

Timetravel and Penrose Stairs

John Bigelow

(Monash University)

There is a famous drawing by M.C. Escher, *Ascending and Descending* (1960), which echoes the “Penrose Stairs” of Lionel and Roger Penrose, *British Journal of Psychology* (1958), and earlier pictures of “impossible objects” by the Swedish artist Oscar Reutersvärd, dating from 1934 (see Seckel, 2004). In Escher’s drawing, some monks are climbing up a spiral staircase, and they are passing other monks, on their right, who are descending on the same staircase. The picture can be seen as depicting a “double helix”. It can also be seen as a timelapse representation of a simple narrative, in which one monk is ascending, and another descending, the same spiral staircase. The distinctive feature of this little narrative is the striking fact that the ascending monk, after taking enough steps *upward* from any given step, will eventually take a last step *upwards* onto that very same step: and likewise the descending monk descends from any given step *downwards* until he returns to the very step he has left so far above him. I will argue that a timetravel story recounts impossible sequences of events, in a manner very similar to that in which Escher’s drawing depicts an impossible world, and an impossible sequence of events in that impossible world.

In a very influential paper, David Lewis argued that it is possible to tell a logically consistent timetravel story: and he tells a very short timetravel story that, he claims, is internally consistent. Lewis’s very short story tells of the adventures of someone called Tim, who travels back in time intending to kill his own grandfather, but by some fluke he fails in this attempt, and his grandfather survives long enough to father Tim’s father. Lewis also refers to longer timetravel stories, like Heinlein’s “By his bootstraps” (1959) (and I think also of Hanley’s *The Metaphysics of Star Trek*, 1997), which flesh out more details but have the same essential structures as Lewis’s own little thumbnail sketch. In the background behind Lewis’s paper, there is quite a lot at stake: the alleged possibility of timetravel is indirectly linked to philosophical positions Lewis holds concerning, for instance, the illusoriness of the passage of time, the nature of personal identity, the possibility of personal survival after death of the body, the compatibility of free will with determinism, the reduction of causation to mere regularities, an atomistic picture of the whole world as consisting merely of a mosaic of “just one little thing and then another”, with no God or laws of nature guiding the motions of the stars, and no human agents acting like little gods, and guiding events in any such manner, each within his or her own limited sphere of influence. If any timetravel story is internally consistent, then Lewis’s story about Tim is; yet I will argue that it is not. Lewis’s story contains a hidden contradiction. But the hidden inconsistency is very hard to spot.

I used to believe otherwise. Indeed, I used to believe not only that you could travel “backward” in time, but even that you could travel “sideways” in time: this being a notion Donald Williams mentioned in his great paper on “The myth of passage” (1951). Williams suggested that there could be regions of spacetime in which what is, for us, a spatial dimension is, for the local inhabitants of that region, a temporal dimension. On this theory, which of the four dimensions is a “time” dimension, within a region, depends only on the direction in which “causal dependences” connect localized events within that region. Williams was a very important influence on David Lewis; and like Lewis, I found “The myth of passage” extremely persuasive. I still find it seductively plausible, but have come to suspect that it just may not be true after all. Some timetravel stories do seem to be internally

consistent; but now I think that only the surface details are consistent, concerning what localized properties are instantiated where, but the underlying “necessary connections”, which implicitly hold according to the story, must conceal a hidden inconsistency.

Thinking about timetravel fictions, and exploring the question of whether there are any hidden inconsistencies, “according to the story”, is a useful way of probing a variety of philosophical problems. In addition, some of the considerations concerning what is implicitly true “according to a story” may carry over into parallel reasoning concerning what is implicitly true “according to a scientific theory”. If I am right about science fiction stories, and if current physical theories do permit the existence of things called “closed timelike curves”, then this jointly suggests that there is something wrong with certain very tempting interpretations of the physical theory, which construe the possibility of “closed timelike curves” as entailing the possibility of timetravel. Either the physics is mistaken, or else it is being misinterpreted, if we take it as entailing the possibility of timetravel.

It is impossible to take a number of steps upwards on a spiral staircase, and yet for the n^{th} step *upwards* to land you on the *bottom* step, where you began. Yet Escher’s “Ascending and Descending” represents any given step as being *above* the one that lies to one side of it, and yet also *below* the one on the other side – which in turn is *below* the next one after that, ... and so on. The picture represents each step as not only being below the one adjacent to it, but also being below the one that *that* one is below. When A is represented as below B, and B is represented as below C, then A is also, indirectly, represented as below C. And so on, for a chain of steps. So each step is represented as being below, and also as being above, the one next to it. This is not possible – at least, not within the *kind* of space that is being represented within this drawing. The whole point of the picture is, that it represents something that is impossible, in the very kind of space that is being depicted. So the drawing depicts an impossible world.

A timetravel story is like that: one event is represented as “resulting from” one thing, and “leading to” another, which in turn “leads to” another, ... and so on: from which it follows that there is an event that “results from” the very thing it “leads to”. This is, I will argue, impossible for the same kind of reason that it is impossible for something to be both above and below another, within the kind of space that is represented in Escher’s “Ascending and Descending”.

Nevertheless, it is possible to have a spiral staircase, with a lowest and a highest step, and yet for there to be a particular vantage-point from which you will obtain the visual *illusion* that you can step *upwards* from what is in fact the highest step onto what is in fact the lowest step. Indeed, there are many equally possible spiral staircases, each with a different lowest and highest step, and each of these possible staircases will look visually indistinguishable from all the others, and from Escher’s “Ascending and Descending”, when viewed from precisely the right angle. Hence, for any two particular steps that appear to be adjacent to one another, it is perfectly possible that they are indeed adjacent to one another. Any particular part of the appearance could be veridical. This makes it very hard to “spot the inconsistency” in the picture. For each particular step it is perfectly possible that its apparent movement upwards – along with the upward appearance for *all* the others *bar one* – could well be perfectly veridical. In these reflections on impossible drawings, I am guided by Max Cresswell’s “A highly impossible scene” (1983), and *Structured Meanings* (1985).

A timetravel story is like that: there are many possible worlds, each of which includes *all the narrative steps bar one*. And for every particular narrative step there is a possible world that includes events precisely matching that narrative step – *and all the others bar one*. There are, as it were, many

consistent timetravel “spiral staircases”, in which there is just one point in the story at which one thing *actually* happens “after” another, but *appears* to come “before”, in some subtle and yet to be explained sense of “after” and “before”.

I will illustrate by describing a logically consistent story that runs very closely parallel to David Lewis’s story of timetraveling Tim. But my consistent story is a “spiral staircase”, whereas Lewis’s story depicts a “closed causal loop”. My story is, in fact, another version of a non-timetravel story that Lewis tells, in the same paper “The paradoxes of time travel”, as a story of a demon who creates two people, Fred and Sam, who jointly, as “Fred *cum* Sam”, *superficially resemble* possible timetraveller, and yet they are not the same person, for lack of the right kind of counterfactual dependences.

Tidy Timelords

There is, I will assume, no internal consistency in a story according to which, at some time in 1920, in a room in an empty warehouse, a book depository, in a room with an open window overlooking the street, a contraption appears out of nowhere, bearing the label “Time Machine”, and a man emerges from this contraption armed with a rifle. At least: something like that could appear from nowhere, and something could emerge is at least an atom-for-atom *duplicate* of a person armed with a rifle. I assume it is not internally inconsistent to suppose that such a thing could happen, even though there had been nothing at any earlier time which could furnish anything like a plausible candidate for a “cause” of this event. It is not inconsistent to suppose that such a thing might “just happen”, uncaused. Certainly Lewis thought so. (Or else, you might imagine the cause to be the whim of some divinity.)

It is possible for such a person to have various sorts of delusions, to think his own name is Tim, and to be trying to kill someone he *believes* to be his grandfather. Meanwhile, we may suppose, on a grassy knoll alongside the warehouse, there is someone else, Tom – whose existence was caused in the ordinary way, by birth and upbringing – someone who has no delusions about his past – and who is, like Tim, aiming to kill someone, namely Grandfather’s partner in a munitions factory. By chance, on one day in 1921, both Tim and Tom fire at their prospective targets, but their bullets collide in midair and Grandfather survives. In due course, in 1922, Grandfather fathers someone who fathers a little boy, and his parents call him Tim.

There may also be, we might suppose, according to my story, a number of “Timelords” in the background, who make sure, in one way or another, that this grandson called Tim grows up to be more and more like the deluded fellow, also called Tim, who appeared in the 1920s. For many of the delusive memories that 1920s-Tim harboured, the Timelords ensure that new-Tim has experiences that match, almost exactly, those delusive memories of 1920s-Tim. These Timelords are aesthetically offended by the appearance of 1920s-Tim, because they want to preserve certain sorts of regularities in the world. For instance, they want it to be the case that, whenever there is someone with a distinctive kind of pain in their toe, then there is a suitably corresponding event of toe-stubbing. Usually, the Timelords can rely on the world to supply an *earlier* toe-stubbing for each toe-pain that you find in the world. But when Tim appeared out of nowhere, with that kind of pain in his toe, the Timelords could find no *earlier* event of the right kind. So the Timelords make sure that there will eventually be the right kind of event in the future, to restore the regularity: “for each pain like that, a corresponding toe-stubbing”. So the Timelords watch the development of Tim, and nudge events in the required directions, whenever necessary. Eventually new-Tim is the spitting image of 1920s-Tim, and steps in something he believes to be a “time machine”, with the intention of killing the person he

rightly believes to have been his own grandfather. As he steps into the machine, the Timelords make sure he stubs his toe, and as soon as he closes the door of this contraption, it disappears into thin air.

This completes the top step in a spiral staircase narrative, in which, at the “top” step, new-Tim has the delusion that the next step “upwards” will take him to the 1920s. He has the delusion that events in the 1920s depend on what he is currently doing. For instance, he has the delusion that 1920s-Tim had a pain in his toe *because* he, new-Tim, has just stubbed his toe when getting into the “time machine” – whereas, in my story, in fact the dependence works the other way around: the Timelords have ensured that new-Tim stubs his toe *because* they know that 1920s-Tim had a pain in his toe, and they wanted to ensure that for every such pain there is a suitably corresponding stubbing.

The Visions of Disembodied New-Tim

There are other possible, consistent, “spiral staircase narratives”, which match all the surface details of Lewis’s story of Tim, but which speak concerning certain matters where Lewis’s story remains conveniently silent – and which do explicitly specify *which one* of the steps, which *seems* to be “below” the next, is really “above” it.

For instance, there is a story like the last one, except that when Tim gets in the “time machine” and *physically* disappears, his soul lives on and has experiences exactly matching all the events undergone by 1920s-Tim. According to this story, things that 1920s-Tim did were actually, physically, caused by 1920s-causes, not by any willings of disembodied new-Tim. Nevertheless, the disembodied new-Tim has the delusion that he is living in the 1920s, and is freely choosing what to do, and is thereby actually causing those events back in the 1920s. In this story, there is a match of many surface details, but a reversal of some of the “orders of dependence”.

In this story, unlike the last one, there really is someone, namely disembodied new-Tim, who has all the *memories* he does because of the things the Tims have done. But in this story – in getting the “right” direction of dependence for *memories* – we reverse the normal dependences between *willings* and actions. In this story, the actions of 1920s-Tim did not depend on the willings of new-Tim – but rather, new-Tim has the experiences as of willing, as he does, *because* of the actions of 1920s-Tim.

Or alternatively, it might be a sheer coincidence that the delusions of embodiment, and of willing, which are experienced by disembodied new-Tim, just *happen* to match the actions of 1920s-Tim. Again we have a “spiral staircase”, in which one of the steps is actually “below” the next – even though, by sheer coincidence (according to the narrative), this step superficially looks *exactly* as it would look if it were in fact “above” it.

Counterfactual dependences

In a timetravel story, for each sizeable chunk of the narrative, there is indeed a possible world in which all these things really do happen, exactly as described – there are many “spiral staircases”, each with a different particular step at which the apparent circle is secretly broken. This explains why it is very hard to point to any particular inconsistency within the narrative. Nevertheless, I will argue, there is no possible world with *all* the causal dependences that are represented as holding, according to the story. I will argue that you cannot have a “closed causal loop”. Lewis argued that timetravel stories do

indeed entail the possibility of closed causal loops; but he maintained that such loops are not internally inconsistent. I claim that they are – although, because they are representations that are closely analogous to Penrose Stairs, it is not easy to see exactly where the inconsistency lurks.

In *Real Time* (1981) *Real Time II* (1998), Hugh Mellor argued that “closed causal loops” are internally inconsistent. His argument develops key ideas that were broached in a paper by Michael Dummett, on “Bringing about the past” (1964). Mellor mounted his argument by first construing causation as requiring a certain sort of “probability-boosting”. This makes each causal, narrative step in the story very much like a step “upwards” on a spiral staircase. You cannot, Mellor then argued, keep going “upward”, in this sense, and eventually arrive back where you started. Things cannot get “pinker and pinker”, in a series, and arrive back at the one we started with. If causation were to require increase in some quantity, which might for instance be in some way analogous to “global entropy”, then you could not consistently have any closed causal loop. I think Mellor’s argument probably works. It relies on a theory that links causation with “probability-boosting”: as does Lewis in his paper on “Counterfactual dependence and time’s arrow” (1979). But I will run a different argument, not relying on “probability-boosting”, but on “counterfactuals”.

Lewis links personal identity with causation, for instance in “Survival and identity” (1976); and he links causation with counterfactual dependence in “Causation” (1973); and for his theory of counterfactuals, see his book *Counterfactuals* (1973). I assume Lewis is somewhere near the truth, in linking causation both with probability-boosting, and with counterfactuals. Even if he is wrong in some aspects of his theories about personal identity, causation, and counterfactuals, for my purposes he is near enough to the truth on all these matters. He is “right enough” to ensure that the counterfactuals I will be citing, in the arguments below, are all true *according to* the relevant timetravel stories. These counterfactuals are, I maintain, true according to these fictions, even if they are not explicitly articulated in those fictions: Lewis may not be entirely right in his theories about “Truth in fiction” (1978), but (again) he is “right enough” for the purposes of the argument I will be mounting.

Many critiques of Lewis over the last decade or so have focussed on questions about what a time traveller “can” and “cannot” do, especially if they set out to kill their own ancestors, as for instance Vihvelin (1996), Vranas (199?) and Kiourti (2007). The concerns in these papers extensively overlap those in the argument I will advance. There is even more overlap in a considerable number of papers that focus explicitly on counterfactuals, as for instance Smith (1997), Sider (2002) and Dowe (2007). My argumentation does, however, have some features that set it a little apart from these recent debates in the literature.

To begin to see why Lewis’s timetravel story is secretly inconsistent, consider the following fairly natural way of articulating “the Grandfather paradox”. I embellish this story into one in which Tim and Tom both try to kill someone Tim is calling Grandfather (even though this fellow is still a virgin schoolboy): but their bullets collide and deflect one another, so Grandfather survives and does indeed grow up to father Tim’s father.

Because this is a timetravel story, Tim is not just seeing visions of events in the 1920s, but rather, there really is a physical 1920s-Tim, who is fully present in the 1920s – and this means that not only can he see and touch people like Tom and Grandfather but they can see and touch him. Otherwise this would not be a timetravel story, but would be like the story of Ebenezer Scrooge, in Dickens’ *A Christmas Carol*, who can see a vision of people in a “Christmas past”, but they cannot see him –

because he wasn't in fact physically present at that occasion. No: Tim is not in that way like Scrooge. Like Scrooge, what he experiences depends on what happens in the relevant places at an earlier time, but in addition, unlike what we find with Scrooge, it is also true according to the story that what happens in the relevant places in that earlier time does depend on what Lewis's Tim actually decides to do.

For instance, perhaps Tim leaves footprints on the dusty floor of the room in the warehouse. This means that there are marks on the floor in the 1920s – and that the presence and characteristics of these marks are not merely accidentally related to the fact that Tim chose to walk precisely where he did, across the dusty room. There are many such *counterfactuals* that are true according to the story: things that, according to the story, *would have*, or *would probably have*, or at least *might have* happened differently, *if* (“contrary to the facts”) Tim had done something slightly different from what he did in fact do, according to the story.

Lewis himself stresses this very aspect of his timetravel story, and he introduces the character Tom into the story precisely order to drum this point home. Tom tries to kill Grandfather's Partner, but by an unexpected coincidence he fails/ But Lewis assumes that it clearly would have been possible for Tom to have succeeded. If Tim had fired just a little earlier, he would have succeeded. Locally speaking, Tim is just like Tom. Tim, like Tom, really is physically present at the scene, and so what happens around him depends at least in part, in the ordinary ways, on what he does. So Lewis concludes that there is at least one significant sense in which it *would* be possible for Tim to have succeeded in shooting Grandfather.

When Tim and Tom fire their bullets, one aimed at Grandfather and the other at his Partner, according to the story they are firing real bullets, and these bullets are moving according to the usual laws of motion. So the following counterfactual implicitly holds, according to the story: that if Tom had fired very slightly later than he did, then his bullet would not have deflected Tim's. But Tim's bullet was directly on-course for a fatal impact in Grandfather's heart. If Tom's bullet had not deflected Tim's, then Tim's bullet would have (or would *probably* have, or at least – surely – *might have*) hit Grandfather in the heart.

Someone who accepts Lewis's thesis, that timetravel stories can be logically consistent – seeing trouble brewing over the horizon – might be reluctant to accept that, according to the story, if Tom had fired earlier then Tim's bullet would (or might) have hit Grandfather in the heart. But to block this counterfactual from being true, according to the story, it would have to be implicitly true, according to the story, that there is some sort of a fail-safe, back-up system in place, to stop the bullet hitting its mark by some other means if Tom were to fire at the wrong time. I do not think it is implicitly true, according to the Lewis story of Tim, that there is any such back-up system in place. But suppose you were to entertain the notion that the story *could* be rendered consistent, if we *were* to suppose such a thing to be implicitly true, according to the story. Suppose that, according to the story, there are back-up systems ensuring such things as that, “If Tom's bullet hadn't deflected Tim's, then someone else's would have ...”.

There are certainly consistent stories, with a back-up system of this kind to guard against the worrying chain of counterfactuals that begins with the possibility of Tom having fired at a slightly different time. But there will be many other similar counterfactuals, worked into thousands upon thousands of chains of counterfactuals, all of them threatening to lead to trouble just over the horizon. Lewis's theory of personal identity, in “Survival and identity” (1976) ensures that there will be thousands of

interlocking chains of counterfactual dependence in any timetravel story, leading round in a closed loop, simply as a consequence of the continuing personal identities of the characters in the story. Even if you do not agree with all the details of Lewis's theory of personal identity, he is right at least about the fact that the personal identity of characters in a timetravel story do go along with thousands of interlocking chains of counterfactual dependence.

Suppose we were to deal with all these counterfactuals by the same strategy: namely, by supposing it to be implicitly true, according to the story, that there are a very great many such back-up systems in place. This will in effect foist on the story a Grand Conspiracy Theory, which is alleged to be implicitly true according to the story. The story will then turn into something equivalent to my story about the Tidy Timelords. Then the story will effectively cease to be a timetravel story. The implicit "closed loops" of a genuine timetravel story will have become mere "spiral staircases". All the "narrative dependencies" will run forwards in time. There will be none of the "backward dependencies" that Lewis needs, in order to ensure that 1920s-Tim really is the very same person – and a "narratively-later version" of the very same person – as the new-Tim that steps into the contraption labelled "Time Machine", with murderous intent.

So we must allow that it is implicitly true, according to the story, that if either Tim or Tom had fired the least bit earlier or later, then a bullet would (or at least might) have passed through Grandfather's heart.

So we may well ask: "what if Tim had killed Grandfather?" Lewis asks this in the penultimate paragraph of his paper. He here explicitly asks what *would* or *might* have happened if Tim had acted slightly differently and the bullet had passed through Grandfather's heart. After asking this crucial question, he gives a brief answer. But Lewis's answer is not satisfying:

Perhaps Tim might have been the time-traveling grandson of someone else. Perhaps he might have been the grandson of a man killed in 1921 and miraculously resurrected. Perhaps he might have been not a time traveller at all, but rather someone created out of nothing in 1920 equipped with false memories of a personal past that never was.

Lewis here gives us three options.

The first is that, according to the story, if Tim had fired slightly earlier, he might then have been the grandson of someone other than the man who is in fact his grandfather. I say that this is definitely *not* true, according to the story. And if it were, that would spell trouble of a different sort for Lewis. It is important to the story, if it is to be a timetravel story, that 1920s-Tim's existence and characteristics depend on 1976-Tim, not *vice versa*. But if we suppose that 1976-Tim's existence depends on Tim firing at exactly the right time, then the direction of dependence between the Tims runs the wrong way around.

Lewis's second option is that, even if Tim had fired slightly earlier, Grandfather would still have fathered Father in 1922, because something would have put the course of events back on track, as for instance if Grandfather were miraculously resurrected. There is indeed an imaginable possibility of this kind, but it belongs in a story with something like Tidy Timelords, not in a timetravel story. It is not true, according to the timetravel story that we take Lewis to have told, that "even if Tim had fired slightly earlier then something would have nudged the course of events back on track for the eventual production of Tim". Or rather, perhaps according to the story it *is* true that something would have to nudge events back on track: yet nevertheless, according to the story it is *also* false. It does need to be

false, as well as true, because otherwise 1920s-Tim would not depend on 1976-Tim in the way required for something to be a timetravel story, but rather, the existence or nonexistence of 1976-Tim would then depend on whether or not 1921-Tim, and Tom, fire at precisely the right moment to ensure that their bullets collide in midair.

And finally, the third option Lewis offers us is: that if Tim had fired very slightly earlier, then there would (or might) have been no conception of Father in 1922, nor of Tim in 1949, nor death of Grandfather in 1957, nor entry of anyone into something labelled “Time Machine” in 1976. I say that this, too, is an imaginable possibility in some story, but it is not true according to the story we have been told: not if the story is to be a genuine timetravel story. Again, if it *were* true, then the existence or nonexistence of 1976-Tim would then depend on precisely when 1921-Tim and Tom fired their rifles. For this to be a timetravel story, the dependences would have to run the other way around: the presence of 1920s-Tim in 1921 should depend on the actions of 1976-Tim, not *vice versa*.

To drive the point home more securely, we need to look not just at one counterfactual conditional in isolation, but at a vast forest of counterfactual conditionals, all of which are (I say) clearly true according to the story. Consider for instance counterfactual questions about Tom, as well as those concerning Tim. In “The paradoxes of time travel”, Lewis assumes that it is unproblematically possible for Tom to have succeeded in killing the Partner. If Tom had fired earlier, the Partner would have died.

Yet this is not as unproblematic as it might seem. If Tom had succeeded in killing the Partner in 1921, then there might have been no one suitably placed to introduce Grandfather to Grandmother, in 1922. It might be true according to the story that if the Partner had been killed in 1921 then Father would not have been conceived in 1922. And then: where would that gunman over there in the book depository have come from? Is it therefore true, according to the story, that if *Tom* had fired earlier, and killed the Partner, then *either* Tim would have been the grandson of someone else, *or else* that the Partner would later have been miraculously resurrected ... or *what?* No, I say, those counterfactuals are *not* true according to the story. (Or perhaps, even if according to the story they must be true, it is also the case that according to the story they are false.)

Indeed, if we grant Lewis that Tom could have killed the Partner, why could he not have killed Grandfather by accident, instead of the Partner? Indeed, if Tom had fired a second shot, and aimed it at Grandfather, surely *then* Grandfather would (or at least might) have died. But if Grandfather had died, then where did that other fellow, over in the book depository, come from? In imagining what would have happened if Tom had killed Grandfather in 1922, would we need to backtrack and rewrite the little bit of history that involved the first, 1920, appearance of Tim, followed by the planning of the assassination, stalking Grandfather’s movements, and so on, right up to the fatal events in 1921? There will be lots of counterfactuals that are true according to the story, about what memories Tom would have had in 1922, of his meetings with Tim in 1921, if indeed Tom *had* met Tim during Tim’s period of planning the assassination of Grandfather. We land ourselves with the same unsatisfactory options, in trying to flesh out many pertinent counterfactual suppositions about *Tom*, that Lewis faced in imagining how the story could have continued if *Tim* had succeeded in killing Grandfather.

As I have argued above, we must allow that, according to the story, it is true that if Tim had fired just a little earlier, then a bullet would (or at least might) have passed through Grandfather’s heart. For the same sorts of reasons, we must allow a long chain of such counterfactuals: if a bullet had passed through Grandfather’s heart, then he would (or might) have died in 1921; if he had died in 1921, then

Father would (or might) never have been conceived in 1922; and so on. And you can see where this chain of counterfactuals is heading: towards a contradiction. It seems as though there will be a chain of counterfactuals – all true according to the story – starting with the counterfactual supposition that Tim fired at a very slightly different time, and leading to a final consequent according to which Tom is a lone gunman on the grassy knoll, with no one who calls himself Tim at some window in the nearby book depository, ready to fire any bullets that might conceivably deflect his own.

But surely it is *not* true, according to the story, that: “if Tim had fired at a very slightly different time, then Tim would (or might) not have been present at that window, and indeed that he would (or might) not have existed anywhere at all”. To furnish a concise abbreviation, we may symbolize anything of the form “ $A \rightarrow \dots$ ” as meaning “If it were the case that A then it would (or at least might) be the case that ...”. Then the counterfactual “if Tim had fired at a very slightly different time, then he would not have existed” entails something of the form “ $A \rightarrow \text{not-A}$ ”. In the penultimate paragraph of his paper in “The paradoxes of time travel”, Lewis clearly affirms that this is not true according to the story:

It is hard to say what is the least revision of Tim’s story to make it true that Tim kills Grandfather, but certainly the contradictory story in which the killing both does and doesn’t occur is not the least revision.

It is pretty obvious that the counterfactual “ $A \rightarrow \text{not-A}$ ” is *not* true according to the story. But I will belabour the point anyway.

The only way Lewis could allow a counterfactual of the form ($A \rightarrow \text{not-A}$) to be true, according to the story, is by supposing that, according to the story, there is no possible world in which A is true. Then, on Lewis’s theory of counterfactuals, ($A \rightarrow \text{not-A}$) will be “vacuously true”. This would mean concluding that, according to the story, it is not possible for Tim (or, for that matter, Tom) to have fired at an even slightly different time. Lewis explicitly denies that it is impossible for Tim or Tom to have acted slightly differently. And indeed, I say it would not really be a timetravel story, properly so-called, if it were the case, according to the story, that no one could consistently have done anything the least bit different from what they did do. So we have to ensure that it is *not* true, according to the story, that “if Tim had fired slightly earlier then he would not have existed”.

Thus, we must ensure that this counterfactual ($A \rightarrow \text{not-A}$) is *not* true, according to the story, even though its antecedent and consequent are linked by a chain of other counterfactuals, each of which is indeed true according to the story. That is, we must have true counterfactuals, according to the story, forming a sequence of the following form:

$(A \rightarrow B)$,
and $(B \rightarrow C)$,
...
and $(Y \rightarrow Z)$,
and $(Z \rightarrow \text{not-A})$;
and yet: not $(A \rightarrow \text{not-A})$.

This chain of counterfactuals could not all be true, according to the story, if the counterfactual conditional were *transitive*. And there are thousands of these chains of counterfactuals that are all true, according to the story. So if counterfactuals were transitive, timetravel stories would be deeply inconsistent. Fortunately for Lewis, however, the counterfactual conditional is not transitive:

$(A \rightarrow B)$ and $(B \rightarrow C)$ does not entail $(A \rightarrow C)$.

According to Lewis's theory of counterfactuals, $(A \rightarrow B)$ is true when all (or most, or at least some) of the least far-fetched ways in which A could be true are ways in which B would be true too. Likewise, $(B \rightarrow C)$ is true when all (or most, or at least some) of the least far-fetched ways in which B is true are ways in which C would be true too. For short: $(A \rightarrow B)$ means "the nearest A-worlds are B-worlds", and $(B \rightarrow C)$ means "the nearest B-worlds are C-worlds".

According to this theory of Lewis's, counterfactuals will not always be transitive. It may be, in some cases, that the nearest A-worlds are B-worlds, but are not C-worlds, so that $(A \rightarrow B)$ is true and $(B \rightarrow C)$ is false. This means that there are worlds where A and B are true but C is false. So *some* B-worlds will be ones in which C is false. But this leaves open the possibility that the *nearest* B-worlds may be significantly less far-fetched than *these* B-worlds, where A and B are true but C is false. So this leaves open the possibility that the *nearest* B-worlds might be C-worlds: and in that case $(B \rightarrow C)$ would be true. So in that case, $(A \rightarrow B)$ and $(B \rightarrow C)$ would both be true, and yet $(A \rightarrow C)$ would be false: and transitivity fails.

This is good news for Lewis. If counterfactuals were transitive, they would be like "upward steps" in the Escher drawing. If a step A is below B, and B below C, then A is below C and C cannot then be below A. Movement "upwards" is transitive. If counterfactuals were transitive, then chains of counterfactuals would lead to contradictions in timetravel stories. So defenders of Lewis can take comfort from the thought that counterfactuals are not transitive: at least under Lewis's theory of counterfactuals. And indeed, even if you were to reject some of the other features of Lewis's theory of counterfactuals, there would still be powerful reasons for thinking that he is at least right in thinking that counterfactuals are not transitive.

Note however that, on Lewis's theory, transitivity may fail, and $(A \rightarrow B)$ and $(B \rightarrow C)$ may be true but $(A \rightarrow C)$ false – *but only* when some of the ways in which B and C could be true are significantly *less far-fetched* than any of the "nearest" ways in which A could be true.

Hence Lewis is not out of the woods yet. Although "ordinary" transitivity fails, there is nevertheless another, somewhat similar, principle that is true according to Lewis's theory of counterfactuals, namely:

$(A \rightarrow B)$ and $((A \& B) \rightarrow C)$ does entail that $(A \rightarrow C)$.

(And indeed, there are good reasons for accepting this quasi-transitivity principle even if you were to disagree with Lewis about other aspects of his theory of counterfactuals.)

Hence, when there is an "ominous" chain of counterfactuals – one which *would* lead to a contradiction if counterfactuals *were* transitive – this chain will also be part of what we might think of as a 'double helix' of counterfactuals, closely echoing Escher's "Ascending and Descending", in which each ascending monk is alongside a corresponding descending monk.

$(A \rightarrow B)$,

and $(B \rightarrow C)$,	and $((A \& B) \rightarrow C)$,
...	
and $(Y \rightarrow Z)$,	and $((X \& Y) \rightarrow Z)$,
and $(Z \rightarrow \text{not-}A)$,	and $((Y \& Z) \rightarrow \text{not-}A)$;
and yet: $\text{not}(A \rightarrow \text{not-}A)$.	

This “double helix” of counterfactuals is logically inconsistent. If Lewis is to maintain the internal consistency of timetravel stories, he will have to maintain that no such double helix of counterfactuals can all be true, according to the story.

Yet I have argued that all the counterfactuals in the left-hand column, in the above double helix, are all true, according to a timetravel story. The only way Lewis can avoid a contradiction is by maintaining that at least one of the counterfactuals in the right-hand column will be false, according to the timetravel story. And, I have argued, there are thousands of such double helixes of counterfactuals whose left-hand columns are all true according to a timetravel story. Lewis can only avoid a contradiction by maintaining that every one of these thousands of double helixes contains a false counterfactual somewhere in the right-hand column, according to the story. Can he plausibly maintain this to be true, according to a timetravel story? There are two reasons why he cannot.

Suppose, in the first place, for the sake of the argument, that we were to grant to Lewis that, for thousands of different double helixes, in each and every case there is at least one false counterfactual in the right-hand column. Granting this would be equivalent to granting that, according to the story, there is in place something equivalent to a Grand Conspiracy. For instance, one way to break the loop is by supposing that hidden forces are at work, an “invisible, intangible hand”, which will ensure that a new-Tim is born, who grows up to become an exact match for the 1920s-Tim. This would turn the story into a version of my story about Tidy Timelords: and that is not a timetravel story. We would have rescued the internal consistency of the story, but only at the cost of its ceasing to be a timetravel story.

Like the Escher Stairs, what appears to be an unbroken loop must actually contain a hidden break, and must actually be a spiral, with a hidden break somewhere along the way: and one way to achieve that is by ensuring that every such apparent double helix contains at least one falsehood in the right-hand column. One place to break the loop is at its last-most point: when new-Tim gets in the contraption labelled “Time Machine” and disappears. That is where my story of Tidy Timelords breaks the apparent loop. However, I have given no reasons why you would need to break the apparent loop at that particular point. I find it easier to imagine a story which breaks the apparent loop at the futuremost point – but, as far as my argument goes, the apparent loop could conceivably be broken at some other point. What I have argued is only that, in every possible world presenting the superficial appearance of an unbroken loop, there must be *some* point at which there is a break *like* the one that comes at the futuremost point in the story of the Tidy Timelords. There must be some such break in *thousands* of apparent loops. This, I say, amounts to a Grand Conspiracy of a kind that would make the story no longer a timetravel story, properly so-called. Because the story *is* intended to be a “timetravel story”, it follows that *according to the story* no such Grand Conspiracy exists. Hence it follows that, implicitly, according to the story such a Grand Conspiracy both does and does not exist.

Timetravel stories, therefore, are essentially inconsistent: if they were not, then they would not be timetravel stories.

Thus, in the first place, the consistency of timetravel stories (*as* “timetravel” stories) cannot be secured, even if you were to suppose that in every one of thousands of apparent double helixes there is a hidden break somewhere in the right-hand column of counterfactuals. Furthermore, in the second place, it would be a mistake to grant Lewis there is a falsehood in each of the right-hand columns of each of the double helixes of counterfactuals that are implicitly true, according to the timetravel story.

Consider what would be required, in order for one of these counterfactuals to be false – even though its counterpart in the left-hand column is true. This requires something of the following form: that $(A \rightarrow B)$ and $(B \rightarrow C)$ are true, but $((A \& B) \rightarrow C)$ is false. And this requires it to be the case that according to the story, some of the ways in which B and C could be true are significantly *less far-fetched* than any of the ways in which A could be true. Yet this is simply not so, according to timetravel stories, for many relevant As, Bs and Cs.

Return then to the story of Tim, and fill in some relevant candidates for A, B and C. Then $(A \rightarrow B)$ might say: “If Tim had fired a little bit sooner, then a bullet would have passed through Grandfather’s heart in 1921”; and $(B \rightarrow C)$ might be “If a bullet had passed through Grandfather’s heart in 1921, then Grandfather would have died in 1921”. These two counterfactuals are true, I have argued, according to the story. Can Lewis plausibly block the threatened *reductio* by “blocking transitivity”?

No. That would require it to be *more far-fetched*, according to the story, to imagine Tim firing a little bit sooner, than to imagine a bullet travelling through Grandfather’s heart and Grandfather dying in 1921. But this is simply not what the story is like. According to the story, Grandfather is unlikely to have had any *closer* shaves, in 1921, than on the occasion when he survived only because two bullets collided in mid-air. The nearest A-worlds, where Tim fired a little bit sooner, are surely – according to the story – among the *nearest of all* the possible worlds where a bullet passed through his heart and Grandfather died in 1921. That could only be so if there were forces at work that make it far-fetched to suppose that Tim fired the least bit sooner, whereas there are less far-fetched worlds in which Tim does not exist at all, and someone else shoots Grandfather in 1921. If this were so, according to the story, then there would have to be some sort of Grand Conspiracy, making the story equivalent to my story of the Tidy Timelords.

Thus, I conclude that Lewis was mistaken: timetravel stories are not internally consistent. In such stories, there are thousands of Penrose-Escher stairways of counterfactuals, which are all true according to the story. In order for the story to be consistent, however, there would have to be at least one deceptive step in each such chain. This would turn all these inconsistent loops into consistent spirals. And then, the story would be structurally equivalent to one in which there are agents or coincidences, which act like the Tidy Timelords who ensure that someone will be manufactured, and shaped, so that his life fits 1920s-Tim’s delusions. Implicitly, according to the story, there both is and is not such a Grand Conspiracy, as Lewis himself acknowledges when he describes the story of “Fred-cum-Sam”.

Hence every timetravel story is essentially inconsistent, like a Reutersvärd-Penrose-Escher picture. When viewed from the right angle, where you are kept ignorant of enough of the relevant facts, it looks superficially very much like a great many possible, and perfectly consistent, more fully-articulated stories. But if it were not inconsistent, it would not be a timetravel story.

References

Arntzenius, F.,

“Time travel: double your fun”, *Philosophical Compass*, 1 (2006), pp.599-616.

Arntzenius, F. And T. Maudlin,

“Time travel and modern physics”, pp.169-200 in C. Callender (ed.), *Time, Reality and Experience*, Cambridge University Press, Cambridge UK, 2002.

Cresswell, M.J.,

“A highly impossible scene”, pp.72-78 in *Meaning, Use and Interpretation of Language*, ed. R. Bäuerle, Ch. Schwarze and A. von Stechow, De Gruyter, Berlin, 1983.

Cresswell, M.J.,

Structured Meanings, MIT Press, Cambridge MA and London UK, 1985.

Dowe, P.,

“Constraints on data in worlds with closed time-like curves”, *Philosophy of Science*, 74 (2007), pp.724-735.

Dowe, P.,

“Does Lewis’ semantics for counterfactuals allow for time travel?”, from his website, University of Queensland.

Dummett, Michael,

“Bringing about the past”, *Philosophical Review* 73 (1964), pp.338-359.

Escher, M.C.,

“Ascending and Descending” (1960), *The Graphic Work of M.C. Escher*, Oldborne, London UK, 1967.

Goddu, G.C.,

“Time travel and changing the past: (or how to kill yourself and live to tell the tale)”, *Ratio*, 16 (2003), pp.16-32.

Hanley, Richard,

The Metaphysics of Star Trek, Basic Books, New York NY, 1997.

Hanley, Richard,

“Miracles and wonders: science fiction as epistemology”, in *From Time Travel to Superintelligence*, ed. Susan Schneider, Wiley-Blackwell, Chichester UK, 2009.

Heinlein, R.A.,

“By his bootstraps”, in R.A. Heinlein, *The Menace from Earth*, Gnome Press, Hicksville NY, 1959.

Kiourti, Ira,

“Killing baby Suzy”, *Philosophical Studies*, 139 (2008), pp.343-352.

Lewis, D.K.,

“Paradoxes of time travel”, *American Philosophical Quarterly* 13 (1976), pp.145-152.

Lewis, D.K.,

Counterfactuals, Basil Blackwell, Oxford UK, 1973.

Lewis, D.K.,

“Causation”, *Journal of Philosophy* 70 (1973), pp.556-567.

Lewis, D.K.,

“Survival and identity”, in Amélie O. Rorty, ed., *The Identities of Persons*, University of California Press, Los Angeles CA, 1976.

Lewis, D.K.,

“Truth in fiction”, *American Philosophical Quarterly* 15 (1978), pp.37-46.

Lewis, D.K.,

“Counterfactual dependence and time’s arrow”, *Noûs* 13 (1979), pp.455-476.

Mellor, D.H.,

Real Time, Cambridge University Press, Cambridge UK, 1981.

Mellor, D.H.,

Real Time II, Routledge, London UK, 1998.

Penrose, Lionel and Roger Penrose,

“Impossible objects: a special type of illusion”, *British Journal of Psychology* 49 (1958), pp.31-33.

Reutersvärd, Oscar: see Seckel (2004).

Savitt, Steve (ed.),

Time's Arrows Today: Recent Physical and Philosophical Work on the Direction of Time, Cambridge University Press, Cambridge UK, 1995.

Seckel, Al,

Masters of Deception: Escher, Dali and the artists of optical illusion, Sterling, New York NY, 2004.

Sider, T.,

"Time travel, coincidences and counterfactuals", *Philosophical Studies*, 110 (2002), pp.115-138.

Smith, N.J.J.,

"Bananas enough for time travel?", *British Journal for the Philosophy of Science*, 48 (1997), pp.363-389.

Vihvelin, K.,

"What time travellers cannot do", *Philosophical Studies*, 81 (1996), pp.315-330.

Vranas, P.,

"What time travellers may be able to do", *Philosophical Studies*, ?? pp.115-121.

Williams, Donald C.,

"The myth of passage", *Journal of Philosophy*, 48 (1951), pp.457-452; reprinted in Donald C. Williams, *Principles of Empirical Realism*, Springfield, Illinois, 1966.