Chapter 14

Lewis on time travel

Jonathan Bennett

In H. Beebee & A.R.J. Fisher (eds), *Perspectives on the Philosophy of David K. Lewis*, Oxford: OUP 2022, 260–8.

Abstract

This chapter, originally written for a symposium commemorating David Lewis in 2002, considers the possibilities for telling stories concerning time travel to the past that do not involve the time traveller coming to see themselves (whether rightly or wrongly) as unable to perform the kinds of action they are normally able to perform, the doing of which would prevent something that did, in fact, occur from occurring—as in the 'grandfather paradox'. Various options are considered—for example, the time-travelling somehow either causing selective amnesia concerning what happens at the past time or causing the time traveller not to want to change the past; or a ban on travelling back to a time about which the time traveller has information

I hoped passionately that David Lewis would—against the odds—live a long life. I did not want the world to lose him early; I didn't want Steffi to lose him; and, selfishly, I didn't want the last part of my life to be without David. Well, here we are.[1]

When I was invited to contribute to this symposium, I could not say no. But, because of some defect in my wiring, I fail when I try to produce philosophy to fulfil a promise or a commission; so I had to rely on work I had already done but not published. This led me to a paper I wrote in 1977, commenting on David's 'The Paradoxes of Time Travel' (Lewis 1976) that had appeared in the APQ the year before. I sent this to him at the time, and his lengthy and typically generous response left me feeling disinclined to work any more on this, or to publish what I had done; but it left me also feeling that there was something here worth airing. Now it has forced its way to front and centre, as my best chance of contributing to this afternoon's symposium—in David's company, almost. I'll be talking for half an hour.

---

[1] This chapter was read as part of a symposium on David Lewis at the American Philosophical Association, Central Division, 26 April 2002.

Let us consider a time traveller named Wells who, in the Belgian town of Plancenoit at this very moment, travels nowhere in space but nearly 187 years into his past, thus 'landing' smack in the middle of the battle of Waterloo.

The abstract structure of such tales has been well described by David Lewis in the paper that I mentioned, and I need only sketch a reminder. If Wells is to be later at an earlier time, there must be two time-series, that of external time and that of Wells's personal time. His presence at the battle occurs in 1815 (external time), because that just is the date of the battle, but so far as his personal time is concerned—his bodily processes, his sense of time passing, and above all the inter-relations of his memories—the battle occurs after this moment in 2002. The battle externally predates now by nearly 187 years, but it, as we might say, Wells personally postdates now by however much of his personal time Wells uses for the journey back. All this was superbly laid out by David twenty-six years ago.

Travel into the past involves someone's person-stages' being ordered one way in external time and differently in his personal time. Not so with travel into the future, which is something we do constantly. Even in accelerated travel into the future, where the traveller makes a personal-time leap to a later external time, there is no difference between the relevant external and personal orderings of his person-stages: the phenomenon is metrical, not topological. My topic is the idea of 'travel into the past'; that is, to a time earlier in external time than the start of the journey.

The concept of a person is interesting to us only if a person is an item with memories of its personal past; and a given epistemic state counts as a memory of another state only if it was caused by it. So any temporal stage of a person must be causally descended from stages of him that are earlier in his personal time. So if the person-stage that announces 'I am going back to 1815' and the one that arrives on the battlefield are stages of a single person, then the 1815 one must be causally descended from the 2002 one. So time travel essentially involves backwards causation.

In his paper David contended that time travel is conceptually possible; but he did not complete his case for this. He had already published the first version of his counterfactual analysis of singular causation statements, according to which 'e1 caused e2' means that if e1 had not occurred e2 would not have occurred; but he had not worked out how to develop this into an account of backward causation. To do so, he needed something that he didn't ever have, so far as I know, namely a strong, coherent theory of temporally backward counterfactuals. I have a lovely theory about those (Bennett 1984); I lifted it straight out of

David's account of forward counterfactuals; but I see no prospect of its helping to make sense of temporally backward causation.

Anyway, I shall not discuss whether time travel is possible. My interest is in David's defence of the notion of time travel against the charge—which used to be levelled quite often—that it is impossible because it essentially involves 'paradoxes', called 'chronoclasms' in some of the fantasy literature: events in which the time traveller causes something to happen which did not happen. The dramatic examples involve the time traveller's killing his grandmother when she was still a child, or intervening to prevent the invention of time travel; but milder examples will serve as well. A story which has me coming back to this moment from some time in 2005 and not raising my hand is just as 'paradoxical', just as contradictory, as one in which I go back to 1666 and kill one of my ancestors while he is still a baby.

David argued that such contradictions can be avoided, without this being brought about by any kind of censor or chaperone. I shall not firmly disagree with that either; but I want to push it around a little, to see what else we can learn from it.

On the face of it, the threat of paradox or contradiction applies to any kind of backward causation, not only time travel. If causes can ever post-date their effects (so the threat runs), then a backwards causal chain might cause an earlier event which then, through a forward chain, forestalled the event that started the whole chain going. Schematically, A at t2 causes B at t1 which prevents A at t2.

But, really, backwards causation is not the crux either. When describing any causal setup, we need to avoid admitting triples of the form:

A is causally possible
A is causally sufficient for B
B is causally sufficient for the non-occurrence of A.

Obviously, all such triples must be avoided; but in what sense are we threatened with them?

Well, consider the case of Tim, whom David described as journeying back in time and trying to kill his own grandfather in Grandpa's infancy. An acceptable story about Tim must not say that he succeeds in his attempt, thus doing something which implies that he never existed; and the storyteller has to abide by this requirement. Accordingly, in David's version, Tim fails in his assassination attempt. Where's the problem?

'Well', it may be said, 'how does Tim fail? Perhaps Lewis has a problem about saying in detail what it is, according to his story, that keeps out all instances of the forbidden triple'. Well, why does Tim fail? 'For some commonplace reason', David wrote. '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' (1976: 150). Whatever explains his failure, David said, it doesn't imply that something in the story has been guarding Tim against committing a 'chronoclasm'; nor does it imply any inability on Tim's part. Tim may have motive, opportunity, means, skill and so on, so that—as David put it—'by any ordinary standards of ability, Tim can kill grandfather'. It is true that the storyteller must not say that Tim kills grandfather, but that 'must not' is observed just so long as in the story Tim does not kill grandfather.

David was right to distinguish constraints on the dramatist from constraints on the dramatis personae. Forget time travel and look for a moment at ordinary history. Since the French did lose at Waterloo, a truthful history of the battle must not report that someone enabled them to win. But still Marshall Grouchy could have enabled them to win, and would have done so if he had thought strategically about glory for France rather than tactically about glory for himself.

Still, let us pause for a moment. When David says that Tim just fails—some little thing goes wrong, as little things often do—that may seem to be the end of the matter; but before we close the book and sit back, let us note that the case described was notable for its epistemic poverty. Tim enters the situation with virtually no relevant knowledge except that somehow his grandfather will survive the attempt on his life: there is one item of knowledge, one attempt to falsify it, and one failure.

One's knowledge of the public past can be vastly fuller than that: there is no limit to the amount of detail of a past event which can in principle be known now, through memory, testimony, films and so on. So let us now consider an epistemically rich case of time travel. In this, someone I'll call Henry travels to a spacetime locus containing an event about which he knows a great deal before reaching it in his time machine—specifically, a great deal about how he behaves in it. Let us suppose that the event in question is a party.

Henry might have attended the party himself at a personally earlier stage in his life. (David was illuminating on the subject of the synchronous existence of distinct stages of a single person (1976: 147).) Or he may be relying on his memories not of the party itself but rather of copious testimony, films, books, etc. in which thousands of details of the party are truthfully recorded. Either way, he embarks on the party with a detailed picture of how it is

going to go: the spilled jug of punch, the theft of the banjo, the girl who is excluded from the poker game, and so on.

Now, suppose Henry tries to falsify this picture: he tries and fails to move the punch jug nearer to the centre of the table, he tries and fails to keep an eye on the banjo, he tries and fails to steer the girl away from the poker game. Indeed, his initial information about the party could be so detailed that we must suppose that he tries and fails to move one foot forward, to bend a finger, to draw a deep breath: the information that at a certain past moment a certain person did not draw a breath or move a foot is the kind of thing that can be reliably established. Before embarking on his time journey Henry had access to detailed records of his conduct at the party; now he arrives at the party, with the details of his future conduct spread out in his mind; and he finds himself running this course—he does not depart from the pre-pictured path, try as he may.

If this happens often enough, every time he makes such a visit to a past time, Henry will become convinced that he cannot move a foot, draw a breath, lift a glass, and so on. Of course, failure when you try is not proof of lack of ability; but if someone fails every time he tries to do things of a certain sort (the sort being defined by their relation to what he knows through memory and other records), he will naturally think that in some way he is debarred, prevented, from doing such things. And when I first wrote on this topic, I was on Henry's side about this. At that point David Lewis dug his heels in. He wrote to me:

> I think it is possible that *any amount* of failure, and even any amount of failure unified by a pattern of frustration in trying to depart from the scenario, is explained by miscellaneous bad luck and nothing else. . . . So a thorough story about time travel *does not* need to include something special, and unitary, as a defence against 'paradoxes'. Of course a thorough story can include a special, unitary defence—what I called a 'chaperone to protect the past'… but I think the most interesting sort of thorough story is one in which the traveler's repeated attempts to depart from the scenario are frustrated by nothing but miscellaneous bad luck.
>
> It's possible that you don't mean to disagree with what I've just said, but only mean to claim something I'm inclined to agree with. … Perhaps what you mean is that, after a certain amount of patterned frustration, *Henry* ought to believe that his failures have a unified explanation that implies lack of ability, and that it would be irrational for Henry to believe that he'd been stopped by nothing but a run of bad luck. … If so, I'm inclined to think you're right. But what of it?

It is possible for someone to have the misfortune that, although A is true, he has evidence that would make it utterly irrational for him to believe A. …

Part of your point, I think, is that *everybody* in the story who knows enough about Henry would rationally believe in his lack of ability. … But that doesn't change things. Just as it's possible for one person to be in a position in which it's irrational for him to believe what's true, so it's possible for everyone to be in this position.[2]

A verificationist, indeed anyone ever so slightly tinged with verificationism, would presumably doubt that it's really possible for everyone to be permanently, irremediably, rationally wrong. From verificationist premises there might indeed be an objection to the possibility of time travel without special and unitary defences.

David went on to say that he didn't think I could be accused of being tinged with verificationism. I think I could, though that's something else I am shy about. In any case, I have come to wonder whether there is anything much to argue about here. When I implied that we would be compelled to conclude that Henry couldn't perform any of the movements that he went to the party intending to make, I hope I didn't mean that from a pattern of total failures we could infer that something further was the case blocking Henry from succeeding. All I should have meant was that the right way for us to describe the situation is in the language of inability: given so many failures and no successes, we ought to say that Henry cannot carry out his chronoclastic intentions. If I hold to that, and David denies it, that is a real disagreement—but it's a disagreement about meanings, not about metaphysics—not about time travel.

Still, I think it was worthwhile to consider epistemically rich time travel. It is, I hope, of at least some interest to realize that when time travel stories are told in a rich and robust manner, the upshot is—for the traveller—an experience that is indistinguishable from one of imprisonment in a scenario. Whatever the metaphysics of it, it will in every way appear to the traveller as though he is present on the scene not as an active agent but as a puppet.

I am also interested in the question of how one might constrain a time travel story so that the time traveller was protected from having any puppet-like experiences. Can we devise a kind of time travel that would not involve such things as a healthy traveller who does not bend his finger when he tries to? There seem to be just three ways of trying to do this.

---

[2] Letter 240. To Jonathan Bennett, 29 March 1977, in Beebee and Fisher 2020: 471–2.

We could stipulate that time travel causes amnesia. Henry embarks on his time journey with a rich stock of knowledge about the party, but by the time he 'lands' he has forgotten it all.

If he suffers only from partial amnesia, forgetting his knowledge about the party but not about anything else, the story will become arbitrary and ad hoc. For example, if Henry spent most of his time in 2001 researching this party, his knowledge about it will be densely interwoven with his knowledge about his own life in 2001; and any selection principle yielding the required kind of partial amnesia will have to be a thoroughly gerrymandered affair.

For that reason, and another that I'll mention shortly, amnesia can work for us only if it is total. We are to suppose, then, that an effect of Henry's time travel is that while he is at the party he has no memories of his personally earlier self (and, we shall have to suppose, time machines won't transport photographs, etc.). This will enable us as storytellers to avoid 'paradox' without constraining Henry. There are many little movements that he won't make at the party, but in our present story he doesn't even try to make them, so that his not making them does not give him any sense of being in a behavioural straightjacket.

But the price for this is high. In this version, Henry's experience of the party is not that of a time traveller but rather that of a puzzled amnesia victim. It is only after he has returned to 2002, if indeed he does so, that this personal experience of the party is linked with his experience of his 2002 life in a single consciousness in which both are remembered. And it won't happen even then unless the return to 2002—whether by leaping forward or just by patiently waiting—cures the amnesia.

Also, a victim of total amnesia is a poor specimen of a person. And if we suppose that time travel deprives Henry of his memories and equips him with a new, false set of seeming memories, this makes him more like a person but at the price of undermining his claim to be the same person.

This is a disappointing version of the time travel idea. It does after all involve a special trouble-avoiding device; it does not permit any characteristically time traveller's experience; and it permits time traveller's memories only if provision is made for the later recovery of the lost memories.

A second way of systematically excluding the experience of seeming impotence is to leave Henry with his epistemic store, but to suppose that his knowledge and his time travel have the effect of sapping or deflecting his will, so that he does not try to act against the given scenario, even if he had initially intended to do so. Even if Henry had at the outset been

consumed by the ambition to try to produce a chronoclasm, we are now to suppose that he arrives at the party with no such desire.

In one sub-version of this story, time travel causes an unshakable mood of passive, resigned acceptance: Henry confronts the party with no desire except to play the role that his memory spreads out before him. This makes him a poor specimen of a person: his behaviour is systematically governed by an *idée fixe*, like someone acting under post-hypnotic suggestion.

We might try to improve on this by equipping him with a lively new set of interests, concerns, and desires, which manifest themselves at the party and deflect him from his former desire to aim at 'paradox'. But this again increases his personhood at the cost of weakening somewhat his claim to be Henry.

There, then, are the first two ways of ruling out attempts at chronoclasm. In one we deprive the traveller of the knowledge that is needed to shape such an attempt; in the second we deprive him of the will to make the attempt. Neither of these outright implies that the item at the party is not a person, or not Henry; but each of them somewhat erodes his status as a person, or as Henry. Either way, we have a time travel story that is less attractively robust—because its protagonist is less robust—than we had hoped.

I agree with David Lewis that 'the time traveller's world would be a most strange one'. I am merely adding that the time traveller himself must be 'strange'. A thorough account of time travel by people can avoid inconsistency only if it specifies that time travel causes loss of memory, loss of motivation, or seeming impotence of the will.

I said there were three possible remedies, and have described only two. The third is to adopt a version of the time travel idea in which no one can terminate a time journey at any spacetime locus about which he knows a great deal. Henry's mind contains a detailed representation of how the party is going to run, and the locus containing the start of the party also bears the marks of how the party will run; so Henry's mind is related to that locus in a characteristic manner which—we are now supposing—renders it causally impossible for him to 'land' there. Just as super-tankers cannot come to land except in deep waters, so time machines cannot 'land' except in epistemic shallows.

If time travel had that feature, any time traveller could retain all his memories, and thus be aware that he was indeed on a time trip; and yet he could also act as freely as we can: he could consider and deliberate and plan how to act, and his deliberations could relate to his decisions, and his decisions to his actions, just as ours do. Because of his initial ignorance of the spacetime locus he is visiting, we are under no pressure to allow that he might try to bend

his finger in a 'paradox'-producing manner, and so we are spared from having to allow that he tries to bend his finger and fails. In short, he need not have any experience of impotence.

A story of this kind—if the time traveller is not constrained or depleted in any way—will have to allow that he sometimes behaves anachronistically. But that's alright. In a world with this kind of time travel, there will be some historical records of people suddenly turning up and behaving in ways that at the time are merely puzzling but are later realized to be anachronistic. For example, the diary of a seventeenth-century bureaucrat might report the mad talk of someone who rambled on and on about 'power', but did not connect it with armies and navies but rather with what he called 'electricity' and 'atomic fission' and 'fossil fuels'. There is no special problem in that.

But there is a difficulty in another part of this field. If the time traveller is not epistemically depleted, he will 'land' at the earlier time with knowledge of events occurring in the public future. So he can at 1665 (say) leave an account of some events occurring in 1865 (say), including a minutely detailed true record of how Queen Victoria (say) behaves on some particular public occasion. The Queen can know of this record, recognize herself in it, and indignantly try to thwart it. She will find that however hard she tries she does not succeed. With enough of this, she will have the experience of impotence, of imprisonment in a scenario—a poisoned bequest from the time traveller.

In short, any difficulty or constraint or awkwardness we encounter in the idea of a fully equipped time traveller, we also encounter in the idea of someone who has received reports from a time traveller. This applies not only for the version in which the traveller can land only in epistemic shallows, but also—as an added difficulty—in each of the other two versions as well. I leave it to you to think about the devices a storyteller might adopt to get around this.

I conclude that a robust, unevasive story about time travel by people must constrain or deplete one or more of the people involve, or else give them the experience of being in a straightjacket, imprisoned by a scenario.

References

Beebee, Helen and A.R.J. Fisher, eds. 2020. *Philosophical Letters of David K. Lewis: Volume 1: Causation, Modality, Ontology*. Oxford: Oxford University Press.

Bennett, Jonathan. 1984. Counterfactuals and Temporal Direction. *The Philosophical Review* 93(1): 57–91.

Lewis, David. 1976. The Paradoxes of Time Travel. *American Philosophical Quarterly* 13(20: 145–52.