Organon F 33 (2) 2026: 196–217 ISSN 2585-7150 (online)
https://doi.org/10.31577/orgf.2026.33205

RESEARCH ARTICLE

Time Travel and the Limits of Logic: Why the Grandfather Paradox Continues to Resist Solution

Yael Loewenstein*

Received: 5 August 2025 / Accepted: 21 March 2026

Abstract: David Lewis's influential solution to the grandfather paradox treats time traveler Tim's failure to kill his infant grandfather as an ordinary case of unsuccessful action. Relative to the facts at the time of the attempt, Tim can succeed, and his failure is no more mysterious than missing a basketball shot. I argue that Lewis's solution is mistaken, and that a more recent alternative—the "strategy by analogy with the impossible," which likens Tim's failure to other failures to achieve logically or mathematically impossible tasks (e.g., crossing each of Königsberg's seven bridges exactly once)—also fails. Drawing a distinction between self-explicable and self-enforcing failures on the one hand, and Tim's failure on the other, it is shown that Tim's failure differs from ordinary failures to achieve the impossible insofar as the latter require no causal intervention: the impossibility itself fully explains and brings about the failure. Tim's case is structurally different. While it is impossible for his grandfather to both survive and not survive, nothing in the impossibility itself explains why his grandfather must survive rather than die—a selection between two contingent-seeming alternatives that ordinarily falls outside logic's purview and within causation's. The grandfather paradox therefore occupies an unrecognized middle ground: a necessary failure

* University of Houston
https://orcid.org/0000-0002-9512-0237
Philosophy Department, University of Houston, 3553 Cullen Boulevard, Houston, TX 77204-3004, USA
yrloewen@central.uh.edu

(relative to any facts) that nonetheless seems to demand causal enforcement, leaving the puzzle's deepest difficulty unresolved by either of the two leading strategies.

Keywords: Time travel; Grandfather Paradox; David Lewis; closed causal loops; backward causation; logical explanation; explanation.

1. The Enduring Mystery of the Grandfather Paradox

Picture this: Time-traveling Tim steps into his time machine in 2025 and journeys back to 1921 when his own grandfather is merely an infant, armed with murderous intent. At first glance, it seems Tim should be able to kill his infant grandfather—what will stop him if he tries? As David Lewis memorably puts it, "The forces of logic will not stay his hand! No powerful chaperone stands by to defend the past from interference" (1976, 149). And yet, assuming time is one-dimensional (i.e., without branching timelines), we know he won't succeed. If he did, Grandfather would never grow up to conceive Tim's parent, and Tim would never be born. In that case, Tim would both exist (a precondition for his time travel) and not exist (no adult grandfather, no Tim): a logical contradiction. This is the classic grandfather paradox.

What should we make of this puzzle? Does it show that, assuming one-dimensional time (an assumption that will be maintained throughout), we ought to reject the possibility of backward time travel? After all, accepting such travel seems to force a choice between two nearly equally unpalatable options: either embrace logical contradiction, or posit mysterious forces that intervene to prevent said contradiction.

Since Lewis's seminal (1976), most theorists agree that accepting the possibility of backward time travel needn't require such disagreeable commitments, though the reasons they give have evolved. Lewis argues that, contrary to appearances, there is nothing mysterious or problematic about Tim's (inevitable) failure to kill his infant grandfather. The mistake that leads us to erroneously think it mysterious, Lewis maintains, lies in a failure to distinguish two distinct contextual uses of 'can.' On the one hand, relative to facts regarding the time of the attempt, including facts about Tim's abilities and opportunities at that time, Tim can kill his grandfather. He fails, but there's nothing unusual about people failing to do things that they

are perfectly capable of doing. It is only when thinking about the 'can'-claim in a context in which we consider certain future-regarding facts, like that grandfather lived until adulthood and Tim was born, that we conclude that Tim can't kill his grandfather. But this is a different context of evaluation, and the 'can'-claim here has a different meaning than it did in the previous context, and so there is no contradiction in saying that Tim can (context 1), but also that he cannot (context 2).

This discrepancy, that on the one hand in some contexts of evaluation Tim can kill Grandfather, but when changing the facts relative to which the 'can'-claim is evaluated, he cannot, is an entirely ordinary phenomenon. Consider the parallel: compossible with facts regarding times sufficiently prior to lunchtime, I can drive to Austin, Texas, for lunch. It is in my capacity to make this choice and to follow through with it. On the other hand, driving to Austin for lunch is not compossible with the fact that, as it happens, I stayed in Houston for lunch; relative to this latter fact, I cannot drive to Austin, since to both drive and not drive to Austin would constitute a contradiction, and I cannot bring about contradictions.

So, Lewis's solution to the puzzle is as follows. Tim can kill grandfather relative to relevant facts at the time of the attempt, including facts about his abilities and opportunities, just as I could have driven to Austin for lunch relative to certain pre-lunchtime facts. Although at the time of the attempt he could have succeeded, he happened to fail for some ordinary contingent reason or other—perhaps a slippery banana peel or a series of missed shots—and his grandfather survived. Tim was subsequently born, enabling him to travel back in time and attempt the murder. No contradiction, no mystery.

Why then does Tim's failure to kill his grandfather seem more mysterious than other commonplace failures to do things one is capable of? Lewis (1976, 151) suggests that it is because the involvement of backward causation leads to confusion over which facts to evaluate the 'can'-claim relative to. At least for those of us who aren't fatalists, when we evaluate 'can'-claims, we standardly don't hold fixed facts regarding the future of the attempt. We think NBA star John Stockton could have made the championship-losing basketball shot in 1998 despite actually missing it. And similarly, when facts such as that Grandfather survived and Tim was born are

excluded from consideration, we see that Tim can kill his young grandfather (he has the ability and the opportunity…), and so his failure is as contingent as Stockton's missed shot.

If Tim's failure to kill Grandfather is contingent prior to the attack, the alleged problem with time travel dissolves. Tim can kill Grandfather, eliminating any need for mysterious logical enforcers to ensure failure. He simply happens to fail, as evidenced by his existence, and his failure enables his own eventual birth and subsequent time travel. Of course, we know that anyone traveling back and attempting to kill their own grandfather as an infant—or indeed, targeting any infant who will become a future grandfather—fails, not because failure was logically mandated at the time, but because there is evidence of that failure: we were told that the potential victim later became a grandfather.

2. Interlude: A Problematic Example

Before saying where Lewis goes wrong, let me pause to file a complaint against the example choice. The "grandfather paradox" has a nice ring to it; the scenario is fun and engaging, but beyond that, there is nothing special about this example. The point it is meant to illustrate is general: that without something like a powerful chaperone to defend the past from interference, time travel appears to allow for logically impossible, self-defeating causal chains.

The problem with using the standard example to illustrate this, however, is that killing one's grandfather before he conceives one's parent is not actually logically impossible (nor need it involve a self-defeating causal chain). For an act or outcome to be logically impossible, it must not occur at any possible world. And yet there are worlds—say, where resurrection occurs—at which killing one's infant grandfather doesn't prevent one's own existence and thus doesn't generate contradiction.

That killing one's own infant grandfather isn't logically impossible has made the issues and arguments surrounding the paradox trickier than they need to be. It invites unnecessary questions like: are the strange worlds where the killing occurs similar enough to the actual world to matter (and if not, why does their dissimilarity make them not matter?)? Do such worlds

undermine arguments that assume that killing one's own infant grandfather is logically impossible? But we need not worry about the answer to these questions: we can simply change the example! In particular, we can sidestep these (non)-issues by using an example that directly confronts the challenge of time travel. Instead of giving Tim the narrow goal of killing his grandfather, let's give him the broader aim of undermining his own existence. This can be attempted by any means appropriate to the world he's in. At some worlds, it may involve trying to kill his grandfather. At others, it might involve preventing a crucial encounter, sabotaging a birth, or any number of alternative actions. What matters is that he tries to bring about a self-undermining causal chain—whatever that takes at a given world.

With this subtle revision to the example, we eliminate the problem of distant worlds making Tim's success logically possible. Tim must fail his attempt—it is logically required that he does—for at no world can one undermine one's own existence. To do so would be to bring about that he both exists and doesn't exist, which is, of course, a contradiction.

So, I wish to change the example: Tim aims to undermine his own existence (whatever that takes), and so it is logically required that he fail. This revised example illustrates the puzzle for time travel just as well as the original: clearly, Tim will have to fail at his attempt to undermine his own existence, but without logical enforcers or similar, what is to stop him at worlds where there is time travel and where the killing would be self-undermining? How can "logic" ensure the presence of some (causal) foiling mechanism or other?

Other than allowing us to sidestep the problem of strange worlds, this revision has no impact on the arguments discussed here. However, revising the example in this manner is dialectically tricky for two reasons. First, in some contexts it is easier to talk about a specific action like trying to kill one's own grandfather. Second, the standard example is the one discussed by my interlocutors. I will thus continue to speak of Tim attempting to kill his own grandfather, but let us understand this merely as the method he uses at the actual world (or at whatever world we wish to evaluate the prospect of time travel). He may use other methods at other worlds. We can think about it like this: call the world of interest w_0. Tim is logically required to fail in his self-undermining endeavor. If killing his young

grandfather at w_0 would constitute such an endeavor, then he must fail to kill him at w_0. Now we need not worry about the problem of strange worlds.

3. The Fatal Flaw in Lewis's Solution

Returning to the main thread: despite its wide acceptance, Lewis's response to the grandfather paradox fundamentally misunderstands the nature of Tim's failure. Failing to complete a self-undermining act, such as killing one's young grandfather, is not a contingent failure comparable to missing a basketball shot. As I have previously argued (2022), even excluding from consideration all facts regarding times at and after Tim's arrival in the past, Tim still cannot kill his grandfather. Relative to past-regarding facts alone, Tim's failure to undermine his existence remains necessary. However, I now think that we can make an even stronger claim: holding fixed only the stipulations that Tim is attempting to undermine his own existence and that killing his young grandfather would be self-undermining at w_0, Tim cannot succeed relative to any facts whatsoever[1]. Not relative to Tim's abilities and opportunities, not relative to anything. That is because relative to no facts is it possible to bring about a logical contradiction. The description <Tim attempts to do something self-undermining (by trying to kill his grandfather)> renders the act impossible in just the way that <both making and missing one basketball shot simultaneously> proves impossible relative to any facts (including Stockton's abilities and opportunities).

There is an important objection to the claim that (contra Lewis) Tim cannot do anything self-undermining relative to any facts because he cannot bring about a logical contradiction relative to any facts. The objection is this: Tim wouldn't be in a position to do something self-undermining, such as attempt to kill his grandfather, if Tim hadn't been born, so the description of the act, e.g., <Tim time-travels from 2025 and tries to kill his grandfather in 1921> already presupposes post-1921 facts, such as the fact of

[1] 'Tim's grandfather' is intended attributively, not referentially, here; that is, it denotes whoever happens to stand in the grandfather relation to Tim, rather than some independently identified individual.

Tim's existence at 2025. We are tricked into thinking that his failure is necessary relative to any facts, the objection goes, because certain post-1921 facts are simply built into, or entailed by, the description of the scenario.

Jenann Ismael (2003) argues along similar lines in defense of the Lewisian-style solution; the idea of the failure being built into the description comes from her:

> …it is built into the description of the class of cases that we are considering [involving self- defeating causal chains] that they are failures, in the same way it is built into the description of the class of cases in which I don't manage to get a hold of my mom on the telephone that they are unsuccessful attempts at reaching her. For, when we describe a self-defeating causal chain, we sneak in, under the guise of the first event [e.g., Tim's attempt to kill grandfather in 1921], a description of the last [e.g., Tim entering the time machine in 2025] which is incompatible with the success of the operation as a whole. (308)

Ismael goes on to say that "In my bid to [travel back in time and] kill my infant self, for instance, there is the…certainty of failure, though failure be neither logically nor causally necessitated by the facts that obtain at the moment that I stand over my crib, gun in hand, finger on the trigger. The inevitability is a *purely epistemic one*, perfectly compatible with the logical and nomological possibility of success." (313, emphasis mine).

Interestingly, the truth of Ismael's two claims here comes apart. Although it is true that the failure is already built into the description, the second claim—that the failure is not logically necessitated by the facts that obtain at the moment she stands over the crib—is false. We can agree with Ismael that the description of the act gives away that the grandfather must have survived the attack, where the "must" here is epistemic. And of course, in most cases in which one fails to kill someone or to otherwise act causally toward a goal, the failure is contingent; so, it would ordinarily be a good assumption that the inevitability of the failure due to its incorporation into the description is merely epistemic. But we should notice that even granting that failure is built into the description and so an epistemic inevitability, it's still a further step to conclude that the inevitability is thereby only epistemic: this doesn't follow. In fact, the

inevitability is not only epistemic here—it is also logical. It is logically impossible to complete a self-undermining act (or to bring about any self-defeating causal chain, more generally), and so it was already impossible to succeed at the moment she stood over the crib.

Let us then be more explicit about the difference between Tim's failure and Ismael's failure to reach her mom by telephone. The description <Tim attempts to do something self-undermining, such as kill his infant grandfather>, does not selectively attend to instances of (contingent) failure the way the description <Ismael tries but fails to reach her mom> selectively attends to only the (contingent) failures out of a larger set of "calling attempts," some of which may be successes. Set language aside for a moment and consider that, regardless of what we write into the description of the act, it's a fact that prior to phoning a living person, it was causally and logically possible to reach them, even if they weren't ultimately reached. So, whether a given attempt is a failure—and thus whether it satisfies the description as an instance of trying and failing to reach her mom—is logically contingent prior to the call. In the time traveler scenario, in contrast, there is no selective attention only to Tim's failures to do something self-undermining, like killing his grandfather, because all attempts must be failures. They must all be failures because it is already impossible to succeed at undermining one's own existence, even prior to the attempt.

We can see the difference in another way. We can understand the description <Ismael does not get a hold of her mom by telephone> as composed of two key components: a description of an attempt for which success is possible (Ismael trying to reach her mother by telephone), combined with the additional stipulation that she failed. Contrast this with the description <Tim attempts to undermine his existence> or <Tim attempts to kill his young grandfather at w_0>. In these, we don't describe an act for which success is possible and then combine it with the stipulation of failure. There is no way to remove the failure from the description without entirely changing the nature of the act. We could make the failure contingent by, say, having Tim try to kill someone else (which, if successful, would not be self-defeating), but this would be a completely different kind of attempt and no longer of relevance here. Rather than describing an attempt that could be

successful and then stipulating failure, the description of Tim's attempt describes a kind of attempt for which the failure cannot be "removed" — success is impossible simpliciter.

So, while we can grant Ismael's (true) claim that the description of Tim's act already incorporates the truth of some future facts, this doesn't undermine our key point—a point that is sufficient to refute Lewis's solution: in contrast to Stockton's missed basket or Ismael's failed attempt at mom-calling, Tim's success is already impossible prior to the attempt in 1921. It is not, and never has been, contingent. Put another way: it is false that the failure merely happened to occur for some contingent reason or other and that we are judging it to be necessary only because we are assessing it relative to future-regarding facts.

4. The Strategy by Analogy with the Impossible

If Lewis and subsequent Lewisians about time travel were wrong about Tim's failure resembling Stockton's contingent miss, then perhaps a better strategy compares Tim's failure with other failed attempts to achieve contradictory or otherwise impossible feats, such as both making and missing the same basketball shot simultaneously, or accomplishing the mathematically impossible task of crossing each of Königsberg's seven bridges exactly once.[2] As it happens, this approach, which I call the "strategy by analogy with the impossible," has been gaining momentum in the recent literature on the topic.[3] The strategy begins by agreeing with Lewis that the question "what will ensure Tim fails to kill the young grandfather?" is misguided. No powerful chaperones defending against logical contradictions are needed, since, contrary to initial appearances, nothing mysterious characterizes

[2] Demonstrated to be mathematically impossible by Euler in 1735.

[3] See, for instance, Baron and Colyvan (2016), (2019) and Smith (2017). The same strategy has recently also been used to try to resolve other paradoxes, such as the Benardete paradoxes (Schmid and Malpass (2025)). In fact, I think it may work for these while still not resolving the grandfather paradox, but that's a topic for another time.

Tim's inevitable failure. But here Lewis's solution and the strategy by analogy with the impossible diverge: according to the latter, the correct analogy draws on other impossible tasks. That the time traveler will fail without help from logical enforcers is no more mysterious than the fact that no such enforcers are required to stop a painter from painting a chair both red and green all over, or to prevent any other impossible task. The intended feat's impossibility is sufficient explanation for why it cannot and will not be achieved.

I will argue that the strategy by analogy with the impossible also fails. This strategy implicitly depends on two key assumptions. First, that Tim's failure to kill the young grandfather is what I call 'self-explicable' in the manner of ordinary failures to achieve the impossible; and second, that Tim's failure is *self-enforcing* like other such failures. Self-explicable failures are failures for which the necessity of the failure is the only explanation needed for why it occurs. Self-enforcing failures are failures for which the mechanism of the failure—what brings it about—is intrinsic to the impossibility: there is no need for any external force, event, or intervention.

Self-Explicable Failure: A failure F to achieve an impossible task T is self-explicable if and only if:

1. F is logically necessary given the description of T;
2. The logical impossibility of T constitutes a complete explanation for why F occurs in every possible instance of attempting T; no additional causal, physical, or contingent facts are required to explain F's occurrence.

Self-Enforcing Failure: A failure F to achieve impossible task T is self-enforcing if and only if:

1. The impossibility of T guarantees F without requiring any contingent causal interventions or any external forces or events;
2. The mechanism preventing success is constitutive of the logical structure that makes T impossible.

Self-explicability concerns explanation: a necessary failure is self-explicable if no further explanation is needed—either for why the failure must occur or for why it does occur in a particular instance—beyond the impossibility itself. Self-enforcing relates to self-explicable but concerns not explanation

but rather what is needed to bring the failure about. If a failure to achieve the impossible is self-enforcing, nothing external to what makes the failure impossible is needed to bring about that failure. For ordinary failures to achieve the impossible, such as failing to both make and miss one basketball shot, these two features—being self-explicable and self-enforcing—make the failures (i) explicable (fully explained) and (ii) not in need of external enforcement to bring them about or to ensure they occur.

To clarify these concepts, let me loosely categorize types of impossible-task failures. These categories are not meant to be exhaustive—perhaps there are other kinds of failures to achieve the impossible that don't fit neatly into either one. One category of failure to achieve the impossible comprises failures arising from the non-existence of something required to accomplish the task: failing to visit a nonexistent location, find a nonexistent object, or use a nonexistent tool. Call these *non-existence failures*. More broadly, non-existence failures can be classified as a kind of failure of opportunity, where the opportunity's absence stems from the non-existence of something. Holding fixed that it doesn't exist, finding Atlantis is impossible: i.e., at all worlds lacking Atlantis, Atlantis will not be found.

Other examples of non-existence failures come from Nicholas Smith (2017), who argues that Tim's failure to kill his grandfather in 1921 is as explicable as other failures to achieve the impossible, including failing to enter a non-existent attic or to find a non-existent fountain of youth. We can classify both examples as non-existence failures because the failures can be explained by the non-existence of the sought thing.

Non-existence failures are clearly self-explicable. Something's non-existence is all the explanation needed for why any attempt to find or use it will fail. Moreover, non-existence failures are plainly self-enforcing. Nothing external—certainly no mysterious force or logical enforcer—is needed to stop searchers from finding what doesn't exist. It is self-enforcing because the thing's nonexistence, by itself, makes it such that it cannot and will not be found.

Smith's examples are obvious non-existence failures, but I believe we can categorize certain other, less obvious examples that theorists claim are analogous to Tim's failure as non-existence failures as well. Consider Sam Baron and Mark Colyvan's (2019) example of Bridget attempting to cross

each of Königsberg's seven bridges exactly once. Euler proved that there is no path crossing all seven bridges without backtracking across at least one. In fact, he proved something stronger: there cannot be such a path. But if there cannot be a path, then trivially, there isn't one. The fact that there isn't one provides all the explanation needed for why Bridget fails.

When explaining Bridget's failure, providing the reason why no path exists is not required, just as the reason no fountain of youth exists is unnecessary for explaining why someone fails to find one. That there isn't a path suffices; no path, no success. Thus, it's appropriate to categorize Bridget's failure as a nonexistence failure: she failed because there was no path that crossed each bridge exactly once. As a non-existence failure, it is both self-explicable and self-enforcing: the path's non-existence itself explains and enforces Bridget's failure to find it.

Returning to Tim, his failure to kill his young grandfather is not best classified as a non-existence failure. Tim doesn't fail because some object, place, or path doesn't exist. Tim doesn't lack the opportunity to succeed. Quite the contrary, what makes the scenario seem so mysterious is precisely that he appears to have the perfect opportunity: we can imagine his helpless young grandfather directly before him, Tim possessing the ability, the means, and every apparent advantage. If Tim's failure isn't a non-existence failure, what kind is it?

A second category of impossible-task failures consists of what I call *inconsistency failures*. Inconsistency failures are failures to achieve both of two mutually inconsistent acts, or to bring about both of two mutually inconsistent states or events. Examples include failing to both make and miss one basketball shot, painting a chair simultaneously both red and green all over, or, using another example from Smith (2017), a Newcastle barber failing each time he sets out to shave all and only the barbers in Newcastle who don't shave themselves. The latter constitutes an inconsistency failure: shaving all is inconsistent with shaving only, since if the barber shaves all, then he has shaved himself and thus not shaved only, while if he shaves only, then he hasn't shaved himself and thus not shaved all.

Tim's failure to kill his young grandfather appears appropriately categorized as an inconsistency failure. By aiming to kill him, Tim aims to bring about two inconsistent states: Tim both exists and doesn't exist. If we're

using the strategy by analogy with the impossible to deflate the grandfather paradox's mystery, inconsistency failures are plausibly the best (i.e., most analogous) examples to employ. Like non-existence failures, inconsistency failures are both self-explicable and self-enforcing. So how do they work? How does the impossibility, in this case the inconsistency, bring about the failure? Certainly, the impossibility is not a causal factor that makes the failure occur. So, what then is the method or mechanism by which the impossible task is stopped?

The answer is so obvious in cases of paradigmatic inconsistency that the question may sound ridiculous. First, we should note that which of the two inconsistent states or events occurs (if either) is causally determined and contingent. For example, whether the basketball shot is made or missed is causally determined and contingent. It could have been different: the other event could have occurred, or, if the two are not mutually exhaustive, it could have been that neither occurred.

Second, by occurring, the contingent state by itself precludes the actualization of events or states with which it is inconsistent. Crucially, this preclusion isn't merely abstract and theoretical—it doesn't just preclude because it's inconsistent, and that's all there is to say. The preclusion is quite tangible: as soon as I paint a chair red all over, it is thereby not green, because to be all red is to not be any other color. Not being any other color (nor uncolored) is an alternative description for the very same state of being red. Since being red, it is already not green; the failure to make it both red and green simultaneously is self-enforcing: nothing (causal) must happen to stop someone from making the chair both at once. The painter need not trip or bring the wrong color paint. It is also self-explicable: that a chair is already not green by being red (or vice versa) by itself explains why anyone who tries to make it both red and green will fail.

Importantly, because inconsistency failures are self-enforcing, no contingent accident or foiling need occur to bring them about. What's more, because they are self-explicable, no contingent accident or foiling need be part of the explanation for the failure. Indeed, I think that it is a mistake to think that a contingent event can be part of the explanation for why

the failure occurs.[4] Whether the basketball player who is trying to both make and miss a basketball shot at once ultimately stops trying because of getting tired or because of twisting her ankle, makes no difference. None of these explains why she fails to achieve something impossible—even why she fails in that token instance, since she could not have achieved it, in any case. That she got tired and stopped trying can be part of a description of what happened during the attempt, but that's all it is. It's not an explanation for why she failed. After all, had she not gotten tired when she did, she still would have failed: her getting tired (or slipping on a banana peel, or…) is no difference-maker and so not part of the explanation for why she failed.[5]

To make it even clearer that the explanation for why the necessary failure occurs is not causal, imagine that our basketball player is just about perfect: immortal, error-free, and with unremitting luck. She can use her collection of infinite basketballs to try to both make and miss one shot repeatedly forever, and she will never succeed (borderline cases don't count as either). Even with never getting tired and with nothing ever happening to stop her, she still won't find success. This makes sense since the failure is self-enforcing in the way that we've seen: if the basketball shot is made, it is thereby already not missed, and if it is missed, it is as such not made. The same is true for Bridget. If Bridget is immortal and error-free and if nothing ever happens to stop her, she will still fail to cross each bridge of Königsberg exactly once. Whatever happens, one cannot succeed at finding a path that does not exist, and so what actually happens is irrelevant. That the path does not exist explains why she fails in any given instance.

[4] Baron and Colyvan (2019), who defend a version of the strategy by analogy with the impossible, argue that when talking about a specific token instance of failing to do something impossible, there will be a contingent causal explanation for why the failure as a matter of fact occurred. I think they're wrong for the reasons given here.

[5] Most accounts of explanation take explanation to be closely tied with difference-making. Baron and Colyvan (2019) themselves seem to endorse the idea that an explanation should be a difference-maker (see their discussion on p. 257), yet strangely, they accept causal explanations of token necessary failures. Strange because causal events are not difference makers for such failures!

5. Why Tim's Case Differs Crucially

Return to Tim's failure to kill baby Grandfather in 1921. Were he to succeed, he would bring about that Grandfather both did and did not survive the attack; that he both did and did not grow to adulthood; that Tim both did and did not come into existence, etc. Of course, this isn't possible. To understand what distinguishes Tim's case from other failures to achieve the impossible, let's test the analogy to standard inconsistency failures. If Tim's failure were analogous to, say, the failure to paint a chair both entirely red and entirely green, we should expect the following:

(i) It is contingent whether Grandfather survives the attack or not (just as it is contingent whether the chair is painted red or not).
(ii) It is logically impossible that Grandfather both survives and does not survive.
(iii) The impossibility of simultaneous survival and non-survival is self-enforcing and self-explicable.

Conditions (ii) and (iii) are met. If grandfather survives, he thereby does not fail to survive, and vice versa. This is precisely the kind of mutual exclusion that makes paradigmatic inconsistency failures self-enforcing: the actualization of one state automatically rules out the other. The problem is that condition (i) fails. It is not contingent whether Grandfather survives. As we've already seen, his survival is necessary even before the attempt.

The difficulty, then, is this. Logic generally doesn't "choose" which of two inconsistent states obtains; it merely prohibits both from obtaining together. In standard inconsistency cases, it remains causally contingent which outcome occurs. For instance, whether the basketball shot is made or missed is determined by causal events; logic simply bars the possibility of it being both. But in Tim's case, logic appears to do more: it not only rules out both outcomes jointly, but it demands that one specific outcome occur.

This makes the grandfather paradox crucially different from other examples of inconsistency failures, and we can now be precise about the difference. There is an inconsistency failure involved here: killing Grandfather would bring about an inconsistency. But the key point is this: it's not the inconsistency failure, or how it's enforced, that's inexplicable! The fact that

if he's dead then he's not alive and vice versa is as explicable as any other inconsistency failure. What is inexplicable concerns the necessity of Grandfather surviving rather than not surviving. Logic is dictating which of two mutually exclusive outcomes must obtain; something it ordinarily doesn't, and arguably can't, do. The survival of Grandfather rather than his death seems to require a causal explanation, yet here it is logically mandated. Logic, in this case, appears to reach beyond its domain and act as a kind of pseudo-causal force.

That Grandfather lives rather than dies is neither self-explicable nor self-enforcing in the way that not both living and dying is. If Tim must fail and there's no self-enforcing mechanism built into the situation, something external/contingent must cause failure—some banana peel, jammed gun, or fluke accident. (Contrast again with Bridget: her failure needs no foiling mechanism at all.) And if, as previously argued, the failure didn't merely happen to (contingently) occur, then even without holding fixed facts about what happens causally downstream, it was already necessary that there would be some kind of foiling event were Tim to attempt the feat. How can logic make the causal sequence behave appropriately?

Here's a helpful contrast. Consider again the painter who attempts to make a chair both red and green all over. We know she will fail, and we know why she will fail—her failure is self-enforcing. Nothing needs to intervene to stop her. But now imagine a strange variant: suppose it is logically required that she fail to paint the chair red. That is, logic itself somehow mandates that she specifically fails to paint it red. What would stop her, then, if she tries? Something must happen each time she attempts it—she'll run out of red paint, get distracted, slip, or be stopped by a mysterious force. Suddenly, we've introduced the need for causal intervention to ensure a logically required outcome, making logic inexplicably tied to the machinery of causation. Yet it is this scenario, where the painter must fail to paint the chair red, to which Tim's failure is relevantly analogous, since his failure is not contingent prior to the attempt. This unnatural hybrid—logic dictating physical outcomes—mirrors exactly what's problematic about Tim's case.

6. Replies to Objections

So far, I have argued that standard failures to achieve the impossible are self-explicable and self-enforcing, whereas Tim's failure to kill his grandfather is neither. If Tim's failure is not self-enforcing, then something external to the impossibility—presumably some causal process—is needed to stop Tim. But how logic can enforce itself through causation remains mysterious. However, while we've seen why failures like Bridget's (to cross each bridge once) or the painter's (to paint a chair red and green all over) are both self-explicable and self-enforcing, more ought to be said about why Tim's failure isn't either one. Could it be that his failure to kill his grandfather is self-explicable and self-enforcing, after all?

The difference I've identified is this: in the case of ordinary failures to do the impossible, the failure and what brings the failure about are internal to the impossibility. A painter trying to paint a chair entirely red and entirely green at once fails because being red just is, in part, not being green. The contradiction lies within the properties themselves. Bridget's failure can be understood as a failure to find a nonexistent path; or just as good, we can equally understand the failure as due to a structural impossibility: Bridget fails to cross each bridge in Königsberg exactly once because of the bridge topology itself. One can draw it out and see that, given the layout, crossing any six bridges leaves one stranded where the seventh bridge is unreachable without backtracking. The very structure that creates the impossibility also explains and enforces the failure.

Now consider Tim's failure. He must fail because to kill his grandfather would be self-defeating. We know his success is logically impossible, but what specifically stops him? We are looking for an answer that is logical or structural and not causal; something intrinsic to the impossibility. The tempting answer is something like, "Tim exists, so grandfather must have survived. Since grandfather lived, he couldn't have also died—in just the way that the chair being red makes it impossible for it to also be green." This response takes Tim's failure to be a normal inconsistency failure: that grandfather lives itself makes it necessary (and brings it about) that he therefore doesn't die; or, that Tim exists itself makes it necessary and brings it about that he doesn't not exist.

But this explanation fails because it assumes what it needs to explain. The question isn't, "given that grandfather survives, what ensures that he doesn't also die?" but rather "what is in place to ensure that grandfather survives—something that was already required prior to the attack—in the first place?" Compare with a scenario in which it is inexplicably logically necessary (relative to any facts) that a particular basketball shot is made. If asked to explain what mechanism ensures the causal chain results in it being made rather than missed, it would clearly be insufficient to reply: "the causal chain does not result in a miss because the shot was made and thus could not also be missed". Like the reply above, this response assumes what needs explaining. We need to know what ensures that the causal chain results in the shot being made in the first place, without presupposing that very event. Similarly, we need to know what ensures that the causal chain produces grandfather's survival without presupposing that survival.

It would be different if grandfather's survival were contingent, which the above reply implicitly assumes. It assumes it by assuming that the necessity of the failure need not be accounted for. If his survival were contingent relative to facts prior to the attack, then the fact that there happened to be a banana peel, or that Tim happened to miss, explains Tim's failure—and the fact that he failed means he couldn't have also succeeded. But we've seen that relative to any possible facts, some foiling mechanism or other had to be in place. Even before the attempt, it was logically necessary that the causal chain cooperate to ensure Tim fails if he tries. This suggests that the (necessary) presence of some foiling mechanism or other still needs to be accounted for.

But Tim's grandfather only has to survive if the attempt was made by Tim (or anyone else for whom grandfather's death would be self-undermining), one may reply. So, grandfather only has to survive on the assumption that Tim is alive, and Tim is only alive because grandfather survives. As such, if Tim is alive, then grandfather does survive; and given that he does, he cannot have also died—nothing mysterious about that. To this I respond just as I responded to Ismael's argument before. It is true that if Tim is alive, then grandfather survived. Tim's existence is certainly evidence that grandfather did. But this still does not entail that it was ever contingent that grandfather survive this specific attack. And, to repeat, if it wasn't

contingent, then an explanation of the enforcement appealing only to "well, he did survive (as we know from Tim's existence)" is not sufficient. Even prior to the attack, he had to survive an attack by Tim, and the causal chain had to cooperate so that he would survive if Tim attacked. The objector may insist: but Tim wouldn't have been there at all if the attack had succeeded, and so Tim's presence indicates that it (contingently) failed. This isn't correct, either. There would have been no possibility of Tim's attack occurring in the first place without Tim, so Tim's existence in itself cannot show that the attack's failure was contingent: the attack was contingent—without Tim, that specific attack couldn't have occurred—but failure was always necessary if it did occur. If the failure is not contingent, we still require an explanation for its enforcement.

Finally, one may object that the non-causal explanation for why Tim fails to kill grandfather is simply that doing so would be self-undermining—i.e., that it would bring about a logical contradiction—and that's all there is to say. But a response like this just sidesteps the issue. No mechanism of failure is given here. The claim amounts to "it is impossible, and so the causal chain will act accordingly". But this brings the mystery right back. How does his grandfather's death "being impossible" ensure the presence of banana peels or the like? I grant that such a reply could have been compelling enough, however unsatisfying, if "it's impossible!" were all that could be said to explain the presence of a causal foiling mechanism for other failures to achieve the impossible. But that is not the case, for a couple of reasons. First, the other failures don't seem to need a foiling mechanism in the same kind of way: we've seen that the immortal painter with unlimited paint and no need for rest could attempt to paint the chair both red and green forever. Nothing would have to stop him, and still, he'd never succeed. Likewise for Bridget: she can walk the bridges of Königsberg forever, with nothing ever happening to stop her, and she would never succeed. That she would continue to be "stranded" after crossing any six bridges—i.e., unable to cross the seventh without backtracking—is not a causal foiling mechanism but a structural one.

Now try to imagine what it would look like for an immortal Tim to continuously try to kill his young grandfather if there were never something preventing him causally. He could repeatedly try to do things that would

ordinarily result in death, and the baby could just repeatedly fail to die. Unless there is a causal explanation for why the baby survives each time, this is as mysterious as ever—just pure magic. Unlike with the others, the explanation for Tim's failure, it seems, will always be causal.

The second reason to reject the explanation "it is logically impossible for Tim to kill his grandfather, so he can't, and that's all there is to say", is that, as I have argued, for other failures to achieve the impossible, if asked to produce a non-causal explanation of why the failure comes about, "it's just impossible" is not all that can be said. Generally, we can describe why the failure occurs without mentioning anything causal. For instance, we can appeal to the nature of color properties, to the topological layout of Königsberg, to the nature of the property of being dead (it is to be not alive), and so on. There seems to be nothing of this kind to appeal to, to explain why Grandfather survives rather than dies.

This suggests that grandfather-type paradoxes occupy a unique logical space: they are neither garden-variety contingent failures (which can be explained causally) nor standard necessary failures (which are self-explicable and self-enforcing). They appear to be necessary failures that nonetheless require causal explanation. And so, we still face the question, "how can logic enforce itself through causation?", and the mystery remains.

Before concluding, let me note that the problem vanishes if the outcome in question isn't self-undermining. Suppose that someone else—say, Tim's friend, John, travels back in time to try to kill Tim's grandfather. We know John will fail since Tim exists. But here, unlike with Tim's attempt, the inevitability of the failure is merely epistemic. Prior to the act, John could have succeeded. Of course, had he succeeded, Tim never would have been born, but that's no problem—Tim's birth was contingent, too. The point is that John's attempt is not self-undermining, so his failure doesn't require logical enforcement; it only requires historical consistency. No paradox arises.

7. Conclusion: The Persistence of Mystery

Where does this leave us? While sophisticated solutions to dissolve the apparent mystery underlying the grandfather paradox have been proposed, each attempt ultimately fails because of an enduring misunderstanding

about what makes the paradox genuinely puzzling. Lewis's original solution, comparing Tim's failure to contingent everyday failures, foundered on the necessity of Tim's failure. The more recent strategy by analogy with the impossible, while initially more promising, fails because Tim's failure to kill his grandfather lacks the self-enforcing character that makes ordinary failures to achieve the impossible unproblematic.

The heart of the difficulty lies in a crucial asymmetry. In paradigmatic cases of impossible tasks, the impossibility itself does all the work. Ordinary necessary failures are self-enforcing because the very nature of the impossible task ensures that success cannot occur. Nothing external needs to intervene. But Tim's case is fundamentally different. Logic demands not merely that some inconsistency be avoided, but that a specific one of two mutually exclusive possibilities obtain rather than the other. This selection between alternatives lies outside logic's purview—it requires causal influence on physical phenomena. Logic is not a force that can causally bring about the presence of well-placed banana peels or fortuitous misses. How, then, can it ensure that some foiling mechanism or other, each contingent, will frustrate the attempt?

If the grandfather paradox cannot be dissolved through analogy with ordinary impossibilities, and if Lewis's contingency-based approach fails to address the paradox's core difficulty, then we face a dilemma. Either we must abandon the possibility of backward time travel in one-dimensional time, or we must accept that such travel requires mechanisms that violate our basic understanding of how logic constrains physical reality. Logic may not directly stay Tim's hand, but it seems to require that something else do so—and that requirement remains the paradox's most troubling aspect.[6]

References

Baron, Sam. and Colyvan, Mark. 2016. "Time Enough for Explanation." *Journal of Philosophy*, 113(2): 61–88.

Baron, Sam., and Colyvan, Mark. 2019. "The End of Mystery." *American Philosophical Quarterly*, 56(3): 247–64.

[6] I am grateful to Brian Garrett for probing comments and discussion, some of which inspired me to write §2.

Ismael, J. 2003. "Closed Causal Loops and the Bilking Problem." *Synthese*, 136: 305–20.

Lewis, David. 1976. "The Paradoxes of Time Travel." *American Philosophical Quarterly*, 13: 145–52.

Loewenstein, Yael. 2022. Against the Standard Solution to the Grandfather Paradox." *Synthese*, 200, 172.

Schmid, Joseph and Malpass, Alex. 2025. "Benardete Paradoxes, Causal Finitism and the Unsatisfiable Pair Diagnosis." *Mind*, 134(534): 397–421.

Smith, Nicholas J.J. 2017. Time Travel and the Grandfather Paradox. In *The Oxford Handbook of Philosophy of Time*, edited by Craig Callender, 456–75. Oxford: Oxford University Press.