The problems of backward time travel

Nicholas J.J. Smith

The so-called paradoxes of time travel have played a significant role in both the physics and philosophy literatures – but how much force do these alleged paradoxes really have?

Most people, when they think of backward time travel, think of a special machine, such as Dr Who's TARDIS, which disappears from one place and time and reappears at an earlier time, in the same or a different place. This, however, is not the sort of scenario that occupies those physicists and philosophers who work on time travel. In the 1940s, Kurt Gödel discovered models of the Einstein field equations in which there exist closed timelike curves (CTCs)[1]. (In string theory, unlike in general relativity, Gödel universes need not contain CTCs[2].) Gödel writes that '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' (Ref. 3, p. 560). Unlike Dr Who, Gödel's time traveller is always oriented towards the local future, and it is not special properties of the vehicle (which is simply a rocket or spaceship) which enable them to travel in time, but rather certain properties of space-time in their universe – viz., the existence within it of CTCs.

Since the work of Gödel, other models of the general theory of relativity containing CTCs have been found. [In fact, such a solution had already been found in 1937, but it was not realized at the time that this model contains CTCs (Ref. 4, p. 21).] These models have not, on the whole, received a great deal of attention from physicists, but in the last ten years there has been considerable interest in the question of whether it is possible to *manufacture* CTCs in universes which (unlike the Gödel universe) do not already contain them. Morris *et al.*[5], Gott[6], Ori[7], Alcubierre[8] and Everett[9], have proposed, respectively, four different sorts of time machine, and each proposal has met with numerous objections. No one has yet produced a practicable scheme for backward time travel in our universe, but nor has it been established that such a scheme cannot be produced: this is an open and exciting question in physics.

One striking aspect of the physics literature on time travel is the weight given to relatively half-baked considerations to the effect that there is some sort of logical or conceptual incoherence in the very idea of backward time travel. Thus, for example, Stephen Hawking writes (Ref. 10, p. 604): 'By travelling in a space ship on one of these [CTCs], one could travel into one's past. This would seem to give rise to all sorts of logical problems, if you were able to change history. For example, what would happen if you killed your parents before you were born. It might be that one could avoid such paradoxes by some modification of the concept of free will. But this will not be necessary if what I call the *chronology protection conjecture* is correct: *The laws of physics prevent closed timelike curves from appearing*.' As is the case here, the considerations in question are usually cited as motivation for producing more rigorous objections to backward time travel, rather than presented as conclusive objections in their own right – but do they really carry *any* weight at all?

That is the question which has been the principal focus of the philosophical literature on time travel, and philosophers have made good progress towards answering it. Many naïve conceptions of time travel and what it would involve have been replaced with more sophisticated conceptions, and more and more arguments, once thought to constitute knock-down objections to the very idea of time travel, have been shown to have no force. These developments are outlined below.

It is often thought that backward time travel would enable one to go back and kill the infamous figures of the past, prevent the plagues of the Middle Ages, and in general change history in countless ways. This idea is often extended to an objection: backward time travel would indeed enable one to change the past; but to change the past is to make it the case that something happened which did not in fact happen, or vice versa – a contradiction; hence backward time travel is impossible. It has been clearly shown, however, that what backward time travel must involve is not changing the past (which is indeed impossible), but affecting the past[11–19]. Thus, for example, the time traveller's setting of various controls at one time (partially) causes their arrival at an earlier time – but, by the same token, a complete chronicle of their destination time tells of their arrival *before* they depart. Their actions have an effect on earlier events, but they do not change those events– they do not cause those events to be one way, while prior to their actions those events were a different way. If a time traveller is going to travel to some past time, then they *have already been there*; if they are going to save a life when they get there, then they have *already* done so. To fail to see this point is to commit the *second-time-around fallacy* – to imagine that backward time travel gives one a second go at the very same events that constitute one's past[18]. [One approach to backward time travel, associated with the name of David Deutsch, involves the idea of time travellers journeying to parallel universes[20,21]. Such a time traveller might emerge from a time machine in 1920 (local time) and kill (the local version of) Adolf Hitler. There is no second-time-around fallacy here: no change is made to the events of 1920 or thereafter in the universe the time traveller leaves behind.]

A second traditional objection to the idea of time travel runs as follows. A time traveller is born in 1964; in 2014 they depart on a journey through time; the journey lasts one year; they arrive in 1984. Thus the time traveller traverses 30 years in one year, and is 51 years old 20 years after their birth. These are straightforward contradictions; therefore backward time travel is impossible. Note that this objection applies just as well to forward time travel: a time traveller is born in 1964; in 1984 they depart; their journey lasts one year, and they arrive in 2014; thus the time traveller traverses 30 years in one year, and is 21 years old 50 years after her birth. These are straightforward contradictions; therefore forward time travel is impossible. Now there must be something wrong here – for forward time travel has actually occurred! The fundamental principle of Einstein's special theory of relativity is that all observers measure the same value for the speed of light in a vacuum – and from this it follows that observers moving relative to one another measure different temporal intervals between the same events. In particular, a clock carried aboard a fast rocket runs slow relative to a similar clock on earth – and the faster the rocket travels, the slower the clock runs. This *time dilation effect* implies that any traveller can, simply by travelling fast enough, become a time traveller. A rocket pilot, coming back to earth after a very high-speed journey, finds that while only a few years have elapsed for them, many years have elapsed on earth. The faster they travel, the further in the future they arrive back – and even aboard existing

Nicholas J.J. Smith, BA (Sydney), MA (Princeton)

Is at the Department of Philosophy, Princeton University, USA.

 0160-9327/98/$19.00 – see front matter PII: SO160-9327(98)01154-5

aircraft atomic clocks have measured small time dilations.

The objection to time travel currently under discussion dissolves when we notice that the interval of time traversed by the time traveller and the duration of their journey – which, apparently paradoxically, have different magnitudes – are measured with respect to different frames of reference[14,17,22]: the former with respect to a frame of reference in which an earth-bound observer of the time machine is stationary; the latter with respect to a frame of reference in which the time traveller is stationary. The former is given in the observer's proper time, the latter in the time traveller's proper time: there is, therefore, no reason why they should coincide. The same goes for the discrepancy between the time elapsed since the time traveller's birth and their age upon arrival: the former is measured with respect to the observer's proper time; the latter with respect to the time traveller's proper time.

A third traditional objection to time travel is this: the time traveller has grey hair when they begin their journey in 2014 and when they end it in 1984; but in 1984 they had black hair; therefore the same individual both does and does not have black hair in 1984; this violates Leibniz's Law (the principle that if one thing is identical with another then any property possessed by the one is also possessed by the other); therefore time travel is impossible. This paradox, like the previous one, dissolves when we carefully sort out with respect to which reference frames different attributions of properties at times are intended[14,17]. If 'X has black hair in 1984' means 'X has black hair in 1984 (earth-proper-time) at *some* X-proper-time', then the time traveller has this property at all times. If, on the other hand, 'X has black hair in 1984' means 'X has black hair at X-proper-time t', then either the time traveller always has this property (if, for example, $t = 20$ X-proper-years after X's birth), or always lacks it (if, for example, $t = 51$ X-proper-years after X's birth).

The objections to backward time travel discussed so far are widely agreed to carry little weight. The next objection, on the other hand, continues to be pressed – although an increasing number of philosophers argue that it, too, should be laid to rest[4,14,18,22–24]. If backward time travel were possible, the objection runs, there would be nothing to stop a person travelling back in time and killing their grandfather before he had a chance to produce offspring – but then they themselves would not be born, and this is a contradiction: the time traveller is both born (and grows up to make a time trip) and not born. This is the *grandfather paradox* (a variant, in which the time traveller kills their younger self rather than their grandfather, is called the *auto-infanticide paradox*). So, the argument goes, if backward time travel were possible there would be nothing to stop contradictions being true – hence backward time travel is impossible. (One often hears the objection put this way: if backward time travel were possible, then a time traveller could go back and kill their younger self; if the younger self died, however, then they would not grow up to become the time traveller; hence the time traveller would not exist to go back and kill the younger self; hence the younger self would not die; but then the younger self would grow up to become the time traveller, and there would be nothing to stop the time traveller from going back and killing the younger self, and so on; this is a vicious circle; therefore backward time travel is impossible. Once the idea of changing the past is rejected, however, the first move in this progression can be seen to involve a contradiction, and the vicious circle does not get started.)

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. As David Lewis showed in the 1970s, however, such chaperones are unnecessary[22]: no strange devices are required to stop the time traveller killing their younger self; rather, they fail 'for some commonplace reason' – their gun might jam, for example, a noise might distract them, or they might slip on a banana peel. Nothing more than such ordinary occurrences is required to stop the time traveller killing their 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.

So much for the grandfather paradox in its traditional form – but Paul Horwich has responded with a new form of the objection[17]. Our time traveller sets off to kill their younger self. No contradiction ensues – this is prevented by a run of coincidences (one or more per murder attempt) which sees the trigger fall off the time traveller's gun, a gust of wind push their bullet off course, and so on. But now, what about this run of coincidences? Horwich argues that such runs are extremely improbable. But, he argues, given travel to the local past, such runs are required to prevent contradictions. So local backward time travel, while not entailing contradictions, entails improbabilities – and hence, although not *impossible*, is extremely *unlikely* to occur.

Horwich's argument is open to criticism on two fronts. True, *if* a time traveller goes back and makes repeated attempts to kill their younger self, a string of coincidences will ensue. But why would a time traveller do such a thing? It makes no sense at all to say that someone is actively trying to do something, and yet feels certain that she will fail. But of course, if the time traveller is trying to kill their younger self, then they *will* fail – so they must have failed to realize this, say because they have fallen for the second-time-around fallacy. Backward time travel alone, then, does not generate improbable strings of coincidences here: we need to suppose, in addition, that time travellers fail to think clearly (that is, fail to avoid the second-time-around fallacy); but a systematic correlation between time travelling intentions and failure to think clearly is itself an improbable coincidence. Nor does it make a difference if we switch to a scenario in which the time traveller knows that they will not succeed in killing their younger self, but fires guns and throws grenades anyway, just out of curiosity: they want to see all the fantastically improbable ways in which their younger self survives. This scenario makes little sense unless the time traveller does *not* already know, prior to firing guns and so on, how their younger self will be saved – and yet it is highly improbable that the time traveller should have *forgotten* the most exciting and dangerous events of their youth (namely, their fantastically improbable escapes from a series of murder attempts by an apparently demented person – their older self). In general, the first objection to Horwich is that backward time travel, in itself, does not entail improbable strings of coincidences. Rather, every argument which purports to derive such coincidences as output, given backward time travel as input, *also* uses as input – in addition to backward time travel itself – occurrences which are *themselves* highly improbable.

Of course it *might* happen that a time traveller becomes convinced that they can kill their younger self, and possessed of a desperate desire to do so (people just *do* sometimes form irrational beliefs and desires for no good reason). Should this happen, the smallish coincidence involved would be amplified, leading to an arbitrarily long run of slips on banana peels and so on as the determined time traveller's auto-infanticide attempts are foiled. Backward time travel, then, does not, as Horwich thinks, *entail* long strings of coincidences, but it does open the way to them: the combination of backward time travel and an event of *mildly* low probability (the development of irrational beliefs and desires on the part of some person) generates a string of events of *very* low probability.

This, however, is not a problem – and this is the second flaw in Horwich's argument. The reason Horwich thinks that auto-infanticide-foiling coincidence-strings are improbable is that they involve 'inverse forks': there is a systematic correlation between intentions on the part of the time traveller to kill their younger self, and the occurrence of events (slips on banana peels and so on) which foil their murder attempts (in other words, every time the time traveller tries to kill their younger self, someone drops a banana peel in front of them, or something else happens which foils the murder attempt) – and these correlated events are associated only with a *subsequent* event (the time traveller's departure from the future). The reason, in turn, that Horwich thinks that inverse forks are improbable is that we do not see them: whenever two types of event are associated with one another (for example, heavy drinking and bouts of depression), there is either a chain of events between them (for example, drinking causes depression, or depression leads one to drink), or else an *earlier* (*not* later) event linked to both by two chains of events (for example, there is a gene which causes both alcoholism and depression). True enough: we do not see inverse forks – but it does not follow that we *shall not* see them. Hence even if backward time travel *necessarily* involved inverse forks, we could not conclude, from the mere fact that we see no inverse forks, that backward time travel will occur at most rarely. To argue otherwise would be like arguing that one has never seen humans fly, and therefore will do so at most rarely – even as the Wright brothers set up in the neighbouring field. (I have presented these two criticisms of Horwich in full elsewhere[18].)

Finally, a popular objection: backward time travel will never occur – for if it were going to, we would *already* have encountered the time travellers involved, and we have done no such thing. Although this argument has been advanced by physicists and philosophers, in fact it has little force. Supposing we have indeed met no time travellers, there might be various explanations of this fact. The longer it takes for time machines to be invented, the more our period of history will appear to future time travellers as four o'clock on 4 March 762 appears to us – that is, fairly low down on the must-visit list. Of course, if there were a great many time travellers, we might expect some of them to end up around here about now – but what if backward time travel is prohibitively expensive, and only a few people ever travel back in time? In any case, can we be so sure that we have met no time travellers? Perhaps we do not know what we are looking for. Consider an isolated society living in a remote part of the world. Some of the members of this society are engaged in a long-running debate concerning the possibility of human flight. Were a jumbo jet to pass overhead, would the debaters necessarily realize that it contained flying humans? The answer to their question might have been staring them in the face for years, without their noticing.

In conclusion, it may well be that backward time travel is a practical impossibility in our universe – this is a question for physicists. No one, however, should feel driven to prove that backward time travel is a practical impossibility because they feel that there is something incoherent or paradoxical about the very *idea* of visiting the past. Whether or not there is incoherence or paradox here is a philosophical question, and one which philosophers are increasingly inclined to agree should be answered in the negative.

References

1 Gödel, K. (1949) *Rev. Mod. Phys.* 21, 447–450

2 Barrow, J.D. and Dabrowski, M.P. (1998) *Phys. Rev. D*

3 Gödel, K. (1949) in *Albert Einstein: Philosopher-Scientist*, Vol. II (Schilpp, P.A., ed.) (Library of Living Philosophers, Vol. VII), 555–562

4 Earman, J. (1995) *Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Spacetimes* Oxford University Press

5 Morris, M.S., Thorne, K.S. and Yurtsever, U. (1988) *Phys. Rev. Lett.* 61, 1446–1449

6 Gott, J.R. (1991) *Phys. Rev. Lett.* 66, 1126–1129

7 Ori, A. (1993) *Phys. Rev. Lett.* 71, 2517–2520

8 Alcubierre, M. (1994) *Classic. Quantum Gravity* 11(5), L73–L77

9 Everett, A.E. (1996) *Phys. Rev. D* 53, 7365–7368

10 Hawking, S.W. (1992) *Phys. Rev. D* 46, 603–611

11 Harrison, J. (1971) *Proc. Aristotelian Soc.* Suppl. 45, 1–24

12 Brier, B. (1973) *South. J. Philos.* 11, 359–364

13 Dwyer, L. (1975) *Philos. Stud.* 27, 341–350

14 Horwich, P. (1975) *J. Philos.* 72, 432–444

15 Dwyer, L. (1977) *South. J. Philos.* 15, 383–385

16 Fulmer, G. (1980) *Southwest. J. Philos.* 11, 151–156

17 Horwich, P. (1987) *Asymmetries in Time: Problems in the Philosophy of Science*, MIT Press

18 Smith, N.J.J. (1997) *Br. J. Philos. Sci.* 48, 363–389

19 Barrow, J.D. (1998) *Impossibility: The Limits of Science and the Science of Limits*, Oxford University Press

20 Deutsch, D. (1991) *Phys. Rev. D* 44, 3197–3217

21 Deutsch, D. and Lockwood, M. (1994) *Sci. Am.* 270, 50–56

22 Lewis, D.K. (1976) *Am. Philos. Q.* 13, 145–152

23 Brown, B. (1992) *Can. J. Philos.* 22, 429–44

24 Riggs, P.J. (1997) *Ratio* (new series) 10, 48–64