Contents

Organon F 33 (2) 2026: 152–153 ISSN 2585-7150 (online)
https://doi.org/10.31577/orgf.2026.33201

Introduction to the Special Issue on David Lewis's "The Paradoxes of Time Travel"

David Lewis's "The Paradoxes of Time Travel" (1976) is the paper in the philosophy of time travel. Lewis's paper has influenced all subsequent discussions of time travel, and the stance of any philosopher on this topic can be gauged by the extent of his agreement or disagreement with Lewis.

In his paper, Lewis laid out distinctions and doctrines necessary to make sense of time travel in general and travel to the past in particular: a distinction between (what he called) personal time and external time; a Four-Dimensional view of persons and other continuants; and an Eternalist framework in which past, present, and future are equally real. Travel to the past requires backwards causation and may sometimes generate causal loops. Travel to the past also gives rise to the famous Grandfather Paradox.

Lewis's overall conclusion is that there are no genuine paradoxes of time travel and that time travel is no more conceptually or metaphysically problematic than space travel. Time travel in our universe may be technically or even physically impossible, but it is not metaphysically impossible.

Now, however convincing Lewis may be at first read, critics can and have dissented from him at every stage. Should the personal time/external time distinction be drawn in the way Lewis drew it, or is there a better way? Can time travel occur on the Three-Dimensional view of continuants, or on a theory of time other than Eternalism? Is backwards causation really possible, and is there not something especially paradoxical about causal loops? Finally, does Lewis resolve the Grandfather Paradox successfully, and might one question the relativistic theory of 'can'-judgements underlying his solution?

Depending on how these questions are answered, critics will range from moderate to extreme: from those who accept the possibility of time travel but disagree on details to

those who conclude that time travel (or, at least, travel to the past) is impossible. But all his critics owe Lewis a great debt for having made explicit the questions that need to be addressed.

The debate continues 50 years after the publication of Lewis's paper. The contributions in this Special Issue, beginning with an overview by Ryan Wasserman, pay homage to Lewis's seminal paper and are testaments to the fecundity and longevity of Lewis's ideas.

The Editors would like to dedicate this Special Issue to the memory of a good metaphysician and a good man: Hugh Rice, formerly of Christ Church, Oxford, who passed away in March 2025.

Brian Garrett (ANU)
Jeremiah Joven Joaquin (DLSU)

Organon F 33 (2) 2026: 154–161 ISSN 2585-7150 (online)
https://doi.org/10.31577/orgf.2026.33202

RESEARCH ARTICLE

Concluding Notes on "The Paradoxes of Time Travel"

Brian Garrett*

Received: 2 August 2025 / Accepted: 13 November 2025

Abstract: I outline some objections to Lewis's (1976) theory of abilities. I then discuss the threat of Logical Fatalism. I criticise Van Cleve's riposte to Lewis's paper and highlight a curious argument of Lewis's. I end by commenting on Lewis's solution to the Grandfather Paradox and matters arising.

Keywords: David Lewis; time travel; abilities; fatalism; Grandfather Paradox.

1. Lewis's Theory of Abilities

From pp. 150–51 of Lewis (1976) it's reasonable to attribute to Lewis the following theory of abilities (Lewis's Thesis):

> (LT) Relative to certain facts, A can (can't) F if and only if those facts are compatible (incompatible) with A's doing F.

Given Lewis's liking for simple ideas I assume that by 'compatible' ('compossible' in the text) he meant 'logically compatible'. The theory would be too unwieldy otherwise.

* Australian National University
https://orcid.org/0000-0002-0604-0260
School of Philosophy, Research School of Social Sciences, Australian National University, ACT 2601, Australia
brian.garrett@anu.edu.au

I record three immediate counterexamples to (LT). (i) I can't speak Finnish, relative to the fact that I have never learnt it; but my speaking Finnish is logically compatible with my not having learnt it. (ii) An ape can't speak Finnish, relative to facts about its larynx and nervous system; but an ape's speaking Finnish is logically compatible with the facts about its larynx and nervous system. (iii) My speaking Finnish is logically compatible with Trump being President. By (LT) it follows that I can speak Finnish, relative to the fact that Trump is President. But this is absurd. The fact that Trump is President is irrelevant to my ability to speak Finnish. Moral: not all compatibilities yield abilities.

There are, of course, other objections to (LT). See, e.g., the next section; Matthew, Schultheis and Boylan 2017; Schwarz 2020; Maier and Kikkert 2025: 4.1–4.3. I should also mention Van Cleve's amusing reptilian objection. On the assumption that no fact is "better" than any other, Van Cleve complained that on Lewis's theory "…there would be no answer to the question whether the man menaced by the crocodile can escape. Relative to some facts he can, relative to others he can't, and that would be the end of the matter. There would no fact about whether he can escape all things considered or absolutely." (Van Cleve, 173)

These objections prompt us to move in one of two directions: either agree with Lewis that 'can'-judgements are always, explicitly or implicitly, relative to background facts but offer a different account of the truth-conditions of "Relative to certain facts, A can (can't) F" or else give up on relativistic accounts altogether (as Van Cleve seemed to recommend).

2. Logical Fatalism

The Logical Fatalist thinks that truths about future actions constrain what we can do, and that we can only do what we will do. To avoid this unpalatable result, facts about the future should be discounted in saying what an individual can or can't do. A simple example illustrates the point. Suppose that the President won't be killed next year. This fact is incompatible with anyone killing him then. But we should not conclude, as (LT) would have us do, that relative to the fact that the President won't be killed, no one can kill him. Rather, the fact that the President won't be

killed, since it's a fact about the future, should be discounted in saying what anyone can or can't do. What is true is that anyone with the means and opportunity can kill the President next year, only he won't, thus avoiding the Logical Fatalist's conflation of 'won't' with 'can't'.

Contrary to (LT), then, not all incompatibilities imply inabilities. In particular, incompatibility with a fact about the future doesn't yield an inability. We thus have another class of counterexamples to (LT). In fairness, Lewis was aware that we do not ordinarily count facts about the future in saying what someone can do. (Lewis 1976, 151) So presumably he would have agreed that facts about the future should be discounted.

There's one exception to this general rule, however. In the case of a traveller to the past, some facts about the future should not be discounted in saying what he can or can't do. In Lewis's story, Tim travelled back to 1921 with the intention of killing Grandfather before Father was conceived. Relative to 1921 there are facts about the future causally necessary for Tim's existence, and hence causally necessary for his presence in 1921. These facts should be counted in saying what Tim can or can't do. Grandfather's continued existence is necessary for Father's birth, and Father's birth is necessary for Tim's existence. Hence, Grandfather's continued existence and siring of Father, though a fact about the future, is a fact relative to which Tim can't kill Grandfather in 1921.

However, facts about the future causally unnecessary for Tim's existence should be discounted in saying what Tim can or can't do. The fact that no one shook Grandfather's hand in 1922, we may suppose, is such a fact. It isn't true that relative to the fact that no one shook Grandfather's hand, Tim can't shake his hand. Tim can shake Grandfather's hand, relative to facts about his means and opportunity, but he won't, since no one shook Grandfather's hand in that year.[1]

[1] I prescind from discussing the challenge to free action posed by the thesis of causal determinism since that challenge, successful or not, concerns all agents, not just travellers to the past.

3. Rejoinder to Van Cleve

J. Van Cleve's recent intervention requires comment. (Van Cleve, 2019) Van Cleve is right to highlight the threat of Logical Fatalism, but wrong to think that Lewis can avoid Logical Fatalism by "relativising", i.e., by holding that 'can'-judgements are not detachable from the background facts relative to which they are asserted. (Van Cleve, 173) Indeed Van Cleve is mistaken, twice over.

First, Lewis is a relativist about 'can'-judgements (*vide* (LT)). Lewis would concede that there's no absolute or all-things-considered 'can'. Van Cleve is thus offering Lewis his own theory as a way out. Second, relativisation doesn't automatically extinguish the threat of Logical Fatalism. Unless facts about the future are deliberately excluded from the class of admissible background facts, Logical Fatalism will supervene.

4. A Curious Argument

At one point Lewis wrote: "Tim cannot kill Grandfather. Grandfather lived, so to kill him would be to change the past." (Lewis 1976, 150) Later in the same paragraph he wrote: "It is logically impossible that Tim should change the past by killing Grandfather in 1921. So, Tim cannot kill Grandfather." (Lewis 1976, 150) These quotations seem to express the same argument. Two readings are possible.

On one reading, the argument endorsed in these remarks is the simple but invalid: A; $\sim\Diamond$(A & B); so $\sim\Diamond$B; where A = Grandfather lived through 1921, B = Tim killed Grandfather in 1921, and A contradicts B. I assume that $\sim\Diamond$B implies that Tim can't kill Grandfather in 1921, and that the impossibility of changing the past implies $\sim\Diamond$(A & B).

Alan Hajek has urged another reading of Lewis's argument which treats its major premise as a counterfactual (corresponding to the 'would' and 'should' in the quotations). We then get:

(i) A; so
(ii) B $\Box\!\!\rightarrow$ (A & B); but
(iii) $\sim\Diamond$(A & B); so

(iv) $\sim\Diamond B$.

One problem with this argument is that (i) doesn't imply (ii). A different example makes the fallacy vivid. From "I lived through yesterday" it doesn't follow that "Had I been killed yesterday, I would have lived through yesterday and have been killed yesterday."

On either reading, then, Lewis's argument is invalid.

5. Lewis's Solution to the Grandfather Paradox

The just mooted confusion has implications for Lewis's solution to the Grandfather Paradox. Lewis wrote:

> We have this seeming contradiction: "*Tim doesn't [kill Grandfather], but can, because he has what it takes" versus "Tim doesn't, and can't, because it's logically impossible to change the past.*" I reply that there's no contradiction. Both conclusions are true, and for the reasons given. They are compatible because "can" is equivocal. (Lewis 1976, 150)

The second italicised sentence, if I am right, is false. However, this doesn't invalidate Lewis's solution since we can replace the defective sentence with "*Tim doesn't, and can't, kill Grandfather because Grandfather's continued existence and siring of Father is causally necessary for Tim's existence.*" As Lewis observed: "No Grandfather, no Father; no Father, no Tim; no Tim, no killing." (Lewis 1976, 152)

There are possible worlds in which Grandfather's continued existence isn't causally necessary for Tim's existence. For example, there are worlds in which Tim travels back in time, kills Grandfather, and Grandfather is immediately resurrected. However, we may assume, the world of Lewis's story isn't one of these worlds.

Lewis's solution to the Grandfather Paradox can thus go through as advertised. If we focus only on local facts pertaining to Tim and Grandfather in 1921 (in particular, facts about Tim's proximity to Grandfather, his means and opportunity, etc.,), we get the (ordinary) sense in which Tim can kill Grandfather. If we focus instead on global facts (including, crucially, the fact that Grandfather's continued existence and siring of Father is

causally necessary for Tim's existence), we get the (extraordinary) sense in which Tim can't kill Grandfather or do anything else that would prevent Father's conception.[2]

6. Four Related Points

(i) Consider Grandfather's business partner, Mycroft. The continued existence of Mycroft isn't causally necessary for Tim's existence. Hence, there's no extraordinary sense in which Tim can't kill him. Moreover, it would be wrong to think that Tim can't kill Mycroft in 1921 because he will be assassinated in 1934, felled by a poisoned dart on the steps of the Diogenes Club. For the same reason it would be wrong to think that Tim can't kill Grandfather in 1921 because he wasn't killed then.

Unfortunately, Lewis made just this mistake. He cited, as a reason why Tim can't kill Grandfather in 1921, "…the simple fact that Grandfather was not killed [then]." (Lewis 1976, 151) There is a reason why Tim can't kill Grandfather in 1921–Tim's killing Grandfather would undermine a condition for his (Tim's) existence–but that's not the reason given here by Lewis.

A simple example may clarify matters. Suppose a certain garden gate was open all day on 1 May 1921. Tim strolls past the gate on that day. Two questions: (1) Is there a sense in which Tim can close the gate? (2) Is there a sense in which he can't close the gate? My answer to (1): Yes, since Tim has the means and opportunity to close the gate, he can do so, though he won't. My answer to (2): No, if Tim has the means and opportunity, there's no sense in which he can't close the gate. The fact that the gate was open all day implies only that no one will close it, not that it can't be closed. Unfortunately, the Logical Fatalist, and, I fear, Lewis, give the wrong answer to (2).

[2] Suppose that Tim's parents met only because of the Second World War and that the war would not have happened if Hitler had not risen to power in Germany. Does this yield a sense in which Tim can't travel to Munich in 1921 and kill Hitler? (No Hitler, no parents; no parents, no Tim; no Tim, no killing.)

(ii) The extraordinary sense in which Tim can't kill Grandfather shows that Lewis was wrong to urge a parallel between the abilities of Tim and Tom. Tom is a normal inhabitant of 1921, similar in age and physique to Tim. (Lewis 1976, 149–51) But there's a crucial difference between them. Unlike Tim, there's no one in 1921 whose continued existence is causally necessary for Tom's existence. If Tom can't kill someone that's for the usual reason, *viz.*, lack of means or opportunity.

(iii) People fall too easily into Logical Fatalist ways of thinking in the case of Tim. But they shouldn't. Suppose Tim travels from 1975 to 3000. We don't have Logical Fatalist intuitions about Tim when he arrives in 3000, so why do we have them when arrives in 1921? (This question has most force if Eternalism–the doctrine that past, present and future are equally real–is assumed.) However, this much is true: the extraordinary sense in which Tim can't kill Grandfather implies that Tim's freedom of action is slightly diminished compared to other inhabitants in 1921.

(iv) In notes discovered after his death, Lewis proposed a new account of abilities. In the Abstract to his notes, written by the Editors of the Special Issue, Lewis's later view is summarised thus: "S is able to A if and only if there are no obstacles to his A-ing, where an obstacle is a "robust preventer": something that would (or does) cause S not to A, and which "wouldn't go away if things were just a little different." (Lewis 2020, 241) However, this account fails to validate the extraordinary sense in which Tim can't kill Grandfather, and it also fails to validate more mundane inabilities such as my inability to sit and stand at the same time. There's no 'obstacle' to success in these cases, certainly not on any causal reading.

7. Coda: The Death of Tim

Lewis never told us the amusing end to Tim's story, so let me do so here. After a series of comical attempts at assassination, Tim and Grandfather meet up. Over a few beers, Tim tells Gramps his whole sorry tale. The dialogue between the smart Grandfather and the not so bright Tim then goes as follows:

> *Grandfather*: Tim, I was just wondering, do I ever get a criminal record?
> *Tim*: Gosh, no. We had a joke in our house that old Gramps never got so much as a speeding ticket!
> *Grandfather*: Oh, that's good to know.

Grandfather then shoots Tim dead.

8. Conclusion

In sum: I outlined some objections to Lewis's theory of abilities. I then discussed the threat of Logical Fatalism. I criticised Van Cleve's riposte to Lewis's paper and noted a curious argument of Lewis's. In closing I made some remarks about Lewis's solution to the Grandfather Paradox and matters arising.

Acknowledgement

Thanks to Alan Hajek, Brian Hedden, Yael Loewenstein, Alex Miller, Joseph Schmid, Nicholas J.J. Smith, Daniel Stoljar, and Ryan Wasserman.

References

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

Lewis, David. 2020. "Outline of '*Nihil Obstat*: An Analysis of Ability.'" *The Monist,* 103(3): 241–44.

Maier, John and Kikkert, Sophie. 2025. "Abilities". In *The Stanford Encyclopedia of Philosophy* (Summer 2025 Edition), Edward N. Zalta and Uri Nodelman (eds.). URL: https://plato.stanford.edu/archives/sum2025/entries/abilities

Mandelkern, Matthew; Schultheis, Ginger and Boylan, David. 2017. "Agentive Modals." *Philosophical Review*, 126(3): 301–43.

Schwarz, Wolfgang. 2020. "Ability and Possibility." *Philosophers' Imprint*, 20(6):1–21.

Van Cleve, James. 2019. "Lewis and Taylor as Partners in Sin." *Acta Analytica*, 34: 165–75.

Organon F 33 (2) 2026: 162–181 ISSN 2585-7150 (online)
https://doi.org/10.31577/orgf.2026.33203

RESEARCH ARTICLE

The Paradoxes of Time Travel: 50 Years Later

Ryan Wasserman*

Received: 6 October 2025 / Accepted: 17 December 2025

Abstract: David Lewis set the agenda for philosophical theorizing about time travel fifty years ago with his landmark paper "The Paradoxes of Time Travel". In honor of that anniversary, I review some of the central themes of his paper and discuss a few of the ongoing debates inspired by that work. In particular, I address some worries that have been raised for Lewis's characterization of time travel, his treatment of backward causation, and his proposed solution to the grandfather paradox.

Keywords: Time travel; David Lewis, backward causation; Grandfather Paradox.

"The Paradoxes of Time Travel" is a philosophical gem with many beautiful facets that reflect the brilliance of its creator. Originating as an informal talk for students and later developed into a book project, the work was eventually published in *American Philosophical Quarterly* in 1976.[1] It would go on to inspire a generation of philosophers and remains the starting

[1] For a history of the project and unpublished book chapters, see Lewis (2023).

* Western Washington University
https://orcid.org/0000-0003-2202-2131
Department of Philosophy, Western Washington University, Bellingham, WA 98225, USA
ryan.wasserman@wwu.edu

point for all serious work on the topic.[2] I cannot hope to do justice to the whole paper here, so I will limit my attention to three central themes: Lewis's characterization of time travel (section 1), his comments on backward causation (section 2), and his treatment of the grandfather paradox (section 3).[3] In each case, I will explain Lewis's position, review what has been said in response, and provide my own assessment of the debate. I will then conclude (in section 4) by reflecting on the continuing impact of Lewis's paper.

1. The Definition of Time Travel

Time travel can seem incoherent by its very definition. After all, ordinary trips through space involve being in different *places* at different times. So, time travel would presumably require being in different *times* at different times. But it is unclear what this would even mean. Worse yet, as soon as we begin to describe a case of time travel, we appear to led into contradictory statements. Suppose, for example, that Marty gets into his time machine in 1985, hits 88 mph, and then–at the very next moment–finds himself in 1955. In that case, Marty's arrival comes just *a moment after* his departure, but it also happens *thirty years before*. This seems impossible. "How can it be that the same two events, his departure and his arrival, are separated by two unequal amounts of time?" (Lewis 1976, 145)

Many thinkers–including philosophers, scientists, and science fiction authors–have tried to avoid these problems by reimagining time.[4] Normally, we think of events as unfolding along a single dimension where any two points are separated by a unique distance.[5] (See the figure below, where the 'D' marks Marty's departure and the 'A' his arrival.)

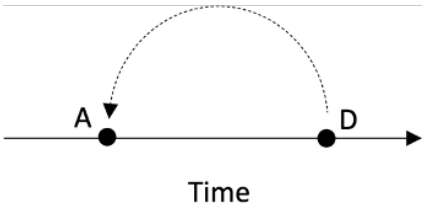
---

[2] For an overview of the literature, see Wasserman (2018).

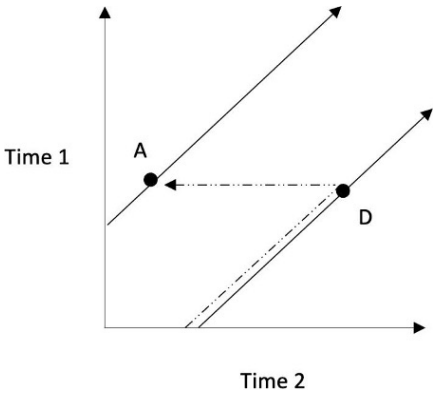
[3] One thing I will not be discussing are Lewis's comments on persistence and personal identity. For a more developed statement of his views, see Lewis (1983). For a discussion of how these things relate to time travel, see Wasserman (2018, Chapter 6).

[4] See, e.g., Meiland (1974), Deutsch (1997), and Daniels (1935).

[5] Lewis understands time to be one dimension of a four-dimensional manifold (ibid., 145–46); for our purposes, a single line will do.



But we could instead think of time as a plane with room for multiple universes to run in parallel. (See Figure 2.):



In this image, the solid angled lines depict different world-histories, and the dashed line represents Marty's journey through time. Importantly, this figure includes two different temporal dimensions, which allows us to say that Marty's departure and arrival are separated by unequal amounts of time– the departure is 30 years after the arrival in Time 2, but just a moment earlier according to Time 1. It also allows us to say that Marty's journey involves being at different times (of Time 1) at different times (of Time 2). In this way, we can avoid both of the initial worries outlined above.[6]

[6] Lewis credits the two-dimensional model to Meiland (1974) and Asimov (1955), but related ideas go back at least to Daniels (1935). For more recent developments of this approach, see Goddu (2003), van Inwagen (2010), and Law (2018).

In his paper, Lewis acknowledges the possibility of this model but questions its usefulness. When a time traveler visits his past, he is sup- posed to visit *his* past–that is, *the past where he previously existed*. Jumping to a parallel universe might provide a similar experience, but it is not the real thing. For this reason, Lewis claims that it does not give us "time travel as we know it from the stories" (ibid., 145).

More importantly, Lewis argues that the two-dimensional model is unnecessary, since we can make sense of temporal discrepancies within a single timeline. The key is to introduce two different ways of ordering events on that line and measuring the distances between them:

> I…distinguish…time itself, *external time* as I shall also call it, from the *personal time* of a particular time traveler: roughly, that which is measured by his wristwatch. His journey takes an hour of his personal time, let us say; his wristwatch reads an hour later at arrival than at departure. But the arrival is more than an hour after the departure in external time, if he travels toward the future; or the arrival is before the departure in external time (or less than an hour after), if he travels toward the past. (Ibid., 146)

Looking back at the first figure above, Lewis would say that, according to external time, Marty's arrival in 1955 comes first. Then, many years later, Marty gets into the time machine and hits 88 mph. That is the ordering of those events in real time. From Marty's perspective, however, the adventure begins with his getting into the machine in 1985. After that, he hits 88 mph and arrives in the past. That is the order of the events–*the very same events*–according to his personal time.

Lewis's picture, then, is one in which all the elements of a time-travel story occur along a single timeline. Moreover, he argues that this timeline is immutable. If Marty *is* (according to his personal time) going to arrive safely in 1955, then it is *already* the case (according to external time) that he arrived. If Marty *will* (personally) interact with his teenage father in 1955, then that is *already* (objectively) a fact before he gets into the time machine.

Importantly, Marty's actions in 1985 will still have an *effect* on the past–his hitting 88 mph will, for example, be a *cause* of his arrival in 1955–but

nothing Marty does will *alter* the past. Backward time travelers, in this picture, simply help to bring about the history that was.

Lewis concludes that the distinction between external and personal time allows us to give a consistent description of time travel without invoking parallel universes, rewritten histories, or other theoretical posits. Moreover, he suggests that the distinction provides for a perfectly general definition of time travel: for someone to be a time traveler is just for there to be a discrepancy between their personal time and the objective temporal relations that hold between events (ibid., 146–47). Marty's 1985-experiences, for example, come *before* his 1955-experiences in his personal time, but *later* according to external time. Or consider H.G. Well's famous example of a time traveler who journeys continuously into the future. That character experiences events in the same order as everyone else, but, from his perspective, things are sped up, so that a journey lasting thousands of external years takes less than a day of personal time.[7] According to Lewis, it is discrepancies of this kind that constitute time travel.[8]

Lewis's picture has proven extremely popular.[9] Nonetheless, worries remain. Here, I will focus on three problems related to Lewis's description of personal time, his definition of time travel, and his appeal to external time.

Lewis initially equates personal time with "that which is measured by [a] wristwatch" (ibid., 146). But, as he himself notes, the mere fact that someone's watch has gotten out of step does not make that person a time traveler. He therefore suggests the following account:

---

[7] Yet another example would be the case of the time traveler who takes an hour of personal time to travel an hour into the past. This would involve a discrepancy of or- der without any difference in distance.

[8] Lewis also famously characterizes time travelers in terms of their four-dimensional shapes (ibid., 146). Marty, for example, would be a "broken streak" with one segment ending in 1985 and another beginning in 1955. Wells's time traveler, meanwhile, would be a "continuous streak" which is "stretched out" relative to ordinary individuals. (The stretching, in this case, means that the mental and biological processes of the time traveler take longer to complete than those of an ordinary individual–see below for further details).

[9] Lewis's definition of time travel has been endorsed by Horwich (1987, 114), Dowe (2000, 441–42) and many others. The main alternative is to characterize time travel using the tools of relativity–see below.

> If you take the stages of a common person, they manifest certain regularities with respect to external time…Memories accumulate. Food digests. Hair grows. Wristwatch hands move. If you take the stages of a time traveler instead, they do not manifest the common regularities with respect to external time. But there is one way to assign coordinates to the time traveler's stages… so that the regu- larities that hold with respect to this assignment match those that commonly hold with respect to external time (1976, 146).

In Marty's case, his 1955 arrival comes after his 1985 departure (according to his personal time), since that sequence shows the hair growth, memory accumulation, etc. that we would normally see in the life of an ordinary person. That is what makes the arrival the "next" event in Marty's personal history. In contrast, imagine that Marty returns to 1985 after spending a week in the past and that he arrives at the precise time and place of his earlier departure. In that case, Marty's 1985 arrival-stages will follow immediately after his 1985 departure-stage in external time. However, to an outside observer, it will look as if Marty has instantaneously aged by a week (with longer hair, fingernails, etc.). Obviously, we do not typically see discontinuities like this in the lives of ordinary people. That is why the 1985 arrival-stages do not come "next" in Marty's personal time.

Lewis's suggestion may seem promising, but it is also problematic. The most obvious issue is that his characterization of personal time only applies to persons and other biological entities that undergo such changes. Moreover, there is no way to generalize Lewis's approach, since some potential time travelers are intrinsically unchanging. To make the worry as vivid as possible, imagine a single persisting electron in an otherwise empty world. Is that particle moving normally toward the future or continuously into the past? Whichever way you look at it, the pattern will remain the same–there is just one unit of negative charge after another. As a result, there is not enough structure to tell us whether there is a discrepancy between external time and the "personal" time of the electron.

A better option–suggested by Dowe (2000, 441), Wasserman (2018, Chapter 1, section 1), and others–is to characterize personal time in terms

of causation.[10] When Marty arrives in 1955, his hair, memories, and other features will closely match those of his 1985 departure. But those features are not *just* similar–they also depend, causally speaking, on what Marty was like when he hit 88 mph. In fact, Marty's very existence in 1955 is the direct causal result of what happened 30 years in the future. The same thing would be true for a lonely electron that is traveling continuously toward the past–at each moment, its existence and features will depend, causally speaking, on what it is like at later moments of external time. This would distinguish it from an ordinary, forward-moving electron, in which each stage depends on those in the external past. Something similar will presumably be true for any other kind of entity, since it is commonly thought that persistence through time (in either direction) requires causal dependence. We can therefore think of personal time, in general, as the order and metric provided by whatever causal relations are constitutive of identity through time.

Even if we grant Lewis his notion of personal time, we might still worry about his definition of time travel. One objection concerns the case of suspended animation. In the Marvel Cinematic Universe, Captain America crashes into the Arctic in the 1940s and is encased in ice for nearly 70 years. When he awakes, Cap is largely unchanged since his biological processes were "frozen" in time. Since these processes are constitutive of personal time for Lewis, he will have to say that there is a discrepancy between external time and Cap's personal time. But that seems problematic, since we would not normally take this to be a case of time travel.

It is less clear what to say about this case on my characterization, since I have not given a precise definition of "causal distance." When a time traveler discontinuously jumps into the future, there are no intermediate stages, so the causal distance from departure to arrival is zero. In Cap's case, his body is present throughout the 70-year period, and causal connections hold at each stage. So, perhaps the causal distance is 70 years as well. That would eliminate any discrepancy with external time, in which case Cap would not be a time traveler.

[10] Lewis himself thinks causation is required for personal identity, but he does not mention it in his characterization of personal time.

But now go back to the Wellsian time traveler. Like Captain America, he is continuously present throughout his journey. But unlike Cap, he is conscious during his trip and sees everything else sped up. However, if an external observer were to pass by the time machine, she would see something different: everything would, from her perspective, be moving at the normal pace, whereas the *time traveler* would appear to be in slow motion. Indeed, depending on the details, the time traveler might look just as "frozen" as Captain America. But, in that case, we seem to have a problem with *any* definition of time travel since we want to say that Wells's character is a time traveler and that Captain America is not. At this point, one might conclude that suspended animation *is* time travel or that Wellsian travel is *not*. But it would be even better if we could identify a relevant difference between the cases. One suggestion is made by Lewis himself in a letter to Jonathan Bennett:

> I'm not sure about cryogenic procedures. They slow down different processes to different degrees, and the rates of the fundamental atomic processes… are unchanged. So it's not clear to me that if we take a functional analysis of personal time… we get a discrepancy between personal and external time. (Quoted in Wasserman 2018, 13.)

Applied to the case of Captain America, the thought would be that his *biological* processes are frozen, but his basic subatomic particles are acting normally. If personal time is characterized in terms of these more fundamental processes, there would no longer be any discrepancy with external time. Moreover, if Wellsian time travel slows down *every* physical process, we would then be able to draw a distinction between the two cases.[11]

A third and final worry for Lewis concerns his appeal to external time. Lewis says little about this notion in his paper, beyond calling it "time itself." But for his definition of time travel to work, the external temporal relations between events would have to be independent of a particular individual's perspective, since that would be the only way to get a mismatch between personal and external time. The problem is that, on the standard

[11] For further discussion of this worry, see Wasserman (2018, Chapter 1, section 2).

interpretation of relativity, there is no objective temporal ordering for all events. In this sense, there is no (global) notion of real time. But if there is no real time, then there cannot be a mismatch between *it* and personal time.

One response to this worry is to say that non-relativistic worlds are possible and that Lewisian time travel could still occur in such worlds. As Lewis himself notes, a time-travel world might have to be "different in fundamental ways from the world we think is ours" (ibid., 145). However, many have argued that relativity allows for journeys through spacetime that appear to be time travel. One of the most famous examples involves the so-called *time dilation effect* of special relativity. To illustrate, suppose that two twins–Alice and Bob–start off at rest with respect to each other on the surface of the Earth. Bob then blasts off in a rocket ship, traveling near the speed of light. In a situation like this, relativity tells us that Bob's clock will appear to be running slow from Alice's perspective (and that Bob will see Alice's clock as being sped up). Importantly, this will be true for both wristwatches *and* biological clocks. So, if Bob travels far enough away and then returns to Earth (maintaining his high speed throughout), he will be visibly younger than Alice–perhaps even years younger–when he arrives. In this respect, he is like a Wellsian time traveler. But Bob's youthful appearance would not be due to a discrepancy between personal time and "real" time. Rather, the two individuals will simply have taken different paths through spacetime, and Bob's interval will be shorter, due to a difference in acceleration. This gives us a mismatch between the "proper time" of the two individuals, but it does not create a discrepancy with external time. Hence, this would not be a case of Lewisian time travel.

It is unclear what Lewis should say about this example. He could try to revise his definition, or he could maintain his account and insist that Bob's journey is not genuine time travel. More plausibly, he could simply say that there is semantic indeterminacy in this case, owing to our limited and sometimes conflicting talk of time travel. Since nothing important hinges on the issue, we can live with that indeterminacy.[12]

[12] For a classic statement of Lewis's view on linguistic use and semantic indeterminacy, see Lewis (1986a, 212). For more on relativity and time travel (including a discussion of general relativity), see Wasserman (2018, Chapter 2, sections 5 and 6).

2. The Problems of Backward Causation

If time travel involves discrepancies between personal time and external time, and the former is understood in terms of causation, then backward time travel would reverse causal sequences. This idea raises several concerns. Here, I will focus on three issues from Lewis's paper: his proposed analysis of causation, his attitude toward causal loops, and his discussion of self-undermining acts.

In his earlier work, Lewis (1973) argues that one event is a cause of another just in case those events are linked by a chain of counterfactual dependence (that is, a series of events in which each member would not have occurred without the previous event). For example, the engine wouldn't have started without a signal from the ignition, and the ignition wouldn't have signaled if I hadn't turned the key. So, my turning the key is a cause of the engine's starting.

To apply this to the case of time travel, we must first say a bit about Lewis's theory of counterfactuals. In general, a counterfactual of the form "A > C" is true just in case all the closest worlds where A is the case are also worlds where C is the case. Closeness, in turn, is taken to be a similarity relation governed by the following set of rules:

(1) It is of first importance to avoid big, widespread, diverse violations of law.
(2) It is of second importance to maximize the spatiotemporal region throughout which perfect match of particular fact prevails.
(3) It is of third importance to avoid even small, localized, simple violations of law.
(4) It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly. (1986b, 47–48)

To illustrate, let w0 be the world in which Marty gets into his time machine, hits 88 mph, and appears in 1955. Let w1 be a world very much like w0, but where–contrary to the laws of w0–a few extra neurons fire in Marty's brain after he gets into the time machine and he chooses *not* to hit 88 mph. In that world, no one shows up in 1955. Finally, let w2 be a world like w1, but where a perfect duplicate of Marty miraculously shows up out of nowhere in 1955 (at the exact time and place of Marty's actual appearance in

w0). Because of that miraculous appearance, w2 will match w0 up until the time at which the extra neurons fire in 1985. So, w2 does a better job than w1 with respect to (2). But we can suppose that the sudden appearance of Marty's duplicate in that world would count as a "large miracle," relative to w0, since it would require billions of atoms to appear in a highly organized way–thus breaking Lewis's very first rule. If this holds for all worlds where a duplicate appears, it follows that no one would have appeared in 1955 had Marty not hit 88 mph. Lewis's theory of causation, therefore, says that Marty's hitting 88 mph was a cause of the appearance in 1955. For this reason, Lewis concludes that his theory of causation does not "rule out causal reversal *a priori*" (1976, 148).

This much is correct, but there is also a worry. Suppose that Marty's friend, Doc Brown, has a teleporter that he uses to send a single electron (all by itself) back to 1955. We can then ask: what would have happened had Doc not turned on his device? Let w3 and w4 both be worlds very much like w0, but where a few extra neurons miraculously fire in Doc's brain so that he does not turn on the machine. In w3, no extra electron appears in 1955; in w4, one does (at the exact time and place that one appears in w0). As in the previous case, the second world will score better with respect to (2), since it includes 30 extra years of exact match. However, unlike the case of Marty, this would *not* require a violation of (1), since a single electron's appearance would not constitute a big, widespread, diverse violation of law. That world *will* score slightly lower with respect to (3), since it will include one additional small miracle (corresponding to the electron's appearance in 1955). But this will not outweigh the decisive win on (2). Hence, Lewis will have to say that an electron would have appeared in 1955, even if Doc had *not* turned on the machine. Given his theory of causation, it would follow that the electron's appearance is not caused by future events. Assuming that causation is required for time travel, it would follow that the relevant electron is *not* a visitor from the future. Hence, Lewis's account seems to implausibly discriminate between would-be time travelers on the basis of their size.[13]

Even if one believes that causal reversals are not ruled out by the nature of causation, one might still worry about the phenomenon for other reasons.

[13] See Wasserman (2015) for further discussion.

One of the most common concerns is about the possibility of *causal loops*. Lewis gives the example of a traveler who goes back in time, meets his younger self, and passes on information about how to build a time machine (where this information is not available any other way). The younger self then uses that information to build a time machine, travel back in time, and meet his younger self. "But where did the information come from in the first place?" asked Lewis. "Why did the whole affair happen? There is simply no answer. The parts of the loop are explicable, the whole of it is not" (ibid., 149).[14]

Lewis's treatment of this case has always struck me as correct:

> Strange! But not impossible, and not too different from inexplicabilities we are already inured to. Almost everyone agrees that God, or the Big Bang, or the entire infinite past of the universe, or the decay of a tritium atom, is uncaused and inexplicable. Then if these are possible, why not also the inexplicable causal loops that arise in time travel? (Ibid., 149)

To this, I would add that a question like "Where did the information come from in the first place?" is ambiguous, since "first place" could refer to an initial point in external or personal time. On the first reading, the question has a perfectly good answer, since the information first shows up (in external time) with the older time traveler from the future. On the second reading, the question presupposes a false assumption, namely that there is a first moment in the information's personal history. From the information's perspective, there is no such time–it came from the older traveler, who got it from the younger man, who got it from the older traveler, who… In this respect, the information is like the "entire infinite past of the universe" that Lewis mentions in the quote above: it does not have a causal beginning.

[14] "Information loops" of this kind can be contrasted with "object loops." Suppose, for example, that Lewis's time traveler gives his younger self a set of printed out plans, which the younger self uses to build a time machine so that he can take back the very same sheets of paper to his younger self. Loops like this will require an object to revert back to its original state when traveling in time (e.g., coffee stains will have to disap- pear, torn corners will need to be replaced, etc.). This kind of process would be highly unusual, but not impossible. See Hanley (2004, section 4).

But, in that case, it seems strange to ask for a cause. If the series never started, why require a starter?[15]

Even if we accept the possibility of a causal loop (in which each event is a cause of the next and where the last is a cause of the first), we might still worry about the possibility of a *self-undermining series* (where each event is a cause of the next and the last is a *preventer* of the first). The standard example from science fiction literature is the traveler who goes back in time and kills her grandfather before her father is conceived. This kind of sequence would seem impossible, since removing the cause removes the effect, which, in this case, would mean removing the remover herself. *Perhaps* we could make sense of this by using parallel universes or rewritable histories (see above).[16] But Lewis rejects these things and therefore owes us a different explanation of what is going on in this case.

3. The Grandfather Paradox

Lewis's discussion of the grandfather paradox focuses on the following story:[17]

> Consider Tim. He detests his grandfather, whose success in the munitions trade built the family fortune that paid for Tim's time machine. Tim would like nothing so much as to kill Grandfather, but alas he is too late. Grandfather died in his bed in 1957, while Tim was a young boy. But when Tim has built his time machine and traveled to 1920, suddenly he realizes that he is not too late after all. He buys a rifle; he spends long hours in target practice; he shadows Grandfather to learn the route of his daily walk to the munitions works; he rents a room along the route; and there

[15] For more on this point (and causal loops in general), see Wasserman (2018, Chapter 5).

[16] For a fuller account of the possibilities, see Wasserman (2018, Chapter 3).

[17] We will focus on the classic version of the paradox, but Lewis also discussed the "epistemic" version of the puzzle in his early seminars. On the latter kind of argument, see Horwich (1987, Chapter 7), Smith (1997), and Wasserman (2018, Chapter 4, section 3).

> he lurks, one winter day in 1921, rifle loaded, hate in his heart, as Grandfather walks closer, closer… (Ibid., 149)

We may not know how the story turns out, but we do know that Tim will not succeed since Lewis stipulates that Grandfather died in 1957, rather than 1921. Still, it seems as if Tim has *the ability* to kill Grandfather:

> Conditions are perfect in every way: the best rifle money could buy, Grandfather an easy target only twenty yards away, not a breeze, door securely locked against intruders. Tim a good shot to begin with and now at the peak of training, and so on. What's to stop him? The forces of logic will not stay his hand! No powerful chaperone stands by to defend the past from interference… In short, Tim is as much able to kill Grandfather as anyone ever is to kill anyone. (Ibid.)

Yet there are equally good reasons for thinking that Tim *cannot* kill Grandfather:

> No Grandfather, no Father; no Father, no Tim; no Tim, no killing. And for good measure: no Grandfather, no family fortune; no fortune, no time machine; no time machine, no killing. So the supposition that Tim killed Grandfather seems impossible. (Ibid., 152)

In this way, the possibility of time travel seems to yield a contradiction: Tim both can and cannot commit *retrogrampatricide*. For Lewis, this is the central problem posed by the grandfather paradox.

His response to this puzzle goes as follows:

> To say that something can happen means that its happening is compossible with certain facts… Tim's killing Grandfather that day in 1921 is compossible with a fairly rich set of facts: the facts about his rifle, his skill and training… and so on… But his killing Grandfather is not compossible with another, more inclusive set of facts. There is the simple fact that Grandfather was not killed… You can reasonably choose the narrower delineation, and say that he can; or the wider delineation, and say that he can't. But choose. What you mustn't do is waver, say in the same breath that he both can and can't, and then claim that this contradiction proves that time travel is impossible. (Ibid., 150–51)

For Lewis, the proponent of the grandfather paradox is like someone who believes that her friend holds contradictory beliefs because she has heard him describe average Americans as both "rich" and "not rich." However, if the first of these descriptions occurred during a discussion of global poverty and the second happened in a conversation with global billionaires, then this inference would obviously be invalid–the two statements would not express contradictory beliefs since "rich" would mean different things on those occasions. In the same way, Lewis suggests that "can" means one thing when we are considering facts about Tim's skills and training, for example, and something else when we are focused on facts about his personal history (e.g., the fact that Grandfather was not killed in 1921). As a result, there are some contexts in which the sentence "Tim can kill Grandfather" would express a truth and others in which it would express a falsehood. But there is *no* context in which it would express *both* a truth and a falsehood, so there is no contradiction.[18]

Lewis is clearly correct about the context-sensitivity of "can." Indeed, "The Paradoxes of Time Travel"–along with his "General Semantics" (1970) and "Scorekeeping in a Language Game" (1979)–helped launch contemporary research into contextualism.[19] However, the above observations do *not* solve the grandfather paradox since we can state the problem using a single sense of "can."

To begin, consider the notion of ability that is typically at issue when undergoing deliberation. In such contexts, we usually assume that some things are genuine options for us and others are not. Walking across the room is an option for me right now; walking across the ocean is not. Arguably, this distinction is closely linked to our evaluation of agents. For example, if I promised you that I would walk across the room and did not, then you could justifiably hold me responsible for that failure. In contrast,

[18] Similar proposals were made, around the same time, by Fitzgerald (1974, 539–40), Horwich (1975, 435–37), and Thom (1975). However, Lewis's account has proven the most popular, having been endorsed by Dowe (2000, 448–50), Sider (2002), and many others.

[19] This includes contextualism in epistemology, where Lewis's "Elusive Knowledge" (1996) has played a major role. Among other things, that paper develops a response to skeptical arguments which mirrors his solution to the grandfather paradox.

it would make no sense for you to blame me for not walking across the ocean–even if I promised to–since that was never an option for me.

Kadri Vihvelin (1996, 2020) argues that this "agential" sense of "can" is what we often have in mind when discussing abilities. After all, we normally assume that something is an option for us only if we would (or at least might) succeed in performing the action, were we to try. But Tim would fail to kill Grandfather, no matter what he tried, since Grandfather's survival is causally necessary (in those circumstances) for Tim to do anything at all. Hence, Tim is not able to kill Grandfather in the ordinary sense of "able." For this reason, Vihvelin rejects Lewis's claim that "By any ordinary standards of ability, Tim can kill Grandfather" (ibid., 150). At the same time, she *agrees* with Lewis that backward time travel is possible. Her view is simply that time travelers are limited in ways that non-time-travelers are not.

I agree with Vihvelin that Tim cannot kill Grandfather in the ordinary, agential sense. But I still find the case puzzling. To see why, consider the following point from Lewis:

> Suppose that down the street another sniper, Tom, lurks waiting for another victim, Grandfather's partner. Tom is not a time traveler, but otherwise he is just like Tim: same make of rifle, same murderous intent, same everything. We can even suppose that Tom, like Tim, believes himself to be a time traveler. Someone has gone to a lot of trouble to deceive Tom into thinking so. There's no doubt that Tom can kill his victim; and Tim has everything going for him that Tom does. By any ordinary standards of ability, Tim can kill Grandfather. (Ibid., 149)

Lewis's analogy is compelling because Tim and Tom are subject to the exact same constraints–each experiences the same gravitational effect, the same psychological pressures, and so on. Moreover, neither of these agents is bound by shackles, subject to coercive threats, or watched over by chaperones from the future. Of course, there is one important difference between the two since Tim–unlike Tom–is ancestrally related to his target. But it is not as if the identity of the target interferes with Tim's decision-making or binds his arms to the wall. We are therefore left with the question: *what is holding him back?*

In my view, this is one of the most interesting puzzles surrounding the grandfather paradox. Here is one potential solution: the preceding argument assumes that causal constraints like threats and bindings are the only things that could limit an agent's abilities. But that seems like a mistake. I am not able to gain knowledge of a proposition without believing it, for example, but that is not due to causal forces. Rather, part of what it is to know something is to believe it, and believing a proposition is logically incompatible with *not* believing it. Hence, my lack of belief in a given proposition prevents me from knowing it by being constitutively incompatible with such knowledge. Perhaps something similar is true for Tim: part of what it is to have an ability (in the ordinary, agential sense) is for the performance of that action to be compatible with the agent's own causal history.[20] Since Tim's failure to kill Grandfather is part of his causal past, it is part of something constitutively incompatible with his success. That is how it renders him unable to kill Grandfather.[21]

4. Conclusion

In this paper, I have outlined some of the central moves in "The Paradoxes of Time Travel" and reviewed some of what can be said both for and against them. But one of the wonderful things about Lewis's work is that it continues to inspire new ideas about a variety of topics–including ones which lie outside the scope of the original paper. In just the last few years, for example, "Paradoxes" has been cited in works on the Christian doctrine of original guilt, the Aristotelian approach to understanding powers, the Molinist view of divine providence, the traditional debate over freedom and determinism, and various issues relating to ethics and moral responsibility.[22]

[20] See Rea (2015). For clarifications and further refinements, see Law and Wasserman (2020).

[21] For other discussions of Lewis's response to the grandfather paradox, see Wasserman (2018: Chapter 4), Garrett (2019), Effingham (2020: Chapter 10), Fernandes (2020), and Loewenstein (2022).

[22] See, respectively, Effingham (2022), Giannini and Donati (2023), Law (2024), Wasserman (2024), Bernstein (Forthcoming), and Cyr and Tognazzini (2024).

The diversity of these topics is a testament to the wide-ranging influence of Lewis's work, and I fully expect that influence to continue for the next 50 years and beyond.[23]

References

Asimov, Isaac. 1955. *The End of Eternity*. Garden City, New York: Doubleday.

Bernstein, Sarah. (Forthcoming.) "Ethical Puzzles of Time Travel". In *The Routledge Companion to Philosophy of Time*, edited by Nina Emery. New York: Routledge.

Cyr, Taylor, and Neal Tognazzini. 2024. "What Time Travel Teaches Us about Moral Responsibility." *Journal of Ethics and Social Philosophy*, 26: 523–46.

Daniels, David R. 1935. "The Branches of Time." *Wonder Stories*, 7: 294–303.

Deutsch, David. 1997. *The Fabric of Reality: The Science of Parallel Universes–and Its Implications*. New York: Penguin.

Dowe, Phil. 2000. "The Case for Time Travel." *Philosophy*, 75: 441–51.

Effingham, Nikk. 2020. *Time Travel: Probability and Impossibility*. Oxford: Oxford University Press.

Effingham, Nikk. 2022. "Fission Theories of Original Guilt." *International Journal for Philosophy of Religion*, 92: 15–30.

Fernandes, Alison. 2020. "Freedom, Self-Prediction, and the Possibility of Time Travel." *Philosophical Studies*, 177: 89–108.

Fitzgerald, Paul. 1974. "On Retrocausality." *Philosophia*, 4: 513–51.

Garrett, Brian. 2019. "Bulletproof Grandfathers, David Lewis, and 'Can't'-Judgments.'" *Acta Analytica*, 34: 177–80.

Giannini, Giacomo, and Donatella Donati. 2023. "Dispositionalism's (Grand)daddy Issues: Time Travelling and Perfect Masks." *Analysis*, 83: 40–49.

Goddu, G.C. 2003. "Time Travel and Changing the Past: (Or How to Kill Yourself and Live to Tell the Tale)." *Ratio*, 16: 16–32.

Hanley, Richard. (2004). "No End in Sight: Causal Loops in Philosophy, Physics and Fiction." *Synthese*, 141: 123–52.

Horwich, Paul. 1975. "On Some Alleged Paradoxes of Time Travel." *Journal of Philosophy*, 72: 432–44.

Horwich, Paul. 1987. *Asymmetries in Time: Problems in the Philosophy of Science*. Cambridge, Massachusetts: MIT Press.

Law, Andrew. 2018. "The Puzzle of Hyper-change." *Ratio*, 32: 1–11.

[23] I thank Brian Garrett for comments on earlier versions of this paper.

Law, Andrew. 2024. "If Molinism is True, What can you Do?" *International Journal for Philosophy of Religion*, 95: 307–22.

Law, Andrew, and Ryan Wasserman. 2022. "Lessons from Grandfather." *Philosophies*, 7: 1–10.

Lewis, David K. 1970. "General Semantics." *Synthese*, 22: 18–67.

Lewis, David K. 1973. "Causation." *Journal of Philosophy*, 70: 556–67.

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

Lewis, David K. 1979. "Scorekeeping in a Language Game." *Journal of Philosophical Logic*, 8: 339–59.

Lewis, David K. 1983. Survival and Identity. Reprinted with postscripts in *Philosophical Papers: Volume I*. Oxford: Oxford University Press.

Lewis, David K. 1986a. *On the Plurality of Worlds*. Oxford: Oxford University Press

Lewis, David K. 1986b. Counterfactual Dependence and Time's Arrow. Reprinted with postscripts in *Philosophical Papers: Volume II*. Oxford: Oxford University Press.

Lewis, David K. 1996. "Elusive Knowledge." *Australasian Journal of Philosophy*, 74: 549–67.

Lewis, David K. 2023. *Philosophical Manuscripts*, edited by Frederique Janssen-Lauret and Fraser MacBride. Oxford: Oxford University Press.

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

Meiland, J. W. 1974. "A Two-Dimensional Passage Model of Time for Time Travel." *Philosophical Studies*, 26: 153–73.

Rea, Michael. 2015. "Time Travelers Are Not Free." *Journal of Philosophy*, 112: 266–79.

Sider, Theodore. 2002. "Time Travel, Coincidences and Counterfactuals." *Philosophical Studies*, 110: 115–38.

Smith, Nicholas J. J. 1997. "Bananas Enough for Time Travel?" *British Journal for the Philosophy of Science*, 48: 363–89.

Thom, Paul. 1975. "Time Travel and Non-Fatal Suicide." *Philosophical Studies*, 27: 211–16.

van Inwagen, Peter. 2010. "Changing the Past". In *Oxford Studies in Metaphysics, Volume 5*, edited by Dean Zimmerman. Oxford: Oxford University Press.

Vihvelin, Kadri. 1996. "What Time Travelers Cannot Do." *Philosophical Studies*, 81: 315–30.

Vihvelin, Kadri. 2020. "Killing Time Again." *The Monist*, 103: 312–27.

Wasserman, Ryan. 2015. "Lewis on Backward Causation." *Thought*, 4: 141–50.

Wasserman, Ryan. 2018. *Paradoxes of Time Travel*. Oxford: Oxford University Press.

Wasserman, Ryan. 2024. "Time Travel, Freedom, and Incompatibilism." *Erkenntnis*, 89: 2953–66.

Organon F 33 (2) 2026: 182–195 ISSN 2585-7150 (online)
https://doi.org/10.31577/orgf.2026.33204

RESEARCH ARTICLE

On Lewis's Way Out

Nicholas J.J. Smith*

Received: 17 September 2025 / Accepted: 29 December 2025

Abstract: On the fiftieth anniversary of the publication of Lewis 1976, this paper reexamines 'Lewis's solution to the grandfather paradox': both the philosophical move often designated using this phrase and the accuracy of this manner of designating that move.

Keywords: Time travel; Grandfather Paradox; changing the past.

1. The Grandfather Paradox(es)

We need to distinguish two different arguments against the possibility of backwards time travel. Both of them have been referred to in the literature as the "grandfather paradox"—and also by other names.

The first argument goes back at least as far as Gödel 1949:

> by making a round trip on a rocket ship in a sufficiently wide curve, it is possible in these worlds to travel into any region of the past, present, and future, and back again, exactly as it is possible in other worlds to travel to distant parts of space. This state of affairs seems to imply an absurdity. For it enables one e.g., to travel into the near past of those places where he has himself lived. There he would find a person who would be himself

* The University of Sydney
https://orcid.org/0000-0003-2632-1032
The University of Sydney, Department of Philosophy, Main Quadrangle A14, The University of Sydney, NSW 2006, Australia
nicholas.smith@sydney.edu.au

> at some earlier period of his life. Now he could do something to this person which, by his memory, he knows has not happened to him. This and similar contradictions, however, in order to prove the impossibility of the worlds under consideration, presuppose the actual feasibility of the journey into one's own past. (560–61)

The argument is a *reductio* of the possibility of backwards time travel:

> Backwards time travel generates contradictions: if backwards time travel were to occur then contradictions would be made true. It is impossible for a contradiction to be true.
>
> Therefore backwards time travel is impossible.

In Gödel's case, the contradiction is 'P and not P' where P says that a certain thing happens to the younger self of the time traveller: something that the older time traveller knows did not happen to his younger self (hence the 'not P' conjunct)—and yet goes ahead and makes happen (hence the 'P' conjunct). For example, P might be: the younger man is given a gold watch—or pinched on the arm—by an older man who looks a lot like him.

P might also be something more extreme: the younger man is murdered by the older man. Thus, Horwich 1987:

> Many opponents of time travel would base their position on a well-known variant of Gödel's argument. They suppose that if time travel were possible, then people would be able to return to the past and murder their infant selves. But this form of suicide is impossible (for only those who fail will ever be in a position even to make the attempt). So it follows that time travel is impossible. (117)

Alternatively, the time traveller might generate a contradiction by doing something (that did not occur) not to her younger self but to an ancestor —for example, her grandfather:

> Another set of objections consists in attempting to show that the logical possibility of time travel unavoidably generates contradictions. One much-used example is this: if you could time-travel, there would be nothing to stop you going back in time and killing

> your paternal grandfather before your father was conceived. But, of course, if your grandfather had died before your father was conceived, your father wouldn't have existed, and neither would you, so you wouldn't be there to kill your grandfather. (Macbeath 1982, 411)

—or her great-grandfather (Stein 1970, 591, n.2) or mother (Dwyer 1975, 348).

This argument against time travel goes by many names. In reference to the second variant above, the terms 'autofanticide' (Horwich 1987), 'auto-infanticide' (Vihvelin 1996, Smith 1997), and 'retrosuicide' (Vranas 2009) have been used. In reference to the third variant, the term 'grandfather paradox' has been used (Deutsch and Lockwood 1994, Smith 1997, Smith 2017, Smith 2024). Riggs 1997 calls it the 'principal paradox' of time travel. I shall henceforth refer to it as the CTP argument or CTP reductio. The initials stand for 'Changing The Past'. The idea behind the name is that the idea behind the argument is that time travel leads inevitably to time travellers changing the past (doing things that did not in fact occur)—and this involves contradictions (the thing that the time traveller does both did and did not occur).

Note that the CTP argument does not turn on issues of *free will* or the *abilities* of the time traveller. To emphasise this point, Earman 1972 presents a version that does not involve human time travellers at all:

> The paradoxes involving closed timelike curves are often stated in terms of agency, e.g., we are invited to consider what would happen if an observer could travel into his own past and shoot himself-at-an-earlier-time. But it is clear that the paradoxes do not hinge on human agency. Thus, consider a rocket ship which at some space-time point x can fire a probe which will travel into the past lobe of the null cone at x. Suppose that the rocket is programmed to fire the probe unless a safety switch is on and that the safety switch is turned on if and only if the 'return' of the probe is detected by a sensing device with which the rocket is equipped. Is the probe fired? We find that the answer is that it is fired if and only if it is not fired, which is a contradiction if standard logic holds. (231–32)

The second argument, by contrast, does turn on issues of free will or the abilities of the time traveller. This argument goes back at least as far as Lewis 1971:

> To take a traditional example, called the *Grandfather Paradox*: Oscar has taken a dislike to his long-dead grandfather. He decides to enter his time machine, go back to the time of his grandfather's childhood, and kill his grandfather as a young boy. He goes back. Here's Oscar in the empty house across the street; here's Grandfather happily playing in the sandbox; Oscar's rifle is loaded and aimed; Oscar is an excellent shot and the distance is short; there's nothing and no one to stop Oscar from shooting; Oscar's heart is filled with hate—and (1) *Oscar cannot kill his grandfather*. For if he did, there'd be no Oscar's father, and no time traveling Oscar to come back and kill his grandfather. But (2) *Oscar can kill his grandfather*. For what's to stop him? He has all he needs: a good rifle, a good view of his target, determination to shoot, a strong and steady trigger finger, no one else around to stop him shooting. *Contradiction*. (39–40)

Note the difference from the CTP argument. In the CTP argument, a contradiction ensues only if and when the time traveller does something that did not happen (but the thought is that time travellers would do such things—so time travel does lead to contradictions and hence is impossible). In Lewis's argument, by contrast, there is a contradiction even if the time traveller does nothing that did not occur. The contradictory claims here do not concern what the time traveller effects or the objects of his actions (for example: the younger self is given a gold watch, and not given a gold watch). They concern the time traveller himself—his abilities: the time traveller both *can* make it the case that (say) his grandfather was murdered and *cannot* make it the case that his grandfather was murdered. Let the contradiction in the CTP argument be 'P and not P'. P describes something that did not in fact happen (e.g. the time traveller's younger self was pinched on the arm) but which the time traveller then proceeds to make happen (thus making a contradiction true). In the new argument, the contradiction is 'the time traveller can make it the case that P and the time traveller cannot make it the case that P', and

this contradiction is true *even if* the time traveller does *not* proceed to make it the case that P.

This second argument has also been referred to as the 'grandfather paradox': see for example the quote from Lewis above, Sider 1997 and Smith 2024. It has also been referred to by other names, for example the 'can and cannot' problem (Smith 2017, Smith 2024). Following the latter usage, I shall henceforth refer to it as the CCN argument—where the initials stand for 'Can CanNot'.

2. Lewis's Solution(s)

So the term 'grandfather paradox' has two different meanings: it can denote the CTP argument or the CCN argument. Likewise, the term 'Lewis's solution to the grandfather paradox' has two different meanings.

The first meaning is straightforward. Lewis 1976 explicitly presents a solution to the CCN problem:

> We have this seeming contradiction: "*Tim doesn't, but can, because he has what it takes*" versus "*Tim doesn't, and can't, because it's logically impossible to change the past.*" I reply that there is no contradiction. Both conclusions are true, and for the reasons given. They are compatible because "can" is equivocal. To say that something can happen means that its happening is compossible with certain facts. *Which* facts? That is determined, but sometimes not determined well enough, by context.…He can and he can't, but under different delineations of the relevant facts. You can reasonably choose the narrower delineation, and say that he can; or the wider delineation, and say that he can't. But choose. What you mustn't do is waver, say in the same breath that he both can and can't, and then claim that this contradiction proves that time travel is impossible. (150–51)

This solution has been widely accepted (although not universally—see, for example, Vranas 2009) and has been referred to as Lewis's solution to the grandfather paradox, for example, by Sider (1997, 143).

The second meaning is more vexed because while there is indeed a solution to the CTP argument that turns on words written by Lewis, Lewis was not—or so I shall argue—presenting this solution when he wrote those words. Indeed, he was not talking about the CTP argument at all—he was talking about the CCN argument—and he was not presenting a *solution* to that problem when he wrote those words: he was presenting and motivating the *problem* itself.

Here is what Lewis 1976 wrote:

> You know, of course, roughly how the story of Tim must go on if it is to be consistent: he somehow fails. Since Tim didn't kill Grandfather in the "original" 1921, consistency demands that neither does he kill Grandfather in the "new" 1921. Why not? For some commonplace reason. Perhaps some noise distracts him at the last moment, perhaps he misses despite all his target practice, perhaps his nerve fails, perhaps he even feels a pang of unaccustomed mercy. (150)

Note in particular the passage "Why not? For some commonplace reason." This can indeed be read as a solution to the CTP reductio—but presenting that solution to that problem is not what Lewis was doing in his paper. Let me address these points in turn.

First, the CTP reductio turns on a tight link between time travel and contradictions: if time travel were to occur, contradictions would be made true. If this link is severed, the reductio is blocked. One way of severing it is to suppose that time travellers are accompanied by chaperones or time lords who prevent them from changing the past. Another way is to posit 'forces of logic' that—like physical force fields—prevent time travellers from doing things that did not happen. But both of these devices seem extremely far-fetched, and in any case, still allow the reductio to block all cases of time travel that do *not* involve such chaperones or forces. Lewis's words describe a better way of blocking the reductio: what stops time travellers changing the past are not exotic things such as time lords or logic forces, but everyday, commonplace occurrences. Contra the CTP reductio, it is *not* the case (the thought goes) that if backwards time travel were to occur then contradictions would be made true: commonplace occurrences would prevent time travellers from doing things that did not happen, and thus

prevent contradictions. Lewis has indeed been seen as proposing this solution to the CTP argument—for example, by the following authors:

> In standard discussions of x-old's encounter with x-young (Lewis, Dwyer, Thom) x-old must fail (for example) to kill x-young, but there need be no particular reason for his failure. Something distracts him, he loses interest, the trigger is somehow not pulled. Which accident fends the contradiction off matters not at all. (Nerlich 1981, 237)[1]

> David Lewis presented an ingenious argument which is considered by many philosophers as providing the solution to the Principal Paradox. Lewis imagines a time traveller, named Tim, who attempts to shoot (and thereby kill) his grandfather at a time prior to the biological conception of Tim's own father. Tim wants very much to kill his grandfather but somehow he fails the attempt. Although Lewis's argument has great merit, it also has serious shortcomings! (Riggs 1997, 50–51)

> David Lewis (1976) showed that the auto-infanticide objection cannot establish the *impossibility* of backward time travel....Some science fiction writers respond to the auto-infanticide objection by saying that backward time travel *is* possible, as long as time travellers are accompanied by chaperones who prevent them from changing the past. Such chaperones are *ad hoc* and unappealing additions to time travel scenarios, however—and also unnecessary. David Lewis argues that no strange devices are required to stop the time traveller killing his younger self; rather, the time traveller fails for some commonplace reason. 'Perhaps some noise distracts him at the last moment, perhaps he misses despite all his target practice, perhaps his nerve fails, perhaps he even feels a pang of unaccustomed mercy' (Lewis 1976, 150). Perhaps his gun jams; perhaps he slips on a banana peel; perhaps he has a cardiac arrest.

[1] Note that Nerlich attributes this solution to the CTP argument *jointly* to Lewis 1976, Dwyer 1978 and Thom 1975. In this connection note also Dwyer 1975 349: "There may be countless reasons why the assassination attempt fails but these reasons have nothing to do with guns not behaving as normal physical objects or with voluntary action not applying in the usual way."

> Nothing more than such ordinary occurrences is required to stop the time traveller killing his younger self. Hence backward time travel does not imply the truth of contradictions, even in the absence of chaperones. Hence backward time travel is *not* impossible. (Smith 1997, 364–66)

> What "logic-bouncers" prevent time travelers from changing the past? What velvet rope of consistency bars them from doing what they seem perfectly able to do? Many have accepted David Lewis's (1976) deflationary answer: nothing stops time travelers from committing paradoxical deeds (from, say, killing their grandfathers). Or as Sider (2002, 116) puts it, *no one thing stops them:* they fall victim to circumstance; they fail for "ordinary reasons". (Slater 2005, 363)

> In order to defend the possibility of time travel in the face of this argument, we need to show that time travel is not a sure route to doing the impossible. So, given that a time traveler has gone to the past and is facing Grandfather, what could stop him from killing Grandfather? Some science fiction authors resort to the idea of chaperones or time guardians who prevent time travelers from changing the past—or to mysterious forces of logic. But it is hard to take these ideas seriously—and more importantly, it is hard to make them work in detail when we remember that changing the past is impossible. Fortunately, there is a better response —also to be found in the science fiction literature, and introduced into the philosophical literature by Lewis (1976). What would stop the time traveler from doing the impossible? She would fail "for some commonplace reason," as Lewis (1976, p. 150) puts it. Her gun might jam, a noise might distract her, she might slip on a banana peel, and so on. Nothing more than such ordinary occurrences is required to stop the time traveler from killing Grandfather. Hence, backwards time travel does not entail the occurrence of impossible events—and so the above objection is defused. (Smith 2017, 154–55)

Second, however, Lewis does not discuss the CTP argument at all in his 1976 paper. Rather, he presents the CCN argument. This argument centres on two contradictory claims: Tim can kill Grandfather; and Tim cannot kill

Grandfather. In order for the argument to have force, *both* claims must be plausible. Lewis's remarks about commonplace occurrences occur in the context of motivating the first claim: Tim can kill Grandfather. Note that if what stopped Tim were forces of logic, or chaperones, we might well be inclined to deny that, nevertheless, Tim *can* kill Grandfather. However, if what stops Tim is simply a commonplace mishap, then this claim still seems highly plausible. This point is more clear and explicit in Lewis 1971 (the posthumously published longer work from which Lewis 1976 was drawn):[2]

> (2) *Oscar can kill his grandfather*.…the reasons for (2) are also convincing! What's to stop Oscar? If Oscar were not a time traveler, there would be no case at all for saying he could not do the killing. If, a block down the road, another sniper is aiming at another child, and the situation is a perfect duplicate of the first except that in that situation the sniper is not a time traveler descended from the child (but imagine that he has been deceived into thinking that he is, so that his mental state is exactly like Oscar's) then we'd certainly think that the other sniper could kill the child he's aiming at. We'd have as much reason to say that as we *ever* have to say that anyone can do anything. And what relevant difference is there between the abilities of that other sniper and the abilities of Oscar? We could, of course, imagine that Oscar is attended by guardian spirits (or Forces) that somehow prevent him from doing anything wrong: if he started to pull the trigger he would find his finger paralyzed, or the bullet would be deflected, or he would undergo a mysterious change of heart. Then Oscar would indeed not be like the other sniper, and would not possess the abilities he seems to possess; and (2) would be straightforwardly false. (40)

[2] A footnote at the end of Lewis 1976 describes it thus: "The present paper summarizes a series of lectures of the same title, given as the Gavin David Young Lectures in Philosophy at the University of Adelaide in July, 1971." Lewis 1971 is described as follows by Janssen-Lauret and MacBride 2023 x: "The first manuscript consists of his Gavin David Young Lectures, entitled *The Paradoxes of Time Travel*. The lectures were delivered in July 1971 at the University of Adelaide."

Thus, Lewis's 'commonplace occurrences' remarks are not presented by him as a solution to the CTP argument. First, he was not discussing the CTP argument at all: he was discussing the CCN argument. Second, the remarks are not part of his solution: they occur in the context of motivating (not solving) the CCN argument by arguing for one of its two contradictory conjuncts.

The time travel literature is full of stories about objects or information 'coming from nowhere'—for example:

> Suppose I steal a time machine from my local museum and use it to travel back in time. When I arrive at my destination, I donate the machine to the local museum—so it turns out that the machine I stole was in the museum only because I put it there. Where does the machine come from? Or consider the Borges story about the time traveller who takes back copies of works by a famous artist to discuss with the artist himself. The time traveller finds on his arrival a hopeless artist, who proceeds to become famous by copying the paintings given to him by the time traveller. Which paintings are copies, which originals? Where do the ideas for the paintings come from? (Smith 1997, 371, n.11)

Ironically, it seems that 'Lewis's solution to the grandfather paradox'—in the sense of the 'commonplace occurrences' solution to the CTP reductio—is another example of an idea that came from nowhere. Those who attribute this solution to Lewis see it in his words—they did not make it up themselves—and yet it is clear from the preceding discussion that it is not what Lewis himself had in mind when he wrote those words: he was talking about the CCN argument (and motivating, not solving, it).

3. A Better Solution to the CTP Argument

My point has not been that the commonplace occurrences solution to the CTP argument is a *chimera*. It is indeed a genuine solution to the problem. My point is rather that its *origins* are unusual in that it was not presented as such by the person to whom it is attributed—yet nor was it thought up by someone else.

In this section I turn from the question of the origins of this solution to the question of its merits. It has achieved the status of the 'gold standard' or 'received' solution to the CTP argument. In this section I'll argue that there is however a much better solution to that problem.

One problem with the commonplace occurrences solution to the CTP argument (henceforth the COS) is that it engenders a new argument against backwards time travel: one according to which backwards time travel is not *impossible* but extremely *improbable* (see Horwich 1987). This is not a knock down objection to the COS, because—as I argue in Smith 1997—Horwich's further problem is soluble.

There is however a deeper conceptual point to be made. The CTP reductio turns on the idea that backwards time travel leads to contradictions. A defence of the possibility of backwards time travel against this argument denies that if time travel were possible, contradictions would be engendered. At this point there is a natural tendency to ask any defender of time travel: *what stops* the time traveller doing something that did not happen (thereby making a contradiction true)? It seems, after all, to require so little. The time traveller need not do anything as terrible as murder anyone—or even find her younger self and shake her hand. She can do *anything* that did not happen: for example, step on a fallen leaf (that was not on that day bruised by any footstep). The COS takes this question seriously and offers an answer: commonplace occurrences! As Sider 2002 puts it:

> We have admitted the possibility of time travel, though not the possibility of autoinfanticide. But these possible time travelers who do not kill their earlier selves: some have the desire as well as the means. What stops them? No one thing. Some have a sudden change of heart. Some fear awful forces they think would be unleashed by a violation of the laws of logic. Some attempt the deed but fail for various reasons: non-lethal wounds, slips on banana peels, and the like. (116)

Returning to my example of the fallen leaf: the time traveller slips when she tries to step on it; or a bird flies by and shrieks, distracting her; or a gust of wind moves the leaf out of her way; and so on. But for someone who took the original question seriously, this answer is liable to seem unsatisfying. Hence Horwich's objection, and related objections about the *inexplicability*—as

opposed to the *impossibility* (CTP reductio) or *improbability* (Horwich's argument)—of time travel scenarios (see Smith 2017 and Smith 2024).

A better response is to point out that the question itself is mistaken.[3] The question should not be answered at all: it should be dissolved—rather than solved—by pointing out that it rests on a confusion. The question 'What stops X occurring?' is sensible only when 'X' is a coherent description of a scenario. Contrast the case where you ask me why I did not *put my shoes in the cupboard and not on the floor*, with the case where you ask me why I did not *put my shoes in the cupboard and not in the cupboard*. In the former case, you give a coherent description of a possible scenario, and so it makes sense for me to say what stopped me from actualising this scenario: for example, I was too tired to be bothered putting my shoes away properly, so I just left them on the floor. In the latter case, there just *is no* coherent description of a possible scenario in play: *no* scenario can be correctly described as one in which my shoes are in the cupboard and not in the cupboard. So I should not proceed to answer the question: I was too tired; I slipped on a banana peel; I got distracted. I should point out to *you* that your question does not make sense. You are asking me why I did not actualise a scenario, your very description of which is incoherent. There *is no* possible scenario corresponding to your description, and so the question why I did not actualise 'it' is a mistaken one: rather than answer it, the clearest thing to do is to point out the confusion behind the question.

Exactly the same can and should be said about the question: what stops the time traveller crushing the leaf that was not in fact crushed (or changing the past in any other way)? The *description* of the scenario involves a contradiction: the leaf is uncrushed and also crushed by the time traveller (at the very same time, in the very same place). Thus *no* possible scenario satisfies the description. Hence, there is no legitimate question why such a scenario is not actualised: there is no there there to actualise. *Whatever* might happen will not be 'that' and hence there is no sensible question as to what *stops* 'that' from occurring. Not only do we not need an *exotic* explanation—time lords or logic forces—we do not need *any* explanation. The question as to what stops the time traveller actualising 'that' scenario is no better than the question of why he does not put his shoes in the

[3] Here I sketch this response; for a complete presentation, see Smith 2017.

cupboard and not in the cupboard. We should not answer it at all: not by appealing to chaperones or forces of logic, or commonplace occurrences. We should reject the question and say to the questioner: before you ask me why X does not occur, make sure that your description 'X' is coherent. For if your description involves an internal contradiction, it describes *no* scenario at all; and if no possible scenario satisfies your incoherent description, then there just *is no* 'that' to ask about—and hence no question what stops 'that' occurring.

4. Conclusion

It is widely held that the gold standard solution to the CTP argument against backwards time travel is 'Lewis's solution.' I have argued that this solution to this problem should not be regarded as having been presented by Lewis 1976 in his remarks about 'some commonplace reason,' because when he made those remarks, he was not in the business of solving the CTP argument—he was setting up (not solving) a different argument against backwards time travel; and should not be regarded as the gold standard, because a better solution is available.

References

Deutsch, David, and Michael Lockwood. 1994. "The Quantum Physics of Time Travel." *Scientific American*, 270(3): 50–56.

Dwyer, Larry. 1975. "Time Travel and Changing the Past." *Philosophical Studies*, 27: 341–50.

Dwyer, Larry. 1978. "Time Travel and Some Alleged Logical Asymmetries between Past and Future." *Canadian Journal of Philosophy*, 8: 15–38.

Earman, John. 1972. "Implications of Causal Propagation outside the Null Cone." *Australasian Journal of Philosophy*, 50: 222–37.

Gödel, Kurt. 1949. A Remark about the Relationship between Relativity Theory and Idealistic Philosophy. In *Albert Einstein: Philosopher-Scientist*, edited by Paul Arthur Schilpp, 555–62. La Salle, Illionois: Open Court (*Library of Living Philosophers*, vol. VII).

Horwich, Paul. 1987. *Asymmetries in Time: Problems in the Philosophy of Science*. Cambridge, Massachusetts: MIT Press.

Janssen-Lauret, Frederique and Fraser MacBride. 2023. Editors' Introduction. In Lewis, David. 2023. *Philosophical Manuscripts,* Oxford: Oxford University Press ix–xvii.

Lewis, David K. 1971. The Paradoxes of Time Travel: The Gavin David Young Lectures at the University of Adelaide. In Lewis, David. 2023. *Philosophical Manuscripts,* Oxford: Oxford University Press 3–56.

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

Lewis, David K. 2023. *Philosophical Manuscripts*. Oxford: Oxford University Press.

Macbeath, Murray. 1982. "Who Was Dr Who's Father?" *Synthese*, 51: 397–430.

Nerlich, Graham. 1981. "Can Time Be Finite?" *Pacific Philosophical Quarterly*, 62: 227–39.

Riggs, Peter J. 1997. "The Principal Paradox of Time Travel." *Ratio*, 10: 48–64.

Sider, Theodore. 1997. "A New Grandfather Paradox?" *Philosophy and Phenomenological Research*, 57(1): 139–44.

Sider, Theodore. 2002. "Time Travel, Coincidences and Counterfactuals." *Philosophical Studies*, 110: 115–38.

Slater, Matthew H. 2005. "The Necessity of Time Travel" (On Pain of Indeterminacy). *Monist* (Special Issue on Time Travel), 88: 362–69.

Smith, Nicholas J.J. 1997. "Bananas Enough for Time Travel?" *British Journal for the Philosophy of Science*, 48: 363–89.

Smith, Nicholas J.J. 2017. "I'd Do Anything to Change the Past (but I Can't Do 'That')." *American Philosophical Quarterly*, 54(2): 153–68.

Smith, Nicholas J.J. 2024. Time Travel. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University, Summer 2024 edition.

Stein, Howard. 1970. "On the Paradoxical Time-Structures of Gödel." *Philosophy of Science*, 37(4): 589–601.

Thom, Paul. 1975. "Time Travel and Non-Fatal Suicide." *Philosophical Studies*, 27: 211–16.

Vihvelin, Kadri. 1996. "What Time Travelers Cannot Do." *Philosophical Studies*, 81: 315–30.

Vranas, Peter B.M. 2009. "Can I Kill My Younger Self? Time Travel and the Retrosuicide Paradox." *Pacific Philosophical Quarterly*, 90: 520–34.

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.

Organon F 33 (2) 2026: 218–237 ISSN 2585-7150 (online)
https://doi.org/10.31577/orgf.2026.33206

RESEARCH ARTICLE

Lewis and the Price of Time Travel: Lessons from *Großvater*

Huw Price*

Received: 13 October 2025 / Accepted: 29 March 2026

Abstract: Responding to Lewis's (1976) defense of the consistency of time travel (TT), Horwich (1987) and Price (1996) claim that TT may nevertheless be shown to be improbable, due to its need for unlikely coincidences. Smith (1997, 2024) and Ismael (2003) reply, correctly, that this begs the question against TT. Where does this leave us, and TT itself? To put the issue in a broader frame, I note (i) a Lewis-inspired "defense" of Aristotelian mechanics against a famous argument by Galileo; and (ii) the relevance of the Duhem-Quine Thesis. With a range of comparison cases thus in view, TT may be assessed by the Quinean pragmatic standards for theory choice that Lewis elsewhere endorses. I conclude that it is not unreasonable to conclude that TT is highly implausible, despite the fallacy of a direct appeal to the unlikelihood of coincidences. By Lewis's own standards, treating Quinean pragmatic virtue as a guide to truth, this amounts to an argument for the epistemic improbability of TT.

Keywords: David Lewis; time travel; causal loops; bilking.

* Trinity College, University of Cambridge
https://orcid.org/0000-0002-9091-760X
Trinity College, University of Cambridge, CB2 1 TQ, UK
hp331@srcf.net

1. Death of the Author

Let's begin with autoinfanticide – *auto*autoinfanticide, in fact, from my point of view, because I don't mean the usual impersonal example. I want to describe a case in which I myself was perpetrator and victim. Happily for both of us, not to mention the present text, the deed was dialectical. Nodding off at the wheel, my young self collided with his own epistemic guardrails, without noticing that he was doing so.

To set the scene, imagine the closing years of the twentieth century, deep in the pages of *Time's Arrow and Archimedes' Point* (Price 1996, hereafter TAAP). In Chapter 7, discussing the so-called bilking argument against backward causation, our young author mentions the parallel with time travel, and Lewis's famous defense.

> [I]t is clear that even given the hypothesis of time travel, we are never actually justified in expecting the experiment to yield contradictory results, for logic alone rules that out. A number of authors have made this the basis of a defense of the possibility of time travel against the bilking argument. See Horwich (1975), Lewis (1976) and Thom (1974), for example. This issue is not directly relevant to our present concerns, which exploit a much larger loophole in the bilking argument. In passing, however, let me record my view – similar to that of Horwich (1987), ch. 7 – that the bilking argument survives the former challenge. Roughly speaking, it shows us that the hypothesis of time travel can be made to imply propositions of arbitrarily low probability. This is not a classical reductio, but it is as close as science ever gets. (TAAP, 278; endnote to p. 171)

This endorsement of Horwich is noted both by Nicholas Smith (1997) and Jenann Ismael (2003), in papers defending Lewis's argument. Both Smith and Ismael also pointed out the mistake to the author in person – probably more than once.[1]

[1] On the most recent occasion, Ismael also pointed out the analogy on which I'm relying below: the similarity between her objections to Horwich and myself, on one hand, and points I make in TAAP, on the other.

To put the point in my terms, I had hit one of the very fallacies that TAAP was trying to expose. For example, TAAP discusses the suggestion that entropy will *decrease* in the future, if the universe eventually recollapses under its own gravity. There are several ways in which this proposal emerges in discussions of links between cosmology and thermodynamics. Most directly, some writers argued that the familiar increase in entropy is *explained by* the fact that the universe is expanding (e.g., Gold 1962; see TAAP, Ch. 4, for discussion).

From this claim – call it the *Gold Hypothesis* – the proposal about a future reduction in entropy is an easy consequence. For suppose that the universe does recollapse in what we think of as the distant future. In that era, the universe may equally be said to be *expanding,* if we simply reverse the sign on the temporal axis. The physics of gravitation doesn't have a preference between the two labellings. Unless we slip in some time-asymmetric principle to break the symmetry, this means that the Gold Hypothesis will imply that entropy *increases* in that region, in the reversed labelling sense. In other words – reverting to our familiar labelling – it implies that entropy *decreases* in that future recollapsing era.

A common response to this argument is that such an entropy decrease would be extraordinarily unlikely, by the usual statistical arguments. But this begs the question against the proposal. The probabilities are time-symmetric, as Boltzmann famously learned from Loschmidt. Towards what we call the past, it is commonly accepted that something overrides the statistics – the so-called Past Hypothesis (Albert 2000). My point in TAAP was that we're guilty of a double standard if we don't allow that the same might be true towards the future. The statistics only show that entropy increases *so long as it is not constrained to decrease* (as it seems to be towards the past).

This is precisely the kind of point that Smith and Ismael make against my remark in support of Horwich. In his more recent SEP piece, Smith puts it like this.

> We can set out Horwich's argument this way:
>
> 1. If time travel were ever to occur, we should see extensive uncaused correlations.

2. It is extremely unlikely that we should ever see extensive uncaused correlations.
3. Therefore time travel is extremely unlikely to occur.

> The conclusion is not that time travel is impossible, but that we should treat it the way we treat the possibility of, say, tossing a fair coin and getting heads one thousand times in a row. As Price (1996, 278 n.7) puts it—in the context of endorsing Horwich's conclusion: "the hypothesis of time travel can be made to imply propositions of arbitrarily low probability. This is not a classical reductio, but it is as close as science ever gets." (Smith 2024, §2.2)

Smith then describes his own argument from (Smith 1997).

> Against the second premise, [Smith 1997] argues that, from the fact that we have never seen extensive uncaused correlations, it does not follow that we never shall. This is not inductive scepticism: let us assume (contra the inductive sceptic) that in the absence of any specific reason for thinking things should be different in the future, we are entitled to assume they will continue being the same; still we cannot dismiss a *specific reason* for thinking the future will be a certain way *simply* on the basis that things have never been that way in the past. You might reassure an anxious friend that the sun will certainly rise tomorrow because it always has in the past—but you cannot similarly refute an astronomer who claims to have discovered a specific reason for thinking that the earth will stop rotating overnight. (Smith 2024, §2.2)

2. Reversing in His Grave?

To tie this to my discussion in TAAP, consider these remarks by the famous Cambridge astronomer, Sir Arthur Eddington (1882 – 1944).

> If someone points out to you that your pet theory of the universe is in disagreement with Maxwell's equations—then so much the worse for Maxwell's equations....But if your theory is found to be

> against the second law of thermodynamics I can give you no hope; there is nothing for it but to collapse in deepest humiliation. This exaltation of the second law is not unreasonable. There are other laws which we have strong reason to believe in, and we feel that a hypothesis which violates them is highly improbable; but the improbability is vague and does not confront us as a paralysing array of figures, whereas the chance against a breach of the second law … can be stated in figures which are overwhelming. (Eddington 1928, 74 – 75)

Eddington is well aware of the puzzle of the time-asymmetry of the second law, and of the origin of low entropy in our universe. Concerning the latter, he says that it requires us 'to admit anti-chance; and apparently the best thing we can do with it is to sweep it up into a heap at the beginning of time' (1931, 452). Eddington does not regard this as a novel suggestion. On the contrary, he regards it as implicit in physics since the mid-nineteenth century.

> There is no doubt that the scheme of physics as it has stood for the last three-quarters of a century postulates a date at which either the entities of the universe were created in a state of high organisation, or pre-existing entities were endowed with that organisation which they have been squandering ever since. Moreover, this organisation is admittedly the antithesis of chance. It is something which could not occur fortuitously (1928, 84).

In other words, Eddington sees the need for a Past Hypothesis, overriding the usual statistical probabilities in inferences towards the past – hence his term 'anti-chance'. So far as I know, he never considers the possibility that there might be anti-chance at some point in the future, too. However, his mocking dismissal of any 'theory … found to be against the second law of thermodynamics' shows what he would think of the suggestion. But here Eddington falls into the fallacy that I described in TAAP, and that Smith and Ismael rightly identified in my expression of support for Horwich. To adapt Smith's argument above, 'you cannot [appeal to the usual statistical arguments to] refute an astronomer who claims to have discovered a specific reason for thinking that' because the (future) end of the universe is *also* subject to 'anti-chance', entropy will eventually decrease.

3. Get out of Jail Free?

This is a strong point, but it may seem too strong. What stops it becoming an all-purpose Get Out of Improbability Free card? After all, the application to the case of a potential Future Hypothesis requires us to stare down a truly phenomenal level of improbability. To give a ball-park figure, Roger Penrose (1989, 444) calculates the improbability of the low entropy initial condition as about 1 in 10 to the power 10^{123}. Symmetry suggests that a low entropy final condition would be similarly 'special'. Bananas enough for time travel (Smith 1997) seems an easy trick, in comparison. If the defense works in this cosmological case, what's to stop it working anywhere at all?

Pursuing this thought, we might note a structural similarity to an argument that Lewis himself mentions, tongue somewhat in cheek, in discussing Putnam's challenge to metaphysical realism. This 'malicious argument', as Lewis calls it (1984, 221), has since made a name for itself as Goodman's Proof that P.[2]

> Zabludowski has insinuated that my thesis that *p* is false, on the basis of alleged counterexamples. But these so-called "counterexamples" depend on construing my thesis that *p* in a way that it was obviously not intended – for I intended my thesis to have no counterexamples. Therefore *p*. (Chalmers 2025)

In a similar (impious and fictional) spirit, here's Lewis's Proof that Probably P.

> Horowitz and Priczki have insinuated that P is improbable, on the basis of alleged unlikely consequences. But the claimed improbability of these consequences depends on assuming that not-P – for, given P, they are not unlikely in the first place. (Therefore probably P.)

[2] I am grateful to Daniel Nolan and Jason Grossman here. In the version given by Lewis, the imagined opponent is a Mr Z, not Zabludowski, but the same unusual initial perhaps speaks to a common ancestor.

The optional final step makes a very big difference, obviously – a huge trade of plausibility for strength. More on that below, but for the moment, these are our questions. Does Lewis's defense of time travel overgeneralise, in a way that requires that it be qualified in some way? If so, what's the upshot of such a qualification, for the time travel case?

Let's take this in two stages. First, to get another example on the table, I want to introduce a famous thought experiment. Like the grandfather paradox, it can be presented as a *reductio.* It is widely regarded as a strong argument, and yet there's an apparent response to it, analogous to Lewis's defense of time travel.[3] Where does this leave us? I'll address that question at the second stage, by setting all three examples – time travel, future low entropy, and the one I'm about to describe – in a familiar general framework.

4. Dumping the Bodies

In the early seventeenth century, reflective people knew that heavier bodies fall faster than lighter bodies. If they wanted an authority for such a self-evident truth, Aristotle gave them one. Famously, Galileo challenged this received wisdom in *Dialogues Concerning Two New Sciences* (Galileo 1954). Let's borrow a formulation of Galileo's argument from John Norton.

> The argument is a reductio ad absurdum:
>
> 1. Assumption for reductio proof: The speed of fall of bodies in a given medium is proportionate to their weights.
> 2. From 1: If a large stone falls with 8 degrees of speed, a smaller stone half its weight will fall with 4 degrees of speed.
> 3. Assumption: If a slower falling stone is connected to a faster falling stone, the slower will retard the faster and the faster speed the slower.

[3] Or, more precisely, to Lewis's explanation of how his imagined time traveler, Tim, will fail to kill his grandfather: 'For some commonplace reason. Perhaps some noise distracts him at the last moment, perhaps he misses despite all his target practice, perhaps his nerve fails, perhaps he even feels a pang of unaccustomed mercy.' (Lewis 1976, 150)

4. From 3: If the two stones of 2 are connected, their composite will fall slower than 8 degrees of speed.
5. Assumption: the composite of the two weights has greater weight than the larger.
6. From 1 and 5: The composite will fall faster than 8 degrees.
7. Conclusions 4 and 6 contradict.
8. Therefore, we must reject Assumption 1.
9. Therefore, all stones fall alike. (Norton 1996, 341 – 42)

Like the bilking argument, then, this is a thought experiment by *reductio.* And it is easy to imagine a response, modelled on Lewis's defense of the possibility of time travel. We know on logical grounds that the result of such an experiment would not actually be a contradiction: the body in question would not fall both slower and faster than 8 degrees of speed. What should Aristotle expect?

Norton himself discusses the possibility of systematic loopholes, arguing, for example, that the argument depends on 'tacit assumptions' such as this:

> 8a. Assumption: The speed of fall of bodies depends only on their weights. (Norton 1996, 342)

One way for Aristotle to escape Galileo's *reductio* is to do so systematically, identifying and rejecting some such tacit assumption. But there's another option. Experiments often fail, for boring, accidental reasons. Here 'fail' covers a usefully wide range, from experimenters slipping on banana skins, to many sorts of error within the apparatus itself. By Aristotle's lights, Galileo has described a case in which, as a matter of logic, the natural tendencies of falling bodies will not be wholly manifest. If there isn't a principle to explain why not, random contingencies will do the job instead. We might object that it is very unlikely that experimental errors will conspire to hide the truth in this way, but this begs the question against Aristotle, in the way we have described.

For all that Galileo's argument shows, then, we may be living in an Aristotelian world. Such a response may seem absurd, but Aristotle has some powerful modern allies. Some of the most influential are intellectual forebears of Lewis himself. Let's invite them to the table.

5. Lessons from *Großvater*

Lewis's *Doktorvater* was W. V. Quine (a feather in both caps, obviously). One of the doctrines that now bears Quine's name is the Duhem-Quine thesis. Here Duhem, as the senior partner, may be regarded as the *Doktrinvater*. By a somewhat gerrymandered *Dokt'vater* relation, then, we may regard Duhem himself as one of Lewis's distinguished intellectual *Großväter*. Let's see what he and Quine have to tell us about Lewis's defense of time travel.

One of the central tenets of the Duhem-Quine thesis is that there is no such thing as a decisive refutation of a single scientific hypothesis. This claim has two aspects to it. One is a kind of holism, pushed to a famous limit in Quine's presentation of the view, as we'll see in a moment. The other, more relevant here, is the fact that any claimed refutation of a scientific hypothesis is bound to depend on auxiliary hypotheses. In the face of recalcitrant experience there will always be an option of abandoning some of those auxiliaries, instead of the target hypothesis.[4]

Duhem himself puts the central point like this.

> In sum, the physicist can never subject an isolated hypothesis to experimental test, but only a whole group of hypotheses; when the experiment is in disagreement with his predictions, what he learns is that at least one of the hypotheses constituting this group is unacceptable and ought to be modified; but the experiment does not designate which one should be changed. (Duhem 1914 – 1954, 187; quoted in Stanford 2023)

For Duhem this was merely a thesis about physics. In Quine's famous development of the idea, it becomes a view about the 'totality of our so-called knowledge or beliefs'. Here is a well-known passage from 'Two Dogmas'.

> The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the profoundest laws of atomic physics or even of pure mathematics and logic, is a man-made fabric which impinges on experience only along the edges. Or, to change the figure, total science is like a field of force

[4] Even worse, there will be the possibility of inventing new hypotheses, with new forms of relevance, and attributing the recalcitrant experience to those.

> whose boundary conditions are experience. A conflict with experience at the periphery occasions readjustments in the interior of the field. But the total field is so underdetermined by its boundary conditions, experience, that there is much latitude of choice as to what statements to reevaluate in the light of any single contrary experience. No particular experiences are linked with any particular statements in the interior of the field, except indirectly through considerations of equilibrium affecting the field as a whole. (Quine 1951, 42 – 43; quoted in Stanford 2023)

Setting aside Quine's willingness to revise even logic, it is easy to see the relevance of this to our current concerns. If you accept this thesis, then of course there's no knock-down argument against time travel, or the hypothesis that one is oneself a time traveller. At least in empirical matters, there are no knock-down arguments against anything.[5]

Is this good news or bad news for time travel? On the one hand, it puts Lewis's defense of the possibility of time travel on solid foundations, or at least widely-shared foundations. (After all, one of the lessons of the general case is that there is less bedrock than we might have hoped for.) But if the effect of this is to confirm that time travel is in the same boat, or raft, as Aristotelian mechanics, that's rather cold comfort. The Duhem-Quine thesis may prevent us from regarding Aristotelian mechanics as *contradictory,* but it doesn't commit us to giving it significant credence.

This is the point left ambiguous by our probabilistic version of Goodman's Proof that P. A sound dismissal of an argument for *improbability* need not be an argument for *probability,* or *plausibility.* With this sorted out, what can we say about time travel? How should loyal descendants of Duhem and Quine address the issue? And do the analogies we have drawn with future low entropy and Aristotelian mechanics have anything useful to say?

[5] This need not diminish Lewis's achievement in defending the possibility of time travel. As Sherlock Holmes reminds us, it is one thing to know that there is a culprit, quite another to find a plausible candidate.

6. Meeting the Market

At this point, we need a convenient path through a large forest. The methods and rationality of theory choice, downstream from Duhem and Quine (not to mention, e.g., Kuhn), are huge topics. How can we find something condensed enough for current purposes, without too much risk of misrepresentation? Answer: by letting Lewis himself be our guide. In well-known passages in *On the Plurality of Worlds* (Lewis 1986), Lewis offers an explicitly Quinean case for realism about possible worlds. As he says, it is modelled on a similar case for the existence of sets. Neither example is empirical, but they give us enough of a sense of Lewis's view of the Quinean method to apply it to our present cases, without serious risk of getting him wrong.

Here are some key passages from (Lewis 1986). First, the case of set theory:

> Set theory offers the mathematician great economy of primitives and premises, in return for accepting rather a lot of entities unknown to *Homo javanensis*. It offers an improvement in what Quine calls ideology, paid for in the coin of ontology. It's an offer you can't refuse. The price is right; the benefits in theoretical unity and economy are well worth the entities. (1986, 4)

Next, the extension to possible worlds.

> As the realm of sets is for mathematicians, so logical space is a paradise for philosophers. We have only to believe in the vast realm of possibilia, and there we find what we need to advance our endeavours. We find the wherewithal to reduce the diversity of notions we must accept as primitive, and thereby to improve the unity and economy of the theory that is our professional concern – total theory, the whole of what we take to be true. What price paradise? If we want the theoretical benefits that talk of possibilia brings, the most straightforward way to gain honest title to them is to accept such talk as the literal truth. It is my view that the price is right, if less spectacularly so than in the mathematical parallel. The benefits are worth their ontological

> cost. Modal realism is fruitful; that gives us good reason to believe that it is true. (1986, 4)

These are not empirical cases, but they give us a sense of what is in play, by Quinean and Lewisean lights. Is a proposal fruitful? Are the benefits worth the costs? Let's think about our two comparison cases, and then about time travel.

Galileo's thought experiment creates an anomaly for Aristotelian mechanics, a place where additional measures are needed to avoid contradiction. We saw that those measures might be systematic, in identifying additional dispensable assumptions on which Galileo's argument relies. Or they might be *ad hoc,* in the sense of banana skins. Either way, Aristotle has his loophole, and it won't do simply to declare it improbable – that risks begging the question against Aristotle, as we've seen.

But the components of Quinean ideology and ontology are a different matter. We may find the additional Aristotelian measures inelegant and expensive. Then the fact that they turn out to be entailed by the theory, in the light of Galileo's argument, gives us no reason not to debit them to its account – quite the contrary. The Galilean alternative needs no such measures, is elegant and economical in comparison, and turns out to lead to Newton's paradise. It's a steal, in comparison.

We may be idealising here, but the point we need is simple. Defensibility *in principle* does not guarantee plausibility. If it did, the Duhem-Quine thesis would be a disaster for theory choice – every option would turn out to be plausible.

Now to the case of future low entropy. Recall that the Gold Hypothesis (GH) postulates that the increase in entropy in our region is due to the expansion of the universe. If it were plausible, GH would perform an immense theoretical service, in linking thermodynamics to cosmology in this way. But from GH it follows directly, as we noted above, that entropy will decrease again if – a big if – the universe eventually recollapses. The conclusion can only be blocked by some move to break the evident symmetry between the two cases – a move that is bound to look *ad hoc,* unless it has some other motivation.

Perhaps even more strikingly, the probabilistic argument *against* entropy decrease is an argument form with two possible applications, in the

two directions of time. In endorsing the Past Hypothesis (Eddington's 'anti-chance') we already concede that it fails in one of these two possible applications, towards the past. From this point, it looks blockheaded to try to insist that it must nevertheless be a good argument in its other possible application, towards the future.[6]

As it turns out, GH seems unsuccessful for other reasons: GH does not guarantee PH (Price 1996), and even if it did, expansion by itself does not do the necessary work (Wallace 2010; Rovelli 2025). Moreover, it now seems unlikely that the universe will recollapse. However, there are other ways in which the combination of (i) an attempt to give a cosmological explanation of the low entropy past, and (ii) a respect for time-symmetry can lead us to similar conclusions.[7]

Summing up: at least in some cosmological models, the proposal that entropy will eventually decrease offers a large gain in symmetry and simplicity. And the probabilistic argument against it is already highly questionable, on symmetry grounds, before this proposal hits the table. In this case, then, it is easy to see how the proposal might emerge as the economical choice, in Quine-Lewis terms.

7. What Price Time Travel?

What about time travel and backward causation? They are certainly more interesting than Aristotelian mechanics, by contemporary standards. It is easy to find reasons in modern physics for bringing them to the marketplace, as Ismael points out.

> Lest it be thought that the bilking argument only has interest as an objection to a class of esoteric physical hypotheses, closed

[6] Compare this sales pitch, from a team of rocket scientists we'll identify only as X. "Here's your spaceship to the stars. It's extremely reliable. We have the best statistical mechanics on the planet, and they've checked every angle. We even fired an identical copy in the opposite direction, just to be sure." "Did it work?" "No, it blew up on launch. But this one, as we say, is extremely reliable."

[7] In TAAP, Chap 4, I criticise both Hawking and Penrose, among others, for paying insufficient attention to the consequences of symmetry in these issues.

> causal loops are less star-trekky than one would think. Godel's solution to the equations of General Relativity forced physicists take them seriously as nomological possibilities, tachyons have been explored as potential explanations for quantum non-locality, and it is not as clear as had once been thought that superluminal signaling or influence of certain kinds is impossible in Minkowski space-time. (Ismael 2003, 306 – 307)

Indeed, readers familiar with TAAP will know that I argue at length for taking retrocausality seriously. I claim that it has the potential to solve one of the deepest puzzles of quantum mechanics (QM). One of the great challenges of modern physics is to reconcile the work of John Bell in QM with Einstein's theory of relativity. On the face of it, Bell's discovery of 'non-locality', now confirmed in many experiments, seems to require some sort of spacelike action-at-a-distance, in tension with relativity.

As Ismael says, some writers have suggested that tachyons might help to explain nonlocality. Tachyons are (hypothetical) particles travelling faster than light. They thus have spacelike worldlines, and so travel backwards in time in some inertial frames; that's why they seem to imply retrocausality. But this is not what I had in mind in TAAP. In my view, if we're paying for retrocausality anyway then it is much more economical to confine it to past light cones. There, we have all the ordinary particles. If the particles involved in EPR-Bell experiments can be influenced by measurements they are to encounter in the future, then there's no need for direct spacelike influence (or tachyons). Instead, there's a zigzag causal path, via the worldlines of the particles themselves and the point at which they interact.

If this proposal worked, it would be a very strong reason for taking retrocausality seriously. However, it is unclear that it would rely on Lewis's defense. Most advocates of the proposal in QM, including me, propose (i) that it would operate at the level of so-called hidden variables, and (ii) that the kind of restrictions that QM puts on what can be measured would prevent access to this information, in the way that would lead to Lewisean causal loops. This is what I meant by 'a much larger loophole in the bilking argument' (TAAP, 278) in the passage quoted above. It is the loophole identified by Michael Dummett (1964): if information about the past effect can't reach the future cause, there's no loop to worry about.

Still, as Ismael says, there are reasons in modern physics for giving causal loops a table in the marketplace. They are not to be driven out by fallacious or question-begging arguments about inconsistency or improbability. But this is not yet a reason for buying what they are selling – not yet an argument for their *plausibility,* or *palatability.* I'm not aware of any strong arguments for admitting time travel, or the causal loop-generating kind of retrocausality, as part of the accepted account of our actual world.

On the other side, several factors may be regarded as excessive costs. Causal loops would open the door to spacelike signalling, violating a principle often felt to be crucial in QM. One reason it is regarded as important is that it seems to offer a path for peaceful co-existence between QM and relativity. Nonlocality *without* signalling seems much less of a threat to relativity than nonlocality *with* signalling.

For this reason, many contemporary physicists are very reluctant to entertain theories that permit spacelike signalling. We may think of this as an aesthetic or pragmatic preference, but that's a good enough reason to avoid the causal loop table (absent any strong countervailing reason for heading in that direction). If we want an authority for relying on a pragmatic principle here, then again, Lewis's *Doktorvater* is the one we need. Here is the closing sentence of 'Two Dogmas'.

> Each man is given a scientific heritage plus a continuing barrage of sensory stimulation; and the considerations which guide him in warping his scientific heritage to fit his continuing sensory promptings are, where rational, pragmatic. (Quine 1951, 46)

It is no argument against a pragmatic preference of this kind to point out that the theories for sale at your table do not respect it. Perhaps if you can point to previous customers, obviously thriving after buying your product, then you have the beginnings of an argument. One way or another, you need some countervailing advantage, to set against the present pragmatic preferences of reluctant browsers.

For comparison, imagine you are selling Thai food. Many customers prefer to avoid it, because they don't like chilli. Unless you can persuade them to change these preferences, you need to offer them some countervailing reason for buying your wares. ("Sure, it's a little spicy, but it is also tasty, economical, and very nutritious, compared to the rubbish on sale

elsewhere in the market!") The mere fact that Thai food is *possible,* and perfectly normal by its own standards, isn't going to do the trick.[8]

This brings us back (appropriately) to where we began: to the claimed improbabilities of time travel and causal loops. Ismael has a characteristically vivid discussion of the kind of phenomena in question, using an example she calls Earman's rocket. This is a device built to exploit closed timelike curves, in order to prevent its own launch. Earman introduces it like this.

> Consider a rocket ship which at some space-time point x can fire a probe which will travel into the past lobe of the null cone at x. Suppose that the rocket is programmed to fire the probe unless a safety switch is on and that the safety switch is turned on if and only if the 'return' of the probe is detected by a sensing device with which the rocket is equipped. (Earman 1972, 23–32)

Ismael argues persuasively that 'there is no contradiction in supposing there are Earman rockets' (2003, 307)

> What is true is that, holding fixed the laws that govern the process, if *everything works as it should,* the rocket fires *iff* it doesn't fire. And we can deduce therefrom that – again, holding fixed the laws – everything never works as it should. We can deduce from the description of the rocket and the laws which are supposed to govern it that something goes wrong, somewhere along the way, every time. Some contingency arises to spoil things, some bug in the program keeps the rocket from firing or some malfunction interrupts the process after the firing, before it is inhibited by the safety switch. If I get a grant from the NSF, set up 70 Earman rockets, and hire lab technicians to sit and press the buttons that activate their programs once a minute, 16 hours a day, for 17 long years, the result will certainly be that each time any of the buttons is pressed, some kind of system failure or malfunction – a different one, perhaps, every time – occurs. (2003, 307)

[8] Not, at least, unless the culinary landscape changes, so that all the self-consistent options are now spicy – Mexican, Pakistani, and Szechuan, perhaps, as well as Thai. Now there's no longer a pragmatic reason for avoiding the Thai option.

Later in her paper, Ismael makes the set-up even more unusual. She adds a button to the rockets which, when it is down, disables the mechanism whereby a fully-functioning Earman rocket would prevent its own launch. '[D]o the correlations in the new set-up constitute a real anomaly?', she asks (2003, 311).

> Certainly, they can be made to look strange; we can arrange a thousand Earman rockets side by side, each manned by a lab assistant who launches his rocket with the hazard button up or down at his whim, and do run after run of the experiment. What we observe – and it is important here that I am not reporting an experimental result; things couldn't go otherwise – is that, although things almost always proceed without a hitch when the button is down, in every case in which it is up, some foiling event – an event with a perfectly determinate and independent causal history, an event which was bound to happen and (if our rockets were well-designed) bound to occur for reasons that had as little to do with the position of the hazard button as the price of tea in China – occurs. (2003, 311)

Ismael goes on to argue that even here, we have no genuine anomalies or inexplicable coincidences. And she's right, provided that we assess the question under the hypothesis in question, namely, that the world permits the construction of Earman rockets. However, this is perfectly compatible with its looking very strange indeed to inhabitants of realms like ours (where, I'm assuming, we have never encountered anything like this).

So, modified or not, Earman rockets are not anomalous by the standards of the time travel table. But to the casual browser, equipped with her own expectations and pragmatic guidelines for theory choice, they may still look very unpalatable indeed.[9]

[9] "Spacelike signalling? No, I'm afraid I don't have the stomach for that. It's worse than coriander!"

8. Conclusion

Summing up, what can we say about time travel? Two things: Price was wrong, but *its* price isn't currently right. We shouldn't ignore the time travel shop, peddling its unconventional theories. But there's no irrationality in regarding what it offers as a very poor deal, in the light, in part, of the kinds of strangeness that Horowitz and Priczki misrepresent as improbability.

In fairness to Lewis, Horwich, and the author of TAAP, let's close with the following note. Lewis himself is explicit that he regards Quine's theory market as a guide to truth. As he says: 'Modal realism is fruitful; that gives us good reason *to believe that it is true*' (1986, 4, emphasis added). By Lewis's own standards, then, the pragmatic unpalatability of time travel theories seems to be a reason *to believe them false* – in fancier words, to assign them low epistemic credence. So there seems to be Lewisean route to the conclusion that time travel is highly improbable – a route that avoids question-begging objective probabilities, appealing, instead, to the alethic implications of pragmatic unsaleability. Like the question-begging argument, this route depends on appeal to the distinctive consequences of causal loops. But it presents them not as objectively improbable coincidences, but as aesthetic and pragmatic reasons to purchase one's theories elsewhere – that being an argument for the *falsity* or *subjective improbability* of the theories thereby rejected.

By my lights, and perhaps by Horwich's, the last step either adds nothing, or relinquishes hard-won pragmatic territory. Either way, I'm not inclined to defend it as an additional argument against time travel. But it does seem fair to leave it on the table as a challenge for would-be defenders of time travel, if they regard Lewis's (1976) argument as a path to anything much stronger than mere consistency.[10]

[10] Warm thanks to Brian Garrett, Jason Grossman, Jenann Ismael, Daniel Nolan, and Nicholas J. J. Smith, for comments on drafts of this material.

References

Albert, David. 2000. *Time and Chance*. Cambridge, Massachusetts: Harvard University Press.

Chalmers, David. 2025. Proofs that p. Webpage accessible at https://consc.net/misc/proofs.html (accessed 4 June 2025).

Duhem, Pierre. 1954. *The Aim and Structure of Physical Theory*, translated by P.W. Wiener. Princeton, New Jersey: Princeton University Press.

Dummett, Michael. 1964. "Bringing about the Past." *Philosophical Review*, 73: 338–59.

Earman, John. 1972. "Implications of Causal Propagation outside the Null-Cone." *Australasian Journal of Philosophy*, 50: 222–37.

Eddington, Arthur. 1928. *The Nature of the Physical World*. Cambridge: Cambridge University Press.

Eddington, Arthur. 1931. "The End of the World: From the Standpoint of Mathematical Physics." *Nature*, 127: 447–53.

Galileo, Galilei. 1954. *Dialogues Concerning Two New Sciences*, translated by H. Crew and A. de Salvio. New York: Dover Publications.

Gold, Thomas. 1962. "The Arrow of Time." *American Journal of Physics*, 30: 403–10.

Horwich, Paul. 1975. "On Some Alleged Paradoxes of Time Travel." *Journal of Philosophy*, 72: 432–44.

Horwich, Paul. 1987. *Asymmetries in Time*. Cambridge, Massachusetts: MIT Press.

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

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

Lewis, David K. 1984. "Putnam's Paradox." *Australasian Journal of Philosophy*, 62: 221–36.

Lewis, David. 1986. *On the Plurality of Worlds*. Malden, Massachusetts: Wiley-Blackwell.

Norton, John. 1996. "Are Thought Experiments Just What You Thought?" *Canadian Journal of Philosophy*, 26: 333–66.

Penrose, Roger. 1989. *The Emperor's New Mind*. Oxford: Oxford University Press.

Price, Huw. 1996. *Time's Arrow & Archimedes' Point: New Directions for the Physics of Time*. New York: Oxford University Press.

Quine, W.V.O. 1951. Two Dogmas of Empiricism. In his *From a Logical Point of View*, 20–46. Cambridge, Massachusetts: Harvard University Press.

Rovelli, C. 2025. "Back to Reichenbach." *Journal for General Philosophy of Science*, 56: 97–115.

Smith, Nicholas J.J. 1997. "Bananas Enough for Time Travel?" *British Journal for the Philosophy of Science*, 48: 363–89.

Smith, Nicholas J.J. 2024. Time Travel. In *The Stanford Encyclopedia of Philosophy*, edited by E.N. Zalta and Uri Nodelman (Summer 2024 edition). https://plato.stanford.edu/archives/sum2024/entries/time-travel/

Stanford, Kyle. 2023. Underdetermination of Scientific theory. In *The Stanford Encyclopedia of Philosophy*, edited by E.N. Zalta and Uri Nodelman (Summer 2023 edition). https://plato.stanford.edu/archives/sum2023/entries/scientific-underdetermination/

Thom, Paul. 1974. "Time-Travel and Non-Fatal Suicide." *Philosophical Studies*, 27: 211–16.

Wallace, David. 2010. "Gravity, Entropy, and Cosmology: In search of Clarity." *British Journal for the Philosophy of Science*, 61: 513–40.