciding the issue. First, we can say that only the actual can be a cause or ground a causal explanation. The merely possible or potential (as such) causes nothing. Second, the primary or focal sense of causation applies to causes and effects that are positive: to the existence of things, or to their being a certain way (their “accidents”). If privations or negations have causes at all, it is in only a derivative sense. Third, a cause must take the form of a powerful particular: a particular, concrete substance with an appropriate active power. The effect is constituted by the exercise of this active power. Whether an active power exists and can be exercised on a particular occasion depends upon the intrinsic nature of the agent and on its external circumstances. In the case of rational agents, the potential exercise of an active power depends on the agent’s having some reason for doing so. If there are adequate reasons for contrary actions on a given occasion, rational agency can introduce an element of contingency.

1.2. Diachronic vs. Synchronic Causation

Another crucial issue concerns the relation between causation and time. In some cases, it seems that the cause must precede its effect in time. This would hold in every case of continuous causation. Where a substance or group of substances is undergoing some continuous process or activity (such as locomotion, e.g., the revolution of the earth around the sun), the earlier phases of the activity are in some sense causes of the later phases.

In other cases, causation would seem to be simultaneous. When one substance acts upon another, the cause and effect are simultaneous, since the effect is identical
with the exercise of the agent’s active power, and a power can be exercised only while it exists. The two modes of causation are intertwined: it is continuous causation that explains why instances of discrete causation occur when they do. Consider, for example, the continuous motion of two billiard balls toward one another, culminating in the discrete interaction that occurs when they collide. The collision, in turn, gives rise to new processes of locomotion. A diachronic chain of such alternating continuous and discrete causation could regress arbitrarily far into the past.

This twofold account of causation gives rise to three different kinds of cosmological argument: (1) those that argue that such a diachronic regress cannot extend infinitely far into the past, (2) those that argue that no act of discrete causation can be fully explained by continuous action in the past, but requires some further synchronic cause in the moment of action, and (3) those that seek to disentangle causation from time altogether, moving ultimately to a timeless cause of the whole cosmos of finite things, regardless of whether its past is finite or infinite in extent.

1.3. Necessary vs. Uncaused

What sorts of things call for or require a cause? One traditional answer is: contingent things, things that could have failed to exist or occur. Successful cosmological arguments of this kind reach one or more necessary being as the First Cause.

However, one might argue (as does Thomas Aquinas in the Third Way1) that such an argument is not sufficiently ambitious. Arguably, proponents of such argument overlook the possibility that some necessary beings might also require a cause. Some necessary beings might be necessary in a dependent way: through or by the necessity of something else. For example, suppose that God gives rise to universal
ideas through a process of thinking. If both God and God’s thinking were necessary beings, then the ideas would also be necessary and yet causally dependent.

It won’t do to adopt as our principle the rule that all dependent things have causes, if ‘dependent’ is understood as meaning simply ‘caused’, since such a principle would be a mere tautology. More promising would be this: to propose that everything that could possibly have a cause actually do so. Only a thing necessary in itself would be absolutely uncausable.

3. The Justification Problem

The first problem to address is that of justifying a global causal or explanatory principle. Why assume that anything has a cause or causal explanation? The simplest move is to appeal to self-evidence. It is simply constitutive of perfect rationality that one look for and expect to find a cause whenever possible. As I mentioned above, it is hard to deny the reasonableness of a defeasible principle of causation, at the very least. However, there are several strategies available for defending the truth of exceptionless causal principles.

3.1 Empirical Inference to the Best Explanation

Alexander Pruss has argued that a global causal principle or Principle of Sufficient Reason is the best explanation for the fact that we do not see bricks and proton clouds appearing ex nihilo. A global causal principle provides the simplest explanation for this fact.

3.2 Epistemological Arguments

The second such strategy appeals to epistemological considerations.

If uncaused facts are possible, we cannot assign any objective probability to their obtaining or not obtaining. It is only when something is caused that we can appeal to the dispositions and propensities of its potential causes to ground a fact about objective probabilities. If uncaused events were possible, their occurrence would
have no well-defined probability. Thus, we could never have good grounds for supposing that any event was likely to have been caused at all, much less caused in any specific way.

This generates a defeater of all empirical claims, since all empirical claims to knowledge depend on the supposition that our experiences and memories have causes of the appropriate kind (veridical and non-deviant). In a world in which uncaused states are possible, a kind of counter-causal Cartesian demon would lurk behind every event, and we would never be in a position to judge with any reliability that such a demon were unlikely in any given case.

Could the laws of nature (contingently) rule out the possibility of uncaused and spatiotemporally bounded events in our world? No, since the laws of nature only specify what can happen in the absence of external influences. If uncaused events and entities were metaphysically possible, the laws of nature could not rule them out in our world either, since they could not, by themselves, rule out the spontaneous generation of new physical objects with new forces. The application of the laws of nature to exclude certain kinds of events always presupposes a global principle of causality.

What sort of causal principle is justified by these considerations? The principle must be strong enough to encompass any possible piece of empirical evidence, without requiring any empirical knowledge to justify its application (to avoid a vicious circularity). In other words, the principle must take the form: any F has a cause, where the description ‘F’ is broad enough to include any possible empirical data, and such that the application of ‘F’ to its instances is always certain on purely a priori grounds. The requirement of a priori applicability means that ‘F’ must take the form of something like ‘appears to be G’ or ‘can coherently be thought to be G’. The requirement of scope means that G should include everything that is causable.
Putting these facts together, we reach the conclusion that there must be an *a priori* knowable principle of the form: whatever is conceivably causable has a cause.

The fact that such a principle excludes those things that are inconceivably causable poses no threat of incipient skepticism, since if one is justified in believing that something exists which is inconceivably causable, that justification cannot depend on the supposition that the thing was in fact caused in some way.

### 3.3 A Modal Argument for the Causal Principle

Building on Sullivan,⁴ Pruss has also argued for a global causal principle from certain principles of modality.⁵ The causal principle that Pruss argues for takes the following form: if E is an actual state of affairs and E could possibly have a cause, then E does have a cause.

It seems plausible to suppose that an effect would not have occurred without the cause. Many, following Hume and David Lewis, have argued that this fact could be the basis for an analysis of causation. However, even if we suppose such counterfactual analyses of causation to be unsuccessful, it still seems that such counterfactual dependency is a necessary condition for causation.

Pruss argues that we can assume something even weaker: if C caused E, then it should be the case that if nothing had caused E, then E would not have occurred. This weaker assumption works even in cases in which E is over-determined to happen.

\[ \diamond (C \text{ causes } E) \land (\neg \square (D \text{ causes } E) \land E \text{ did not occur}) \]

The so-called Brouwer axiom of modal logic states that if some proposition p is actually true, then it is necessarily possible. This corresponds to a symmetry condition on the accessibility relations between worlds: if any world w is possible relative to our actual world, then the actual world must also be possible relative to w. The Brouwer axiom is entailed by but is weaker than the standard modal logic for metaphysical necessity, S₅.
If we apply the principle behind the Brouwer axiom to counterfactual conditionals, we should get the following rule: if \( p \) is actually the case, then if \( p \) were not the case, then some world \( w \) would have been actual, such that, in \( w \): if \( p \) were the case, the actual world might have been actual. This rule could be captured by the following axiom:

\[
(2) \quad (q \land p \land \neg p) \rightarrow (\neg p \rightarrow (p \rightarrow q))
\]

As it turns out, Pruss’s axiom (2) is not valid in the standard semantics for counterfactual conditionals as developed by David Lewis and Robert Stalnaker, but Pruss argues persuasively that this points to a flaw in that semantics, not to any grounds for doubting (2). In addition, Pruss’s proof requires two more obvious axioms involving conditionals (axioms that are valid in all standard theories):

\[
(3) \quad (p \rightarrow q) \rightarrow (p \rightarrow q)
\]

\[
(4) \quad ((p \rightarrow q) \land (p \rightarrow \neg q)) \rightarrow \neg p
\]

In these axioms, ‘\( \rightarrow \)’ represents logical entailment.

Suppose that \( q \) is the true proposition that some state of affairs \( E \) occurs, and suppose that \( E \) can have a cause but does not in fact have one. Let \( p \) be the proposition that nothing causes \( E \). Thus, both \( p \) and \( q \) are true in the actual world.

Since \( E \) can have a cause, we have \( \neg p \).

By (2), we know that if \( \neg p \) were the case, then if \( p \) had been the case, then \( q \) might have been the case. Let \( w \) be a world in which \( \neg p \) is true, in which \( E \) has a cause. Let’s call this possible cause \( C \). Applying (1) to world \( w \), we get the result that it is true at \( w \) that had \( E \) had no causes, i.e., had \( p \) been true, \( E \) would not have occurred. But this means that if \( \neg p \) were true, it would have been the case both that \( (p \rightarrow \neg q) \) and \( (p \rightarrow q) \), which is impossible. So, \( \neg p \) is impossible, contrary to our hypothesis.
Consequently, any state of affairs that can possibly have a cause must have a cause in fact.

Pruss goes on to argue that all wholly contingent, positive states of affairs can have a cause. If a state of affairs $S$ is wholly contingent, then we can conceive of a world in which something exists which has the power to bring that state of affairs into existence, and which exercises that power in that world, resulting in the state of affairs $S$. Thus, any wholly contingent, positive state of affairs would seem to be causable. There is only one reason for doubting this conclusion: $S$ might include some entity $E$ for which essentiality of origins holds. That is, $E$ might be the kind of thing such that if any duplicate of $E$ were to come into existence in any other way, it would have to be numerically distinct from $E$. Thus, if $S$ has no cause in the actual world, $E$ could not exist in a world in which its existence (and thus $S$ itself) had any cause at all. Consequently, Pruss proposes a weaker principle, (5), which explicitly excludes this case:

(5) Every wholly contingent, positive state of affairs that does not de re involve contingent entities for which essentiality of origins holds can have a cause.

Principle (5) encompasses all states of affairs that are characterized in purely qualitative terms. It seems plausible that if a state of affairs $S'$ is the purely qualitative counterpart of $S$, and $S'$ has a cause, so must $S$. Consequently, we can endorse (6):

(6) If all wholly contingent, positive states of affairs that do not de re involve entities for which essentiality of origins holds have causes, then all wholly contingent, positive states of affairs have causes.

Pruss argues that (6) will be true in any ‘nice’ world: one lacking indiscernible but distinct entities. Our world appears to be ‘nice’. Putting (5) and (6) together with
Pruss’s global causal principle, we reach the result that every wholly contingent, positive state of affairs has a cause.

3.4. The Problem of Grounding Modality

As Alexander Pruss has argued,7 narrowly logical, Lewisian, Platonic (Robert M. Adams, Alvin Plantinga), and non-causal Aristotelian essentialist accounts are all problematic. The best alternative is Aristotelian: possibilities and necessities are all grounded (made true) by facts about the powers of actual things. Immanuel Kant developed a similar argument in his pre-critical phase.8

Suppose, for reductio, that there were a fact p that obtains without causal explanation. Let q be the fact that p obtains without explanation. Consider a world w in which p is false. Given the Brower axiom B, if p were false, q would still be possible. However, nothing in world w could ground the possibility of q. Hence, the grounding of modality requires a global principle of causation.

Pruss also provides an argument that doesn’t depend on axiom B:

(7) If the causal principle is true in all possible worlds with the possible exception of the actual world, then it is true in all possible worlds.

If (7) is true, then either the causal principle is true in the actual world or there is some non-actual world w in which the causal principle is false. We can generate a contradiction from the second disjunct. Let p be some fact that obtains in w without explanation. Let r be some contingent fact that obtains in w but not in the actual world (if p does not obtain in the actual world, just let \( r = p \)). Then \( (r \& p) \) is a contingent fact that obtains in w without explanation and does not obtain in the actual world. But this means (on the causal-Aristotelian account) that there is some entity with some causal power in the actual world capable of bringing it about that \( (r \& p) \) be true without causal explanation, an obvious absurdity.9
3.5. Relying on a Weak Principle of Sufficient Reason

Richard Gale and Alexander Pruss showed\textsuperscript{10} that a valid cosmological argument could be build by means of a weak principle of sufficient reason:

\textit{Weak Principle of Sufficient Reason}:

It is possible that every contingent state of affairs have a sufficient explanation.

From the Weak PSR we can get the result that it is possible that a powerful necessary being exists. We can then use S5 (including the principle that whatever is possibly necessary is necessary simpliciter) to get the result that such a being actually exists.

There are two general worries with the Gale-Pruss move:

1. Do we have any grounds for believing the weak versions of the CP that aren’t equally good grounds for believing the strong versions?
2. Is S5 too strong? The argument won’t work with S4, or anything weaker.

In a series of recent articles, Joshua Rasmussen has created several variations on the Gale-Pruss theme. In his 2010 paper “From States of Affairs to a Necessary Being”, he assumes the following:\textsuperscript{11}

(8) If $x$ is an intrinsic type of concrete particular, then it is possible that the fact that there is at least one member of $x$ has a causal explanation.

Consider the type of being a contingently existing concrete particular. By (8), it must be possible for something to cause this type to have at least one member. Such a cause must be a necessarily existing concrete particular. By S5, we can infer that this necessarily existing being exists also in the actual world.
In his 2011 paper “A New Argument for a Necessary Being”, Rasmussen assumes instead:

(9) Any intrinsic property $\pi$ that (i) possibly begins to be instantiated, and (ii) is possibly instantiated by something that has a cause is such that it is possible that there is a cause of $\pi$’s beginning to be instantiated.

It is plausible to suppose that the property of being a contingent concrete particular (call it property $c$) is an intrinsic property that possibly begins to be instantiated and is possibly instantiated by something that has a cause. Therefore, it is possible that this property $c$ is caused to begin to be instantiated. This cause could not itself instantiate $c$, on pain of a vicious circularity: something that instantiates a property $\pi$ cannot be the cause of $\pi$’s beginning to be instantiated. So, it is possible that there exist some powerful necessary being (and if powerful, then concrete).

By S5, this necessary being exists in the actual world.

John Turri raised the following worry about Rasmussen’s arguments: is the necessary being powerful in every possible world, including the actual one? In response to this worry, we consider the property of being either a contingent concrete particular or a necessary being with some contingent power. This seems to be an intrinsic property that could begin to be instantiated. Consequently, it must be possible for this property to be caused to begin to be instantiated. Thus, there must be some necessary being with some essential power in some possible world that causes this property to begin to be instantiated in that world. This enables us to infer (in S5) the actual existence of a necessary being with essential powers.

Another worry about Rasmussen's argument concerns his assumption that the property of being a contingent particular a property that is only contingently instantiated, or one that could begin to be instantiated. Suppose that the past is infinitely long, and that there always have existed contingently existing particulars.
We might argue that, under those assumptions, it is metaphysically necessary that the property be instantiated, and metaphysically impossible that it should begin to be instantiated. This would follow if we accepted something like Sydney Shoemaker’s Branch Principle:\(^1\)

\textit{Branch Principle:}

\begin{quote}
For every possible world w, there is a time t such that w agrees with the history of the actual world at every time before t.
\end{quote}

According to the Branch Principle, if the past has always existed, then anything that has always been true is necessarily true, since any possible world must “branch off” from the actual world some finite amount of time ago. The Branch Principle should be attractive to Aristotelians, since it accords with the idea that all mere possibilities are grounded in the powers of things in the actual world.

3.6 Objections to the Global Principle of Sufficient Reason/Causal Principle

3.6.1 Hume’s imagination argument.

David Hume provides the source for an argument against the necessity of an exceptionless causal principle: we can imagine some event occurring uncaused, and whatever we can imagine is possible.

However, it is not at all obvious that we really can imagine some event occurring without any cause at all. Is the absence of the imagination of a cause necessarily the imagination of its absence?

3.6.2 The inconsistency of a sufficient reason for contingencies.

Peter van Inwagen,\(^1\) James Ross,\(^1\) and William Rowe\(^1\) have all argued that it is inconsistent to hold both (i) that there are contingent facts, and (ii) that there is a sufficient reason for every contingent fact. The sufficient reason for the sum total of all contingent facts must be a necessary fact. If that necessary fact is truly a sufficient reason for any other fact, the second fact must also be necessary, since whatever is
necessitated by a necessary truth must be necessary itself. Hence, the Principle of Sufficient Reason entails that all truths are necessary, and so there is nothing for the supposed First Cause to cause.

This objection depends on two crucial assumptions: (1) no necessary proposition explains a contingent proposition, and (2) no contingent proposition explains itself.

Both are disputable. First, as Alexander Pruss has pointed out,\(^8\) there is a good ad hominem objection to (1): even if the PSR were false, it would still be a good explanation of the non-occurrence of some event to cite the non-occurrence of any possible cause of an event of that sort. However, given the falsity of the PSR, the non-occurrence of possible causes does not necessitate the non-occurrence of the event.

Second, statistical explanations seem fine, even though in those cases the explanans does not necessitate the explanandum. If it is very likely for a radium atom to decay in a given period of time, then that fact is a perfectly good explanation (a “sufficient reason”, in the relevant sense) of the atom’s actual decay.

In addition, anyone who accepts libertarian freedom and who rejects the randomness objection (the so-called Mind argument) must take the agent to be a non-necessitating cause of any of his free actions and so must either deny (1) or (2). For example, one could take a free action of a contingently existing being to be a self-explanatory fact, given explanations for the existence, freedom, knowledge and power of the agent.

The defender of the cosmological argument who rejects (1) must deny that every adequate or sufficient explanation is a contrastive explanation—in the strong sense of being logically incompatible with the explanation of any contrary fact. In denying (1), one need not suppose that each of the contrary choices would have been explained by the very same facts as the actual choice was. However, each actual choice has to be explained by a set of actual facts (concerning the beliefs and
values of the agent) despite the co-obtaining of certain other facts about the agent's beliefs and values that would have equally well explained the contrary choices.

3.6.3 Restricting the scope of the causal principle

It is possible to block the cosmological argument by replacing a broad causal principle with a much narrower one. For example, Graham Oppy has proposed the principle that all “non-first events” have causes.19 A “non-first” event is one that began at a certain time and was preceded in time by other events. Applying Oppy’s principle leads to the conclusion that the first causes consist entirely of first events: events that either stretch back infinitely far into the past or that occur at the first moment of time. In either case, such first events are not likely to lead to interesting theological conclusions.

However, Oppy’s proposal still leaves us vulnerable to the epistemological argument considered above. How do I know a priori that my current set of beliefs or experiences is not a first event? How do I know that the universe didn’t just begin a moment ago? If I have to take this seriously as a real possibility, one whose probability I cannot judge to be low, then I have an effective defeater for any claim to empirical knowledge.

Here’s an alternative proposal that might seem to be more successful: all events with a finite temporal bound in the past have causes. Once again, however, this generates skeptical possibilities: perhaps my present mental state has existed changelessly for an infinite amount of time.

A third possibility: all events that do not include an infinite regress of causes have a cause. This still fails the epistemological test. For all I could know a priori, my present state of consciousness could include such an infinite regress: an infinite series of mental states, each caused by and indistinguishable from its predecessors.

Laying aside these epistemological considerations, the skeptic might suppose that we have good reason to take beginningless events with infinite duration to be
plausible exceptions to any causal principle, on the grounds that it is obvious that such events couldn't be caused. There are at least three reasons for doubting this:

(a) Past unbounded and past bounded intervals are mathematically isomorphic. Given relativity, whether an interval has an infinite duration or not can be frame-relative. It's unlikely that such frame-relative features could have metaphysical implications.

(b) It's not obvious that a past-unbounded interval couldn't be preceded in time by a further set of events, each infinitely far in the past—just as the finite natural numbers are succeeded by infinitely large numbers in non-standard models of arithmetic.

(c) It's not obvious that causes must precede their effects in time. Both simultaneous causation and causation by atemporal entities seems possible.

3.6.4 Causing the causing

James Ross objected to the cosmological argument on the grounds that the causal principle leads to an infinite regress. The proposition that God caused the actual universe (U) is a contingently true state of affairs, call it C(U). So, C(U) needs to be caused by God, and this causing is a further state of affairs, C(C(U)). This needs to be caused, ad infinitum.

One might argue, as I did in my 1997 paper “A New Look at the Cosmological Argument”,20 that C(U) isn't some further state of affairs, but is just the sum of God and the existence of the universe. Or, one might deny that it is a wholly contingent state of affairs, since it seems to include God's existence. Its wholly contingent part C* might not be distinct from E. Following Thomas Aquinas, we might suppose that C(U) is simply identical to U itself, since “the actuality of the cause qua cause is the effect”. 

18
In addition, as Pruss has argued,\textsuperscript{21} the regress might be real but not vicious. C’s causing E is not a cause of E itself. C’s causing E is not more fundamental causally or explanatorily than C itself. These further causal facts are “epiphenomenal”, not part of the causal explanation of the event.

Finally, we could identify God’s causing the universe with God’s causal activity, which is identical to God himself, a necessary being. To think that this makes the universe itself necessary is to commit the de re/de dicto fallacy.\textsuperscript{22}

4. The Regress Problem

If we suppose that there is some version of a global Causal Principle or Principle of Sufficient Reason that can be known to be true, then the next problem to face is that of the Regress Problem: why doesn’t the principle entail an infinite regress of causes, with no stopping point? There are two basic approaches to solving the problem: arguing that infinite regresses are impossible, and arguing that they would be irrelevant, even if they were possible. The second strategy involves some kind of aggregation or conjunction of all of the wholly contingent or causable facts into one gigantic whole. If it can be shown that this whole, whether or not it contains one or more infinite regress within it, is itself wholly contingent (or causable), then the relevant Principle can be applied to the whole, resulting in a First Cause that is necessary and uncausable.

I’ll look at a Thomistic version of the first strategy in section 4.1,\textsuperscript{23} followed by three versions of the second strategy (aggregation, plural logic, set theoretic, and Leibnizian conjunction).

4.1 Aquinas’s Appeal to Per Se Causation

Aquinas denied the existence of an infinite chain of per se causes, not of merely per accidens ones (ST I, q. 46, a. 2, reply 7; SCG 2, 38). Aquinas argued that a chain of causes going backward in time could be infinite, since it would be a chain of merely
accidental or *per accidens* causes. The real or *per se* cause of the existence of each thing in the chain would be the timeless God: the previous members of the chain would be merely instruments used by God.

Aquinas clarifies what he means here by means of his hammer illustration. Consider a shoemaker who has made a pair of shoes. The shoemaker and his craft is the essential efficient cause of the existence of the pair of shoes. Let’s suppose that the use of a hammer is an indispensable part of the shoemaker’s craft. Then the involvement of at least one hammer would be part of the essential cause. However, the number of hammers would not be. Suppose that the shoemaker used several hammers in making this pair, because the first hammer wore out, the second was lost, the third borrowed by a neighbor, and so on. The number of hammers involved makes no difference to the origin of the shoes. Similarly, the number of ancestors that a person has is only an accidental feature of his cause, since human parents are only instruments God uses in causing the existence of particular men. God could have created an infinitely old universe, using infinitely many ancestors as instruments in the creation of each human being without introducing a *per se* causal regress, since the chain of essential causation would in each case terminate in God.

Suppose the skeptic believed that my present existence is *per se* caused by my own past existence, and, more remotely, by the past existence of my ancestors. In this case, Aquinas would insist that the causal chain could not be infinite, even if it were extended in time. Thus, the important distinction is between *per se* causation and accidental or *per accidens* causation, not between simultaneous causation and causation through time. Aquinas assumed that causation through time is always accidental causation, since he did not accept that a past event could be the essential cause of a present event (like my present existence). However, if a skeptic were to deny this and argue that we do receive our present existence from the past (by a kind of “inertia of existence” principle: i.e., that whatever exists tends to go on existing),
then Aquinas would deny that this chain of existence-receptions can go back to infinity—once again, because such a chain would fail to explain why anything in the chain (and the chain as a whole) has come to exist.

Why are per se causes needed at all? In the absence of any per se cause, the existence of a thing would be unintelligible. Thus, Aquinas’s argument depends on a tacit appeal to a version of the Principle of Sufficient Reason.

4.2 The Aggregation Move. (Al-Fārābī)

In my 1997 article,24 I recommended a revival of al-Fārābī’s solution to the regress problem. Al-Fārābī recommends that we take the aggregate of all contingent entities and then apply the causal principle to that aggregate, resulting in one or more necessary beings that collectively cause the aggregate of contingent things. I introduced the notion of being “wholly contingent”: an entity is wholly contingent if both it and all of its proper parts are contingent. I proved that every contingent entity contains a wholly contingent part. I applied al-Fārābī’s strategy to the aggregate of all wholly contingent things, which I called “the Cosmos”. I restricted the causal principle to wholly contingent “facts” or “states of affairs”, understood as concrete entities. I then proved that the Cosmos is itself wholly contingent. If we assume that all wholly contingent things have causes, we can conclude that the Cosmos has a cause. Since causes and effects are disjoint, the cause of the cosmos must have no wholly contingent parts. Hence, the cause of the Cosmos must be a necessary being.

Alexander Pruss has suggested a variant, defining the Cosmos as the aggregate of all positive, wholly contingent states of affairs. If we assume that all positive, wholly contingent states of affairs have positive causes, then we can reach the conclusion that there is a positive, necessary state of affairs that is the cause of the Cosmos.
The aggregation strategy dissolves the problem of possible infinite regresses by making them irrelevant. Even if some wholly contingent things do stand in an infinite causal regress, the whole regress will be part of the Cosmos, and so the causal principle will apply to the whole, forcing us to jump beyond the bounds of the Cosmos entirely.

The cogency of the aggregation move depends on assuming Mereological Universalism: the thesis that for every non-empty set S, there exists an aggregate of the members of S. Following Peter van Inwagen, many contemporary metaphysicians challenge the truth of Mereological Universalism. The defender of the aggregation move has several options. First, one could argue that universal aggregation is a true principle when applied to sets of facts or states of affairs, even if false when applied to arbitrary sets of material objects. Second, the cosmological argument would work if we could assume that there is an aggregate of the members of the set of the wholly contingent causes of any single fact exists. Third, we could abandon the aggregation strategy and employ instead one of the two strategies to be considered in sections 4.3 and 4.4 below.

4.3 Pluralizing the Argument

In the 1980's, logician George Boolos argued for treating plural quantification as logically primitive and indefinable (Boolos 1984). Plural quantification corresponds to the use of quantified plural pronouns in English: some sets, some facts, etc. It is easy to construct a version of the cosmological argument within plural logic. The causal principle would have to take the following form:

**Plural Causal Principle:**

if the x’s are some wholly contingent things, then there are some y’s such that the y’s are (collectively) the causes of the x’s, and none of the y’s overlap any of the x’s.
Plural logic tells us that if there are any F’s, then there are some things that include all and only the F’s. Thus, if there are any wholly contingent things, then there are some things that are (collectively) all and only the wholly contingent things. These things must have some causes that are disjoint from them. Thus, there must be some necessary things that are collectively the causes of all the wholly contingent things.

4.4 Leibnizian Conjunction

The German philosopher Leibniz introduced a logic-based method for effecting a kind of aggregation of contingent states of affairs. If we think of states of affairs as proposition-like entities, then we use the logical operation of conjunction to aggregate contingent propositions or facts. Here is the standard Leibnizian schema:

1. Every contingent fact has an explanation
2. There is a contingent fact that includes all other contingent facts, the “Big Conjunctive Fact”.
3. Therefore, there is an explanation of this fact.
4. The explanation must be a necessary fact, and one that involves a necessary being with causal power.
5. This powerful necessary being is God.

We can raise an immediate worry about this strategy: can we really form the Big Conjunctive Contingent Fact? Davey and Clifton raised some set-theoretic worries about such “big” conjunctions. Let p be the conjunction of all true propositions that do not contain themselves as proper parts (sub-formulae). Let q be the proposition that p is true. Is q a sub-formula of p? Either way a contradiction ensues.

Alexander Pruss suggested a strategy that avoids the need for such all-inclusive conjunctions. We have to assume instead that all contingent objects belong (of
necessity) to one or more natural kinds, and that there is a set $Q$ containing all of these natural kinds.\textsuperscript{27}

Let $p$ be the proposition that at least one of the kinds in $Q$ has at least one contingently existing member. This clearly can have an explanation: namely, the existence of some thing of a natural kind not in $Q$ that causes the existence of some member of $Q$.

Since the kinds in $Q$ are natural kinds, we can assume that no member of a kind in $Q$ is a member of that kind as a result of its own causal activity. Thus, there must be something with causal power that does not belong to $Q$. By definition of $Q$, this must be a causally powerful necessary being.

5. First Cause Arguments Without a Global Principle of Causal Explanation

In this section, I present several arguments for a cosmic first cause that do not appeal to a global causal principle or Principle of Sufficient Reason.

5.1 The Perpetuation of Time (Aquinas’s First Way)

The first argument is based on Thomas Aquinas’s First Way\textsuperscript{28} and depends on assuming the truth of the A-Theory. The A-Theory of time is the theory that there is a metaphysically unique Present Moment that is constantly in motion. The A-Theory implies the pure passage of time. As I argued above in section 4.1, the Kalām argument provides strong grounds for the A-Theory, regardless of whether we embrace the Branch Principle. Hence, there are real events constituted by the arrival of the Present to each new moment. If the A-Theory is true, then we can sensibly ask for the cause of the perpetual forward motion of time.

The hypothesis of inertia of movement or of existence does not solve the problem. Inertia presupposes the perpetuation of time—it cannot explain it.

If we assume, on Aristotelian grounds, that the cause of the forward motion of time consists in the activity of a substance (the “agent”), then we must ask whether
The agent acts in time or timelessly. Once again, the possibility of the activity of the agent in time can be excluded, since the agent’s activity has to be actual at a time in order to act as a cause at that time. Therefore, it cannot cause the actuality of the Presence of a moment of time t by acting at t, since it could act at t only if t is “already” (that is, independently of its own action) actual then. It cannot actualize at its own activity at t without vicious circularity. However, the cause cannot actualize t by acting at any other time, since there can be no action at a temporal distance—no action over a temporal gap.

Consequently, the agent that causes the arrival of the Present to each moment of time cannot be in time: it cannot act according to some time-parameterized law of nature. The cause of time’s motion must be timeless, and its mode of causation must be non-physical and non-mechanical.

The effect of this First Cause must be in time. In fact, it must be extended in time, since it consists in the progressive movement of the present through time. The only experience we have of a cause whose effect is temporally extended is that of a person effecting a plan or complex intention. Hence, we can reasonably infer that the activity of the First Cause is analogous to intentional action.

In order for the First Cause to be timeless, it must be not only unchanging but also absolutely unchangeable. Hence, the First Cause must be of necessity in a perpetual state of complete actuality of all of its potentials. In fact, the very distinction between actuality and potentiality would be inapplicable to it.

5.2 A Neo-Platonic argument from the Unity of the World

Even if we assume that the A-Theory is wrong, adopting the alternative B-Theory, we still confront a question that invites an inference to a First Cause: namely, what unifies the world’s spatiotemporal structure?
We don’t need to appeal to the Principle of Sufficient Reason or a global causal principle in order to justify the search for the unity of space and spacetime. The unity of space is constituted by an infinite number of precisely coordinated relational facts. For example, whenever A and B are separated by a distance of \( j \) meters, and B and C are separated by a distance of \( k \) meters, then A and C must be no closer than \( |j - k| \) and no farther apart than \( (j + k) \) meters. (If we adopt the block universe of the B-Theory, similar facts hold concerning the intervals separating the events of spacetime.) If there were no causal explanation of these facts, we would have an infinite number of brute, unexplained coincidences. This would be a problem, whether we adopt Relationalism or Substantivalism about space (or spacetime), since the two cases involve isomorphic sets of coincidences.

It is obvious that any possible explanation of the unity of space must be “outside” of space: a non-spatial and hence immaterial agent. In fact, we can go further, following the lead of Plotinus. Any composite entity involves coincidences: correlated relations among the parts and the whole. Suppose, for example, that some entity X contains two proper parts, Y and Z. Given the mereological principle of Companionship, the fact that Y is a proper part of X entails that X has some other proper part, and, similarly, the fact that Z is a proper part of X entails that X has some other proper part. The causal explanation of these coincidences cannot be a composite entity, since the unexplained unity of the cause would render it incapable of explaining the unity of its effect. Hence, we can infer the existence of at least one absolutely simple cause of the unity of space.

6. FROM FIRST CAUSE TO GOD

If there is a necessary being (perhaps even a being necessary in itself and hence uncausable in principle), what can we conclude about this being? Does such a being have any of the characteristics we traditionally associate with God? I have already
given reasons for some of the traditional characteristics: eternity (atemporality), immateriality, and absolute simplicity. We have some reason to believe that the First Cause acts in a way analogous to the execution of an intention. Can we go farther?

Jerome Gellman has produced a new argument for the uniqueness and the omnipotence of the First Cause. Gellman defines being a “creator” in a world as follows:

(10) \( N \) is a creator in \( w \) iff \( N \) is a necessary being whose activity (and inactivity) in \( w \) by itself explains all the contingent truths of \( w \).

He also assumes the following Iterative Principle:

Iterative Principle:

if \( x \) has the power to gain the power to do A, then \( x \) already has the power to do A.

From the Iterative Principle, it follows that if \( N \) is a creator in any world, then all of \( N \)'s powers are essential to it. Suppose, for reductio, that \( N \) is a creator in \( w \) and that \( N \) has some contingent power \( P \) in \( w \). Let \( \text{BP} \) be the aggregate of all of \( N \)'s contingent powers. \( N \)'s activity must explain the fact that \( N \) has \( \text{BP} \) in \( w \), since \( N \) is a creator in \( w \). This means that \( N \) has some causally prior power \( P \), whose exercise provides \( N \) with the powers in \( \text{BP} \). Since \( P \) is causally prior to \( \text{BP} \), and since \( \text{BP} \) includes all of \( N \)'s powers, \( P \) must be one of \( N \)'s essential powers. However, this prior power \( P \) would by itself constitute \( N \)'s having all the powers in \( \text{BP} \), given the Iterative Principle. Thus, \( N \) cannot have any contingent powers.

Gellman then proves that if \( N_1 \) is a creator in world \( w_1 \), and \( N_2 \) in world \( w_2 \), then \( N_1 = N_2 \). Here is Pruss’s version of the argument.
What explains in \( w_1 \), we may ask, why it is that \( N_1 \) exercised none of its powers to prevent \( N_1 \) from engaging in the kind of activity it engages in in \( w_1 \)? It must be that the explanation lies in the exercise of some power \( P \) by \( N_1 \) in \( w_1 \). But then \( N_1 \) also had this power in \( w_2 \) and did not exercise it, and its failure to exercise it must be explained by \( N_1 \)'s exercise of some preventative power \( Q \). But \( Q \) is one of the powers that enables \( N_2 \) to prevent something \( N_1 \) does in \( w_1 \), and so \( P \) includes the power to prevent \( N_1 \) from exercising \( Q \). Repeating the argument with the two entities and worlds swapped, we conclude that each of \( N_1 \) and \( N_2 \) has the power to prevent the other from preventing the other. But that is, surely, absurd!

The Gellman-Pruss argument depends on the plausible principle that if two entities possess essential causal powers and can co-exist in some possible world, and neither power is causally prior to the other in all worlds in which they both exist, then it must be possible for both to exercise those causal powers simultaneously. It is clearly impossible for each of two creators (in Gellman’s sense) to exercise the power to prevent the other from exercising any of its powers.

A more traditional argument for God’s uniqueness and perfection is the Thomistic argument from simplicity. St. Thomas attempted to show\(^{31}\) that a First Cause must be a being of pure existence, that there can be only one such being, and that any being of pure existence must possess every purely positive attribute. A being with every purely positive attribute is by definition “perfect”. Hence, there is of necessity exactly one necessary being.

St. Thomas’s argument\(^{32}\) turns on the claim that the First Cause's essence (or nature) is identical to its existence. Aquinas argues that, first of all, if the First Cause had some nature distinct from its existence, then the original state of affairs—the state of affairs of the FC's existing—would be composite, consisting of at least two elements, viz., that nature and its actual existence. Such a composite state of affairs cannot be necessary in itself, since any composite state of affairs must have a cause of its composition (on Plotinian grounds). Second, if the nature of the First Cause is distinct from its existence, then that FC could be a mere possibile. Its
nature could, in principle, be unactualized. Thus, there would have to be a cause of its actuality, contradicting its being necessary in itself.

If a thing’s nature were identical to its actual existence, that thing would have to be a being of “pure” existence. Its nature would simply be to be. Aquinas argues that, given this conclusion, we can infer that simple existence is a rich property, of such a kind that any being of simple existence would be supremely powerful. There are just two options: (i) to exist simply is to exist to an absolutely minimal degree, or (ii) to exist simply is to exist to an absolutely maximal degree. We can rule out the first option, since the First Cause is supremely powerful, and a being that exists to a minimal degree would be powerless.

The property of actual existence must be metaphysically compatible with all possible properties, whether positive or negative. If there were a logical incompatibility between existence and some property, then it would be impossible for a thing with that property to exist. However, it is possible that the property of existence defeasibly implies certain other properties. That is, it could be that we can reasonably infer that an existing thing has by default certain other properties, properties that follow from existence in the absence of contrary indications. A being of pure or simple existence would then possess all of the default implications of existence. Since such a being must exist to a maximal degree, we should assume that the property of existence defeasibly entails every purely positive property.

In order to define purely positive properties, let us start by trying to define the simply positive and simply negative properties. I will assume that we are working within a sparse theory of universals, according to which only certain properties (the most metaphysically fundamental ones) correspond to real universals. On this view, there are no negative or disjunctive universals, and no universals corresponding to adventitious concepts.
1. If U is universal, then the instantiation of U is a simply positive property, and the non-instantiation of U is simply negative.

2. The conjunction or disjunction (including infinite conjunctions and disjunctions) of simply positive properties is simply positive, and similarly for simply negative properties.

3. The partial existential generalization of a simply positive relation is simply positive, as is the partial universal generalization of a simply positive relation. Similarly for negative relations. That is, relational properties of individuals that are founded on simply positive relations are simply positive, and similarly for negative relational properties.

4. The possibilification and necessitation of a simply positive property is simply positive, and similarly for negative properties.

5. The negation of a simply positive property is simply negative, and the negation of a simply negative property is simply positive.

Given a definition of simply negative properties, we can define the class of purely positive properties:

**Definition of Purely Positive Properties:**

A property is purely positive if and only if exemplifying it does not entail (either strictly or by default) exemplifying any simply negative property.

Here are some plausible examples of purely positive properties: any power, any act of knowledge or understanding, love, goodness, existence and necessary existence.

Let us say that an essence or nature “includes” a property if having that essence entails (either strictly or defeasibly) having that property. Consequently, the essence of any species of bird includes having wings, even though it is possible for a bird to lack wings. It is enough that all bird natures have the property of being winged as a
default implication. Given this notion of inclusion and the definition of purely positive properties, we can now define ‘degree of existence’:

**Definition of Degrees of Existence:**

\[ x \text{ exists to a greater degree than } y \text{ iff } x \text{’s essence includes strictly more purely positive properties than does } y \text{’s (that is, } x \text{’s essence includes every purely positive property in the essence of } y, \text{ plus some additional ones not contained in } y \text{’s essence).} \]

There is a unique essence that corresponds to the maximum degree of existence: the essence that includes all and only the purely positive properties. There are many essences that correspond to minimal degrees, and there is probably no essence that is absolutely minimal, containing no purely positive properties.

Beings of pure existence have simple existence as their essence. Such beings exist to the maximum possible degree, which corresponds to what Aquinas describes as “perfection”. In contrast, all imperfect beings have an element of non-being in their essences. Aquinas argues that there cannot be two distinct perfect beings, since each would be numerically identical to the one, common essence. If both beings were distinct from their common essence, then neither could be a First Cause (necessary in itself), since the state of affairs of each one’s existing would be compound.

A perfect being would have essentially every possible power and would be disposed to know all actual truths and to love all actual beings, since such dispositions entail no simply negative properties.

**NOTES**

1 *Summa theologicae* I, q. 2, a. 3, co. [henceforth ST]

6 Ibid., 44ff.
8 I. Kant, The One Possible Basis for the Demonstration of the Existence of God, ed. and transl. G. Treash (Lincoln, Neb.: University of Nebraska Press, 1763/1994).


In both ST I, q. 3, a. 4, and the Summa contra gentiles I, 22.

Summa contra gentiles I, 22.
BIBLIOGRAPHY


