Brit. J. Phil. Sci. **48** (1997), 363–389

Bananas Enough for Time Travel?

Nicholas J. J. Smith

ABSTRACT

This paper argues that the most famous objection to backward time travel can carry no weight. In its classic form, the objection is that backward time travel entails the occurrence of impossible things, such as auto-infanticide—and hence is itself impossible. David Lewis has rebutted the classic version of the objection: auto-infanticide is prevented by coincidences, such as time travellers slipping on banana peels as they attempt to murder their younger selves. I focus on Paul Horwich's more recent version of the objection, according to which backward time travel entails not impossible things, but improbable ones—such as the string of slips on banana peels that would be required to stop a determined auto-infanticidal maniac from murdering her younger self—and hence is itself highly improbable. I argue that backward time travel does not entail unusual numbers of coincidences; and that, even if it did, that would not render its occurrence unlikely.

1 Introduction

According to the objection to backward time travel taken most seriously in the philosophical literature, backward time travel implies the *auto-infanticide paradox*—or *Grandfather paradox*, as a variant is known. If backward time travel were possible, the objection runs, there would be nothing to stop a time traveller travelling back and killing her younger self. However, the fact that the time traveller was present in such a case would entail that she had not died young.[1] Hence her killing her younger self would involve a contradiction: she

[1] This assumes that intermittent existence of persons (the kind of thing involved in cases of resurrection) is impossible. If intermittent existence is *not* impossible, then there is no contradiction (see on), hence no paradox, and hence no objection to backward time travel.

both would and would not survive into adulthood. So if backward time travel were possible, there would be nothing to stop contradictions being true. Hence backward time travel is impossible. David Lewis [1976] showed that the auto-infanticide objection cannot establish the *impossibility* of backward time travel. Lewis's response, however, left many—including Lewis[2]—with the feeling that there is something very strange about backward time travel; and Paul Horwich ([1987], Ch. 7) fashioned from this uneasy feeling a new version of the auto-infanticide objection, according to which backward time travel is not *impossible*, but highly *improbable*. In this paper I criticize Horwich's new version of the objection. My conclusion is that the auto-infanticide objection to backward time travel should be permanently laid to rest: it can play no useful role in discussions of time travel. In arguing for this conclusion I hope, furthermore, to dispel somewhat the widely held view that worlds in which backward time travel occurs are very odd places. For all the difference it need make to what we see around us, time travellers could be in our midst today.[3]

The paper proceeds as follows. Section **2** makes two preliminary points: that prejudices against backward causation must be laid aside when considering the auto-infanticide objection to backward time travel; and that backward time travellers cannot *change* the past. Section **3** presents the arguments of Lewis and Horwich. Sections **4** and **5** criticize Horwich: Section **4** argues that even if backward time travel is strange in the way Horwich supposes, still that does not render its occurrence unlikely; Section **5** argues that backward time travel is not in fact strange in the way in question. Section **6** expands upon—and deals with objections to—the view presented in Section **5**.

2 Backward causation and changing the past

Backward time travel involves backward causation.[4] The time traveller pulls a lever in 1984, causing her time machine to materialize in 1969; she spills sugar in her time machine in 1984, causing it to become infested with ants in 1969; she drinks too many farewell beers in 1984, causing herself to have a hangover

[2] Lewis ([1976], p. 145) writes that 'a possible world where time travel took place would be a most strange world, different in fundamental ways from the world we think is ours'.

[3] There is an old argument to the effect that while backward time travel may be *possible*, it will never actually occur—for if it were going to occur, we would *already* have encountered the time travellers involved, whereas in fact we have done no such thing. (Horwich ([1987], p. 1) asks, 'If the necessary technology [for time travel] will ever become available, why haven't we encountered visitors from the future?' See also Fulmer ([1980], p. 155).) But 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 747 to pass overhead, would the debaters necessarily recognize it as containing flying humans? The answer to their question might have been staring them in the face for years, without them realizing.

[4] I mean by this simply that backward time travel involves causes occurring at times later than the times at which their effects occur. Some types of backward time travel—e.g. Gödel-style backward time travel—involve only this global kind of backward causation; others—e.g. Doctor Who-style backward time travel—involve, in addition, a local kind of backward causation.

in 1969; and so on. If one thinks that there is something wrong with backward causation, then one *a fortiori* thinks that there is something wrong with backward time travel. Now the auto-infanticide objection cannot even be set up if backward time travel is *already* ruled out: for purposes of *reductio*, the objection begins by assuming that backward time travel is possible. It follows that if we are going to consider the auto-infanticide objection to backward time travel at all, we must set aside all prejudices against backward causation.

Backward time travellers can affect the past—but can they *change* it? It is widely believed that if backward time travel were possible, one could go back in time and, with the benefit of foresight, prevent one's past mistakes and the terrible events of history. One could kill Adolf Hitler when he was young, urge Ned Kelly to cover his knees, and so on. Sadly, this idea is incoherent—I call it the *second-time-around fallacy*. There can be no first time around of a set of events, with the time traveller absent, followed by a second time around of the *very same events*, with the time traveller playing a role: for either there is no second time around; or else the second time around is a genuinely distinct series of events, to be involved in which is to *avoid* rather than *change* the original series of events. To see this, consider that to say that an event (for example, Australia losing the America's Cup in 1987) both did and did not occur, *simpliciter*, is to assert a contradiction. To be able to say without contradiction that some event both did and did not occur, one must posit at least two times or places such that the event occurred at one, but not at the other. Suppose that times are multiplied. One might posit two temporal dimensions, saying that time is like a plane rather than a line.[5] Then the year 1987, for example, is a line across the plane, and the time traveller can travel to a point on that line (1987 at hyper-time *b*) at which she prevents the Americans from winning back the America's Cup. She cannot, however, prevent the very loss which she witnessed as a young woman: at the point on the 1987 line which the time traveller experienced as a young woman (1987 at hyper-time *a*) Australia loses. Suppose, then, that places are multiplied. One might posit universes parallel to our own. In this case the time traveller can travel to a universe in which she sabotages *Stars and Stripes*. Once again, however, she cannot prevent the loss she witnessed before departing: in the universe from which she departs, Australia loses in 1987.[6] So the idea that time travellers can

[5] Meiland [1974] develops this suggestion for making sense of the idea of changing the past.

[6] There is also a third option. An event can both occur and not occur at *one* time and place: it can occur in some respects but not in others. A war, for instance, can still be going on (there is sporadic fighting) and yet already have ended (a peace has been signed). The idea in the time travel case would be that the time traveller can travel back and make it the case that in 1987 Australia only lost in some respects (Australia had the moral victory, say). But this suggestion just leads back to the two cases discussed above. For did Australia have the moral victory when the time traveller was young? If so, then she has not changed anything; if not, then the story is incoherent, *unless* one distinguishes two 1987s—and if one tries to distinguish two 1987s in terms of *respects* then one faces an infinite regress, which is itself a threat to coherence.

change the past is incoherent. If there is no bifurcation, of time or place, then there can only be contradiction, not change. Yet even if there *is* such bifurcation, still there can be no change, only *avoidance*: in this case what the time traveller causes to happen is *not* what she wanted to happen (she wanted *Adolf Hitler* to die, say—she did not want some analogue of Adolf Hitler in another universe or temporal dimension to die).

Time travellers can *affect* the past; and if there are parallel universes, or additional temporal dimensions, then time travellers can *avoid* the past.[7] They can *change* the past, however, no more than anyone can do anything contradictory, such as prove that 17=7. If a time traveller is going to travel to some past time, then she has already been there. If she is going to save a life or prevent a birth when she gets there, then she has already done so. A complete chronicle of events occurring at the time to which she travels describes her arrival and her actions, before she departs.[8]

3 Lewis and Horwich

Some science fiction writers respond to the auto-infanticide objection by saying that backward time travel *is* possible, as long as time travellers are accompanied by chaperones who prevent them from changing the past. Such chaperones are *ad hoc* and unappealing additions to time travel scenarios, however—and also unnecessary. David Lewis argues that no strange devices are required to stop the time traveller killing his younger self; rather, the time traveller fails for some commonplace reason. 'Perhaps some noise distracts him at the last moment, perhaps he misses despite all his target practice, perhaps his nerve fails, perhaps he even feels a pang of unaccustomed mercy' (Lewis [1976], p. 150). Perhaps his gun jams; perhaps he slips on a banana peel; perhaps he has a cardiac arrest. Nothing more than such ordinary occurrences is required to stop the time traveller killing his younger self. Hence backward time travel does not imply the truth of contradictions, even in the absence of chaperones. Hence backward time travel is *not* impossible.

Paul Horwich has criticized this response to the auto-infanticide objection. Horwich argues that if time travel to the local past were to occur, there would

[7] I make no claim regarding the existence of parallel universes or multiple temporal dimensions. In what follows I write as though there are no parallel universes, and only one temporal dimension. This is because if there *are* parallel universes, or multiple temporal dimensions, then the auto-infanticide objection to backward time travel loses its bite. If all that is meant by 'auto-infanticide' is killing one's younger self *in a parallel universe*, or *at a different hyper-time*, then auto-infanticide implies no contradiction. (Of course, there would be a problem if 'auto-infanticide' were taken to mean killing one's younger self in *this* universe, or at the *same* hyper-time as that at which one grew up—but in *this* case the points made below apply.)

[8] Many writers assume that backward time travel would involve changing the past. See, for example, Gödel ([1949], pp. 560–1) and Hospers ([1967], p. 177). The difference between changing the past and affecting or influencing the past is discussed by Brier ([1973], p. 363), Dwyer [1975, 1977], and Horwich ([1975], pp. 435–7; [1987], p. 116).

be bilking attempts—that is, attempts to initiate self-defeating causal chains (auto-infanticide attempts, for example). Any such attempts would be thwarted by slips on banana peels and so on. According to Horwich, however, the 'regular thwarting of bilking attempts [would] involve an endless string of improbable coincidences' (Horwich [1987], p. 123). Therefore, if time travel into the local past occurred, 'there would occur certain phenomena that we have empirical reasons to believe will not in fact occur' (ibid., p. 124). The conclusion is that the hypothesis of time travel into the local past is a bad one, because it implies propositions of extremely low probability. As Huw Price puts it, in the context of endorsing Horwich's conclusion, 'This is not a classical reductio, but it is as close as science ever gets' (Price [1996], p. 278, n. 7).

Here is a longer quote from Horwich:

> If there were a regular practice of travel into the past, then there would have to be a correlation between, for example, *the time traveller having an intention to bump off the child who lives at his old address* and the existence of *circumstances that will frustrate this intention.* And we know enough about human motivation (specifically, about the factors that might produce this bilking inclination) and the kinds of phenomena that could cause this plan to fail (amnesia, gun-jamming, brilliant surgeons, etc.) to claim that any such correlation would be an improbable coincidence (Horwich [1987], p. 125; emphases in original).

Horwich's point, then, is that time travel to the local past—the type of time travel, that is, that would make room for bilking attempts—is extremely unlikely to occur because it would require the occurrence of improbable coincidences. Backward time travel is not *impossible*—Horwich goes that far with Lewis—but *local* backward time travel is extremely *improbable.*

In the next section I argue that even if local backward time travel did entail long strings of slips on banana peels and other such coincidences, that would not show that such time travel is improbable, and will occur at most extremely rarely. It would show *only* that such time travel has not occurred in our vicinity, and will not occur in our vicinity within the next few generations—a far weaker conclusion. In Section 5 I argue that in any case, it is not true that local backward time travel entails unusually long strings of slips on banana peels and so on.

4 How often do coincidences occur?

What follows from the supposition that backward time travel entails coincidences? Horwich distinguishes two types of coincidence: a 'non-Humean' type (which will not concern us directly) and a type which involves violations of the *Principle of V-correlation*, which is as follows (call it 'PVC' for short):

> if events of type *A* and *B* are associated with one another, then either there

> is always a chain of events between them . . . or else we find an earlier event of type *C* that links up with *A* and *B* by two such chains of events. What we do not see is . . . an inverse fork—in which *A* and *B* are connected only with a characteristic subsequent event, but no preceding one (Horwich [1987], pp. 97–8).

Horwich's argument is as follows. We do not see inverse forks—they are highly improbable. It follows that coincidences of the PVC-violating type are highly improbable (Horwich [1987], p. 99). Time travel to the local past entails coincidences of the PVC-violating type, rather than the non-Humean type (ibid., p. 120). Therefore time travel to the local past is highly improbable.

Assume for the present that time travel to the local past *does* entail phenomena that violate PVC. The question is: why should we think that PVC-violating coincidences are improbable? One way to motivate this thought would be to derive the improbability of PVC-violating phenomena from some well-established principle or theory—but in fact no such derivation is available. Horwich suggests that the fork asymmetry can be traced to the initially micro-chaotic condition of the universe. It is not known, however, that the initial micro-chaos condition obtained: the main evidence for initial micro-chaos is the observed fork asymmetry itself.[9] Horwich concludes:

> I should make it clear that nothing in what follows will depend on the speculative physical hypotheses that I have suggested here [that is, hypotheses concerning what might *explain* initial micro-chaos, *assuming* it obtained]. What *will* be important for explanatory purposes is the fork asymmetry (Horwich [1987], p. 76).

It seems that Horwich has got things back to front. PVC concerns simply what *has been observed*: it is a purely *de facto* principle, and has no modal force. Horwich is relying not on some independently established principle or theory to establish the improbability of PVC-violating coincidences; rather, he is relying simply on the fact that we have never seen the latter. But arguing that local backward time travel will occur at most rarely, because it involves the occurrence of phenomena which we have never seen, is no better than arguing that one has never seen humans fly, and therefore will see humans fly at most rarely—even as the Wright brothers set up in the neighbouring field.

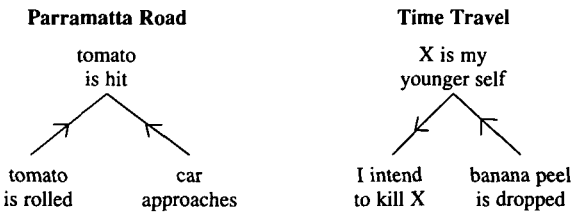
Consider an argument that runs parallel to Horwich's. Parramatta Road is a road in Sydney that was once very quiet and is now very busy. Imagine someone attempting to roll ten tomatoes across Parramatta Road at two second intervals, starting at some randomly chosen time. In the olden days,

[9] Cf. Horwich ([1987], p. 127): 'the observations that have led us to the second law of thermodynamics and to the principle that correlated events are causally connected also suggest that the initial conditions in our world exhibit a certain randomness.'

it would have been a coincidence if even one of the tomatoes had failed to reach the other side. Nowadays, however, it would be no coincidence if *all* the tomatoes were squashed. Squashings are no longer coincidental because they are no longer improbable; they are no longer improbable because there are more cars on the road nowadays. (If you don't agree that being improbable is a necessary condition for being a coincidence, then you can say that a squashing is still a coincidence, even today—but that a squashing is no longer an *improbable* coincidence.) Now suppose some old-timer once argued as follows. If the number of cars on Parramatta Road were to increase, persons with time on their hands and tomatoes in their pockets would make tomato-rolling attempts—and their attempts would entail long strings of coincidences (namely, many out of the set of ten tomatoes getting squashed, attempt after attempt). Hence, an increase in the number of cars on Parramatta Road entails arbitrarily long strings of coincidences. Long strings of coincidences are improbable. Hence an increase in the number of cars on Parramatta Road is improbable, and will very likely not occur.

This is a poor argument—but it is analogous to Horwich's argument. (Substitute the occurrence of backward time travel for the increase in the number of cars, bilking attempts for tomato-rolling attempts, and banana-peel-slipping coincidences for tomato-squashing coincidences.)[10] The fault of both arguments is as follows. We observe that certain phenomena occur only rarely, and hence we conclude that such phenomena are improbable. Now in so far as we can, in general, trust inductive inferences, we can trust that in contexts similar to those in which we have observed certain phenomena to occur only

[10] There is one difference between the Parramatta Road case and the time travel case:



In the Parramatta Road case, the correlation between tomatoes being rolled and cars approaching rises when the number of cars approaching rises; the correlated elements have only a common *effect*. In the time travel case, a causal connection is introduced to raise the correlation between murderous intentions and stray banana peels, which normally occur together with only a common effect (namely, the survival of the potential victim). In the time travel case, the coincidence is a partial cause of the potential victim being the time traveller's younger self, and it is the fact that the potential victim *is* the time traveller's younger self that causes the time traveller to want to murder him. This difference between the two cases is not, however, problematic for my argument—for the difference is solely a matter of the presence in one case of backward causation. Only someone with an *antecedent* prejudice against backward causation could argue that forks of the Time Travel type are *more* improbable than forks of the Parramatta Road type, and hence that the analogy drawn above fails.

rarely, the phenomena in question will not occur very often. We can, however, draw no conclusions concerning the frequency with which the phenomena will occur in contexts *unlike* those in which we made the observations. Those observations give us no reason at all to suppose that the phenomena in question will never and nowhere occur frequently—and if the phenomena *were* to start occurring regularly, then we could and would call them improbable no longer. Contexts involving lots of cars on Parramatta Road are unlike the context in which tomato-squashings on Parramatta Road were observed only rarely. Hence the latter observations are no basis for conclusions concerning the likelihood or otherwise of there ever being lots of cars on Parramatta Road. Similarly, contexts involving time machines are unlike the context in which slips on banana peels, and other such coincidences, are observed only rarely (assuming that time travel entails long strings of slips on banana peels and so on). Hence the latter observations are no basis for conclusions concerning the likelihood or otherwise of there ever being time machines.

An analogy may be useful here. Suppose a friend of mine tells me that she has invented a device which allows humans and lions to converse in English. I can well object that she can have done no such thing—*if* there exists a well-confirmed, independently established theory which tells us, say, that if a lion could talk, we could not understand him. (Such a theory may turn out to be false, but it provides at least a good prima facie reason for rejecting my friend's claim.) In the absence of any such theory, however, I cannot simply say that my friend can have invented no such device, because no one has ever seen a lion and a human converse in English. 'Of course no one has,' my friend can reply, 'I have only just completed my device!' In general, it is important to distinguish two types of inexplicable situation: situations involving entities which we take to be covered by a particular theory, but which are not behaving as that theory dictates; and situations involving phenomena of which there is no existing theory. (There can be borderline cases, of course.) Other things being equal, one should reject hypotheses which entail the first type of inexplicability, but not hypotheses which entail the second type of inexplicability. Now PVC is a response to the fact that only certain patterns of correlation have been observed in the world. It is fair to argue that in the realm in which PVC has been developed, phenomena violating PVC will only rarely be encountered—for such phenomena would be inexplicable in the first of the above senses. But one cannot argue *from PVC* that phenomena violating PVC will only ever be encountered extremely rarely—for such phenomena might be inexplicable only in the second of the above senses. In short, one cannot conclude from the supposition that local backward time travel would bring with it what we ordinarily regard as improbable coincidences, that such time travel will occur only rarely: for the only reason we regard the events in question as improbable coincidences is that within our experience, they have

not occurred very often—and our experience does *not* (apparently) encompass backward time travel.[11]

Suppose that any day prior to tomorrow lies outside the local past of the departure of the first backward time traveller. We have not the slightest reason to suppose that we will not see extraordinary coincidences occurring tomorrow, as our time traveller attempts to murder her grandfather on the occasion of his kindergarten Christmas pantomime (that being the earliest recorded event in which an ancestor of the time traveller takes part). Thus the most that Horwich can argue—assuming that local backward time travel does entail coincidences—is that no time traveller from our local future *has visited us*—that is, that time travel to the local past will not occur within the next few generations, at least not around here. And that is something we know anyway, simply on technological grounds.

In this section it was assumed, for the sake of argument, that backward time travel entails long strings of slips on banana peels and so on. It was argued that it does not follow from this assumption that backward time travel can be expected to occur only rarely. In the following section I question the assumption itself: I argue that backward time travel does *not*, in fact, entail unusual numbers of coincidences.

5 Coincidences and backward time travel I

Horwich claims, in the passage quoted above, that 'If there were a regular practice of travel into the past, then there would *have* to be a correlation between, for example, the time traveller having an intention to bump off the child who lives at his old address and the existence of circumstances that will

[11] Horwich himself seems almost to realize the significance of this point: 'Not only would such an uncaused correlation be an inexplicable coincidence—and highly improbable, *given what we see of the world . . .*'; 'highly implausible, *given our experience of the actual world*' (Horwich [1987], pp. 121, 127; my emphases). The point applies in relation to causal loops, also. Suppose I steal a time machine from my local museum and use it to travel back in time. When I arrive at my destination, I donate the machine to the local museum—so it turns out that the machine I stole was in the museum only because I put it there. Where does the machine come from? Or consider the Borges story about the time traveller who takes back copies of works by a famous artist to discuss with the artist himself. The time traveller finds on his arrival a hopeless artist, who proceeds to become famous by copying the paintings given to him by the time traveller. Which paintings are copies, which originals? Where do the ideas for the paintings come from? These cases are puzzling, but they by no means show that the time travel scenarios in question are impossible or incoherent, *or even improbable*. We think it very improbable that objects or information should come from nowhere—but only because this does not happen very often. It does happen *sometimes*—for instance, when you say something and I mishear you. I think that you said something very profound—something which neither of us would, in fact, ever have thought of. Where does the idea come from? If this sort of thing were to start occurring regularly, then we would simply accept it without raising an eyebrow. And to say "Well, of course—*if* it started occurring! But it is not *going* to start occurring—for it is too improbable!" is to make the mistake of thinking that causal loops need always be inexplicable in the first sense outlined above, rather than the second sense. (Fulmer ([1980], p. 154), Levin ([1980], p. 70), and Lewis ([1976], p. 149) have interesting things to say on the subject of causal loops.)

frustrate this intention' (Horwich [1987], p. 125; my emphasis; original emphases omitted). Why does Horwich think that there would *have* to be such a correlation? An attempt to do something only entails occurrences which cause its failure if it is an attempt to do something impossible. That is, only attempts to do the impossible are *guaranteed* to fail. Is it impossible for the time traveller to kill the child who lives at his old address? One might think it is: the child who lives at the time traveller's old address *is* the time traveller; therefore, if the time traveller were to kill the child, a contradiction would hold (absent resurrection)—namely, the time traveller would both have died young *and* survived to make a time trip. There is no possible world in which a contradiction holds. Therefore there is no possible world in which the time traveller kills the child who lives at his old address.

The crucial step in this argument is the assumption that the following counterfactual is true:

A If the time traveller were to kill the child who lives at his old address, then a contradiction would hold.

If counterfactual **A** is true, then it is impossible for the time traveller to kill the child who lives at his old address, because doing so would involve changing the past, and would entail a contradiction. Therefore, if counterfactual **A** is true, any attempts by the time traveller to kill the child who lives at his old address must fail. Chaperones aside, the only thing available to cause such attempts to fail is a proliferation of stray banana peels. Therefore, if counterfactual **A** is true, attempts by the time traveller to kill the child who lives at his old address entail long strings of slips on banana peels. Horwich claims that human beings being what they are, local backward time travel entails attempts by time travellers to do things such as kill the children who live at their respective old addresses (Horwich [1987], pp. 121, 123). It follows that if counterfactual **A** is true, local backward time travel entails arbitrarily many banana peel slips—that is, arbitrarily many PVC-violating coincidences.

But is counterfactual **A** true? It might seem obviously true; but there is an alternative—just as good, if not better, and yet almost entirely overlooked in discussions of auto-infanticide:[12]

B If the time traveller were to kill the child who lives at his old address, then the child would not be his younger self.

Note that the truth of **B** is in no way incompatible with the fact that the child who lives at the time traveller's old address *is* in fact his younger self. We may suppose that in the actual world, 'the child who lives at his old address' designates the time traveller's younger self. If counterfactual **A** is true, then

[12] Lewis ([1976], p. 152) mentions counterfactuals like **B**, but does not discuss their significance. No one else I have read even mentions them.

there are no possible worlds in which the time traveller kills the child who lives at his old address. If counterfactual **B** is true, however, then there are such worlds: 'the child who lives at his old address' designates in these worlds, not the time traveller's younger self, but someone else.[13] This is not to say that **B** involves the idea of changing the past—the *second-time-around fallacy*—which we have seen to be impossible. The idea is not that the time traveller's killing the child who lives at his old address causes that child not to have been his younger self, when before—'the first time around'—the child *was* his younger self. Rather, in worlds in which the time traveller kills the child living at his old address, the killing is simply the cause of the victim's not being *anybody's* younger self. So if counterfactual **B** is true, then *if* the time traveller were to kill the child, then the child would never have been his younger self, and so the killing would not *change* anything—and hence is not impossible. Hence, if counterfactual **B** is true, any attempts that the time traveller might make to kill the child who lives at his old address are *not* guaranteed to fail. Hence such attempts—and hence local backward time travel itself—do *not* entail PVC-violating coincidences.

Isn't there a problem here? Even granting that it is not *impossible* for the time traveller to kill the child who lives at his old address, still, we know—if the child really *is* the time traveller's younger self—that all the time traveller's murder attempts *will* fail. Why is this? *Why* do all his attempts fail? Isn't it indeed an incredible coincidence that there should always be a banana peel on hand to trip up the time traveller whenever he tries to kill the child who lives at his old address? There just do not seem to be enough bananas around for that to occur—hence not enough bananas for local backward time travel, just as Horwich claims.

In fact there is no problem here. To ask 'why do coincidences always foil the time traveller's attempts to kill the child who lives at his old address?' is to get things back to front. If counterfactual **B** is true, it is only *because* the murder attempts fail that the child *is* in fact the time traveller's younger self in the first place. To get a better feel for how things look if **B** is true—and to see that if **B** is true, local backward time travel really does not entail coincidences—consider three cases, each of which is analogous to the case of the time traveller.

First Case. A time traveller sets off to kill John Austin in 1934. Austin lived until 1960. Therefore the time traveller will fail. There is nothing more problematic about attempts at auto-infanticide than about other attempts to

[13] The description 'the child who lives at his old address' is thus being read in **B** as a non-rigid designator. If 'the child who lives at his old address' is taken as a *rigid* designator, then the contrast between **A** and **B** is lost: the consequent of **B** states a contradiction, and hence **B** implies **A**. The distinction between rigid and non-rigid designators is due to Kripke ([1980], p. 48). A rigid designator denotes the same individual in all possible worlds; a non-rigid designator denotes different individuals in different possible worlds.

change the past. If attempts at auto-infanticide entail coincidences, then the attempt to kill Austin in 1934 should entail coincidences too. But it can easily be seen that to say that a series of coincidences is required to prevent the time traveller from killing Austin in 1934 is a very backward way of putting the matter. The question is not what *must* occur if the time traveller tries to kill Austin—it is what *did* occur. Whatever the time traveller is going to do has already been done. If Austin survived 1934 it is in part *because* the time traveller failed to murder him. If she had succeeded, then Austin would have died in 1934 rather than in 1960. Historians can research the question whether there were attempts made on Austin's life in 1934, before the time traveller departs—and her journey cannot change the truth values of their statements. If there were such attempts, then perhaps the time traveller arrived safely and it was she who made them. If there were no such attempts, then perhaps the time traveller never arrived; or perhaps she did arrive, but came to like Austin so much that she changed her mind about murdering him. Whatever is the case, there is no problem here. If a series of coincidences occurred which saved Austin's life in 1934, then good luck to him—he was simply lucky. The coincidences were not *entailed* by the time traveller's mission. On the contrary, it is *because* the coincidences just happened to occur—as coincidences sometimes do—that the time traveller's mission failed. If, on the other hand, no coincidences occurred, then the time traveller obviously decided not to try to kill Austin in 1934. For if she *had* tried to kill him, *and* no coincidences had saved him, then he would have died in 1934—and we know that that did not happen.

Someone might object that it is *impossible* for the time traveller to kill Austin in 1934; therefore if she goes back and tries to kill him, a whole series of coincidences will in fact be required to prevent her from succeeding. Now why would anyone say that it is *impossible* for the time traveller to kill Austin in 1934? Because they picture Austin living until 1960. It is impossible to change the past; therefore, they say, it is impossible to kill Austin in 1934—for he lived until 1960. But this line of thought is question-begging. It is trivially true that it is impossible to kill Austin in 1934, *given* that he lived until 1960. In just the same way, it is impossible for you not to be reading this paper today, *given* that you are reading it. The point is that one is not entitled to say that you are not sky diving today *until* one has checked whether or not you are reading this paper today. Similarly in the Austin case: one is not entitled to say that Austin lived until 1960 *until* one checks whether or not Austin was killed in 1934, by a time traveller or by someone else. It is getting things back to front to say that the time traveller's attempts to murder Austin in 1934 *must* fail—and that therefore they entail the occurrence of coincidences which ensure that they do fail—*because* Austin lived until 1960.

Yet that is exactly the sort of thing which one must say in the auto-infanticide

case, if one is to conclude that attempts at auto-infanticide entail coincidences. And it might seem to be a *better* thing to say in the auto-infanticide case, for the following reason. Not only is it trivially true that the time traveller cannot kill the child who lives at his old address, *given* that the child survives to make a time trip; but furthermore, we are entitled to assume that the child survives *before* we check on the outcome of the time traveller's attempts to murder him. We are so entitled, it is thought, because the mere fact that the time traveller is alive *guarantees* that the child does not die young. The idea is that the auto-infanticide case is different from the Austin case, in that Austin is not guaranteed to live until 1960—whether he does or not depends upon what the time traveller does in 1934—whereas the child *is* guaranteed to live beyond his encounter with the time traveller. The mere presence of the time traveller provides the guarantee. In the Austin case, if the time traveller fails to kill Austin, then there is no problem; if she does *not* fail, then the history books are wrong and Austin in fact died in 1934. According to the present view, the second option is not available in the auto-infanticide case: the history book (so to speak) is in this case infallible. The presence of the time traveller *guarantees* that the child will not die. But does it really? Well, *only if counterfactual* **A** is true. If we accept counterfactual **B** on the other hand, then we say that there is no *guarantee* that the child survives. He *does* survive, but only *because* of what happens next—namely, he gets genuinely lucky. His getting lucky is not bound to occur: if he were not going to get lucky, then he would not be the time traveller's younger self.

So, the time traveller's attempts to kill Austin in 1934 do not entail coincidences; and the only thing which could make us say that the auto-infanticide case is any different is adherence to counterfactual **A**.

Second Case. Suppose that several years before I was born my parents were booked on a train that ended up travelling past a leaking nuclear facility, causing everyone on board to become sterile. It is somewhat of a coincidence that they missed the train because the ferry they were taking to the station was delayed by fog. Suppose that had my grandmother decided to visit Benny rather than Lenny one day when she was a little girl, she would have perished in the fire which killed Benny. It is quite a coincidence that she was annoyed with Benny that day because the latter had spilt ink all over her collection of cigarette cards the day before. Suppose that had my father been driving into work as usual one day, he would have been crossing the local bridge at the moment it was hit by a ship. It is a coincidence that my father stayed at home that day to look after his brother, who had contracted food poisoning after eating a fish won in a raffle. Now suppose someone were to argue that my birth *requires* the occurrence of all these coincidences. One would reply as follows: 'Well, yes—in a sense; but that's a very backward way of putting the matter. For we cannot get into the position of holding my birth fixed *until* we check

whether or not the coincidences occurred. If we just pull my birth out of a hat, then it seems to require an extraordinary string of coincidences. But if we pay proper attention to the causes of my birth, then there is no problem: the coincidences involved are not *required*, but purely *de facto*. It is *because* they just happened to occur that my birth was able to take place.'

Clearly, if such an argument could be run in the auto-infanticide case, then the view that local backward time travel entails coincidences would collapse. And the only thing that could stand in the way of running an analogue of this argument in the auto-infanticide case is adherence to counterfactual **A**. If one just pulls out of a hat the fact that the time traveller and the child who lives at his old address are one and the same person, then one can argue that attempts by the time traveller to kill the child *require* the occurrence of foiling coincidences. But if one holds counterfactual **B** true, then one cannot simply pull the identity of the time traveller and the child out of nowhere. According to **B**, the identity holds in part *because* certain coincidences occur which stop the child dying, allowing him to grow up to become a time traveller. The coincidences contribute to the identity, rather than the identity requiring the coincidences. The only sense in which the identity requires the coincidences is the back to front sense in which my birth requires my ancestors not to have died childless. This can be hard to see: the requirement can seem more compelling in the auto-infanticide case. This is because while we should not hold fixed my birth without looking to see what comes *before* it, we may do so without looking to see what comes *after* it—and we have a tendency to carry this habit over to the auto-infanticide case: to hold fixed the time traveller being there, facing his younger self, without any reference to what happens *subsequently*. But this tendency is the product—or perhaps the origin—of a prejudice against backward causation, and hence is out of place here. If we turn around, as it were, and face the auto-infanticide case from the other temporal direction, we see that it is *exactly* like the case of my birth, except that the causal arrow points against, rather than with, the temporal arrow. In the latter case the contributing coincidences occur *before* my birth. If counterfactual **B** is true, then in the former case the contributing coincidences occur *after* the time traveller's fronting up to his younger self.

Third Case. Suppose that every object has written upon it the date on which it will cease to exist.[14] The idea is not that the dates are self-destruct devices which *cause* the objects on which they figure to cease to exist at some particular time. Rather, the dates simply *inform* us as to when, *as it happens*, the objects on which they figure will cease to exist. Perhaps God, on creating the world, looked into the future to see when objects will—*as it happens*—cease to exist, and then wrote the corresponding dates on the objects. Or

[14] I am grateful to Huw Price for suggesting that I consider this imaginary state of affairs.

perhaps a time traveller travelled into the future, observed the demise of objects, and then travelled back and wrote the dates.[15] Now, one tries to destroy things before their expiry dates, and to preserve things beyond their dates, but all one's attempts fail. Suppose my pen bears the date 2003. I take it outside to place it under the wheels of a passing train, but there is a train strike that day. The telephone rings just as I am about to drop the pen into a vat of acid. I slip on a banana peel on my way to put the pen in the microwave. My dog eats my designs for a pen grinder. And so on, for as long as you please. However many attempts I make, the attempts in no way *require* the occurrence of the coincidences that foil them. The coincidences just happen, as coincidences sometimes do. God looks ahead, sees them happen, sees that the pen survives until 2003, and *therefore* writes '2003' on the pen. The fact that my pen has '2003' on it does not *prevent* me destroying it, does not make its destruction prior to 2003 *impossible*. My destruction attempts are not *required* to fail in order that the date remain correct. Rather, the date is only there *because* the attempts fail. The existence of the date does not entail the coincidences; on the contrary, it is *because* the coincidences occur that the pen bears the date '2003' in the first place. Future events do not have to conform to the date. It is the date that conforms to future events.

What we regard as a probable frequency of banana peel slips and other coincidences depends upon how many such things we are used to seeing. In our world that's not many—so we might think it unlikely that a string such as that described above, in relation to my attempts to destroy my pen prior to 2003, would ever occur. No problem: if such a string really is unlikely, then very likely my pen would not bear the date 2003—assuming I try hard to destroy it prior to 2003. One or two coincidences might occur, but then either I would stop trying to destroy the pen, or else the pen would be destroyed. The only way to feel that there is a problem here is to imagine that I make a *few* attempts—or no attempts at all—to destroy my pen; that God sees my attempts fail and writes '2003' on the pen; and that *then*—the second time around, as it were—I make lots of attempts to destroy the pen: attempts which *must* now fail, and hence entail coincidences to foil them. But to imagine this is, of course, to succumb to the second-time-around fallacy.

The date hypothesis is of interest because a backward time traveller is, at the end of her journey to the past, someone who can (as it were) see the expiry dates on many objects. Just as my pen's bearing the date '2003' does not *require* coincidences in order to stop the pen being destroyed prior to

[15] Note that it is a peculiarity of worlds such as this that God and the time traveller, upon examining the future, see there the dates which they will write as a *result* of examining the future. Because changing the past, present or future is impossible, there can be no *first* future, devoid of dates, which God or the time traveller observes, then writing dates and thus making way for a *second* future, complete with dates. (This point does not apply in contexts involving multi-dimensional time or parallel universes.)

2003—because the occurrence of the coincidences is part of the causal history of the writing of the date—so, if counterfactual **B** is true, the time traveller fronting up to the child who lives at her old address—who is in fact her younger self—does not *require* the failure of any attempts the time traveller might make to kill the child. The time traveller's murder attempts do not *require* the coincidences that foil them, for the reason that it is *because* the coincidences just happen to occur that the child *is* the time traveller's younger self in the first place, that the child as it were bears an expiry date that lies in the future of her encounter with the murderous time traveller. If long strings of slips on banana peels really are unlikely to occur, then most probably time travellers will either try only a few times to kill persons living at their old addresses, or else will succeed in their murder attempts. The only way one could see a problem with this would be if one fell for the second-time-around fallacy: if one supposed that murder attempts are made on our time traveller at most a few times when she is young; that any attempts that *are* made are foiled, unproblematically, by one or two lucky coincidences; and that *then* the time traveller goes back in time, and—the second time around, as it were—makes a great many attempts to murder her younger self—attempts which this time are *required* to fail, and hence entail foiling coincidences.

It has been argued in the present section that if counterfactual **B** rather than counterfactual **A** is true, local backward time travel does not entail coincidences. If **B** is true, time travellers could be here amongst us now, for all the difference it would make to the numbers of slips on banana peels and other such coincidences we see occurring around us. The next section expands upon—and deals with possible objections to—the argument of the present section.

6 Coincidences and backward time travel II

It might be conceded that if counterfactual **B** is true, attempts by time travellers to kill persons living at their old addresses do not entail coincidences, but the claim be made that this is beside the point: that although I have followed Horwich's wording, I have not followed his intention. I have supposed that we are concerned with what would happen if the time traveller tried to kill *the child who lives at his old address*, with the description 'the child who lives at his old address' read as a non-rigid designator. That is why I have been able to say that it is possible for the time traveller's murder attempts to succeed, and hence that those attempts do not entail foiling coincidences: in worlds in which the time traveller succeeds in killing the child who lives at his old address, the child who lives at his old address is not the time traveller's younger self (even though in the actual world the child who lives at his old address *is* the time traveller's younger self). The objection is that our real concern is not with what

would happen if the time traveller tried to kill the child who lives at his old address—who is in fact his younger self—but with what would happen if the time traveller tried to kill *his younger self*. Killing one's younger self is impossible, the objection continues—hence attempts to kill one's younger self *do* entail foiling coincidences.

At this stage, two distinctions are needed. First, distinguish three readings of 'kill his younger self'. One can read 'his younger self' here as a non-rigid designator (a person could want to kill whoever it is that satisfies the description 'his younger self'), as a rigid designator (a person could want to kill *this person*, who happens to be his younger self), and as a bound variable (a person could want to be a younger-self-killer—could want to be an x such that x kills the younger self of x).[16] Second, distinguish two senses of 'trying': what we might call the 'de dicto' sense and the 'de re' sense. What I am trying to do in the *de dicto* sense is whatever I take myself to be trying to do: drink a glass of milk, say. In the *de re* sense I may be trying to do something quite different: drink a glass of white paint, say—if what I have in my hand is in fact a glass of white paint.

Note some features of these two types of trying. First, observe that it makes no sense to say that someone is trying in the *de dicto* sense to do something, and yet feels *certain* that she will fail. Suppose that one has just read a proof of Pythagoras's theorem, and accepts it as definitive. It makes no sense at all to suppose that one could then try, in the *de dicto* sense, to *dis*prove Pythagoras's theorem. One could try to do such a thing (in the *de dicto* sense) only in so far as one thought that the original proof might have been flawed in some way. Again, suppose that one has just been told that it is impossible to eat an ordinary slice of bread in one minute, without taking any fluid. One can only try, in the *de dicto* sense, to perform this feat in so far as one does not believe what one has been told. The idea of someone *fully accepting* that some task cannot be performed, and yet trying (in the *de dicto* sense) to perform that task, is simply incomprehensible. Second, consider the following case. I am a collector of aluminium cans, and I happen to know that several Fanta cans with colour-reversed labels—extremely valuable pieces—have recently been thrown into garbage bins in the Sydney area. I myself am averse to venturing outside my home, and unfortunately my assistant has a deep hatred of Fanta cans, considering them evil and dangerous: not only would he never own one himself, he will not even bring Fanta cans to me. Accordingly, I ask my assistant, not to search through the bins of Sydney until he finds the cans, but rather to bring me the entire contents of the bins. I will do the searching myself. Now making off with the entire contents of a garbage bin is a difficult and illegal process, and often my assistant will have to have several tries at a

[16] I am grateful to David Lewis for drawing my attention to the third of these readings: I had originally noticed only the first two.

given bin. My assistant will never, in the *de dicto* sense, be trying to bring me a Fanta can. But some of the time—namely, when the bin he is trying to raid contains one—my assistant *will* be trying, in the *de re* sense, to bring me such a can. But only sometimes, and only by accident. If I want a high correlation between my assistant raiding bins and my receiving Fanta cans, I will have to hire an assistant who is prepared to try in the *de dicto* sense to bring me Fanta cans. Likewise, it would require many coincidences to make it the case that every time my assistant went out, he brought back the contents of a briefcase rather than a garbage bin—but that is because he is trying in the *de dicto* sense to bring me the contents of garbage bins. No strange coincidence is involved if he fails to bring me one of the desired Fanta cans—for he will only ever try in the *de re* sense to bring me one of them, and hence will try to do so only sometimes, and then only by accident.

Now, to return to the claim that our concern should be not with what would happen if the time traveller tried to kill the child who lives at his old address, but with what would happen if the time traveller tried to kill *his younger self*. Which reading of 'his younger self' is in play here? Suppose we read 'his younger self' as a non-rigid designator: understood in this way, it is indeed impossible for the time traveller to kill his younger self—for anyone he kills is *ipso facto* not *anybody's* younger self, and hence does not satisfy the description 'his [the time traveller's] younger self'. But clearly, if our concern is with whether or not the time traveller can kill his younger self, with the latter description read as non-rigid, then our concern is trivial. Compare the old 'can you eat my dinner?' problem. Of course any food you eat *ipso facto* fails to satisfy the (non-rigid) description 'my dinner'. But the interesting question is 'can you eat *this food*?' Perhaps you can; but perhaps my dog guards my food, in which case you cannot.

That leaves us with the rigid designator and bound variable readings of 'his younger self'. Consider the former. 'His younger self' denotes the time traveller in the actual world; hence (on the rigid reading) it denotes the time traveller in all possible worlds. A world in which the time traveller kills his younger self and yet goes on living is, therefore (absent resurrection, parallel universes, and so on), a world in which the time traveller is not himself. There are no such worlds. Therefore, the objection runs, the time traveller's attempts to murder his younger self *cannot* succeed, and hence—contra Section 5—*do* entail foiling coincidences. On the bound variable reading, too, it seems that the time traveller's attempts must fail, and hence do entail foiling coincidences: for (again absent resurrection, parallel universes, and so on) a younger-self-killer is an impossible object, and the time traveller can never succeed in becoming one.

The objection just presented is the first of several which I shall consider in this section, each of which tries to show that backward time travel really does

entail a great many slips on banana peels and so on. My replies to all of these objections will have the same form, and before replying to the current objection, I shall outline that form. My reply will be to point out that backward time travel, *in itself*, does *not* entail slips on banana peels and other such coincidences. Rather, each 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* as rare and apparently improbable as long strings of slips on banana peels. Hence, in order to derive large numbers of output coincidences, the objections need to stipulate the occurrence of large numbers of *input* coincidences. This means that if we are confining ourselves to the actual world—in which, we are supposing (leaving aside the considerations of Section **4**) events in as scarce supply as slips on banana peels will continue to occur only rarely—then the arguments in question cannot get started; while if we instead suppose that we live in a world in which there is no reason to expect that the kinds of input coincidence in question will not occur often, then we *ipso facto* suppose that we live in a world in which there is no reason to expect that long strings of slips on banana peels and so on will not occur often—and hence in which the entailment of such strings by backward time travel would give us no reason to suppose that backward time travel will occur at most rarely. In short, the only worlds in which backward time travel *and* large numbers of slips on banana peels occur are worlds in which such slips can*not* be deemed improbable. For the sake of brevity, when I wish to make this point in future I will write simply 'improbable outputs can be derived only on the assumption that equally improbable inputs occur'.

Now to reply to the first objection—namely, that killing one's younger self (as opposed to the child who lives at one's old address) is impossible, and hence that the time traveller's attempts to kill his younger self do entail foiling coincidences. The first point to note is that we cannot suppose that the time traveller recognizes the impossibility of younger-self-killing, and yet tries—in the *de dicto* sense—to kill his younger self. So, the time traveller must fail to realize that younger-self-killing is impossible; or else he must try only in the *de re* sense to kill his younger self. Consider these alternatives in turn.

First the *de dicto* case. One can tell any number of stories in which time travellers end up, for one reason or another, making *de dicto* attempts to kill their younger selves. A time traveller might, for instance, fall for the second-time-around fallacy, and fail to realize that it is impossible to kill one's younger self: she just goes on trying and trying to kill the person before her, even though she is sure that she is facing her younger self. Or again, our time traveller might believe in resurrection: he tries and tries to murder his younger self, in the belief that if he succeeds, the younger self will be brought miraculously back to life. These stories could be multiplied without end—but they all suffer from the same defect, namely: improbable outputs can be

derived only on the assumption that equally improbable inputs occur. Take the first story: true, if a high proportion of time travellers were to fall for the second-time-around fallacy, then local backward time travel would indeed entail a great many slips on banana peels and so on. But it seems very improbable that there should be a systematic correlation between time travelling intentions and fallacious reasoning. So, if improbable things happen only rarely, then only a few time travellers will fall for the second-time-around fallacy, and hence only a few will try in the *de dicto* sense to kill their younger selves, and hence only a *few* banana-peel-slips and so on will occur—and we cannot baulk at *that*: it is not as if such slips *never* occur! The alternative is that highly improbable things such as mass irrationality on the part of time travellers *can* be expected to occur—but then, if the improbability of such things is no bar to their occurrence, how can we argue that backward time travel will not occur, for the reason that it is too improbable? Similarly for the second story: it would be a remarkable coincidence if a large number of time travellers were to believe in resurrection—*unless*, of course, resurrection were actually observed. But then in the latter case, in which resurrection is a possibility, auto-infanticide attempts need not fail, and hence do *not* entail coincidences after all.

Second, the case of *de re* auto-infanticide attempts. It seems clear that time travellers would sometimes try in the *de re* sense to kill their younger selves: a time traveller might, for instance, try to land her time machine in just the spot where, unbeknownst to her, her younger self sits playing with a pocket calculator. But normally, time travellers would make such *de re* attempts *only sometimes*, and then only by accident, and an abnormal number of stray banana peels would not be required to prevent the attempts from succeeding (compare the case of the Fanta can). Furthermore, such attempts could be expected to have a self-regulating effect. Suppose that our time traveller is on a mission to kill her identical twin, from whom she was separated at birth. She locates her twin—or so she thinks—and throws a grenade, but it fails to explode; her gun jams; she slips on a banana peel. The greater the improbability of such happenings, the less of them it will take to convince our time traveller that she is facing her younger self, rather than her twin—and when she realizes this, her killing attempts will cease (assuming that she recognizes the impossibility of younger-self-killing, the case in which she does *not* recognize this having been treated in the previous paragraph).

Of course, this is only the normal case: again, we can cook up all sorts of scenarios in which time travellers end up making arbitrarily large numbers of *de re* attempts to murder their younger selves. Suppose that some military regime programmes as many soldiers as you please to kill their younger selves, and wipes the soldiers of their memories, so that they do not recognize their younger selves when they meet them (and hence do not try in the *de dicto* sense to kill those younger selves). Once these time travelling soldiers are turned

loose, a *huge* number of coincidences will greet their killing attempts. There are many stories similar to this one that could be told. The problem with all of them is that they build in as premises not simply backward time travel, but also things such as military regimes that programme soldiers to kill their younger selves. Now such things are perfectly *possible* (assuming that backward time travel is possible)—but it seems incredibly *unlikely* that they should ever occur. Philosophers trying to come up with objections to backward time travel think up such scenarios, but we have no reason to expect that someone with time machines at her disposal would ever attempt to actualize one of them. Once again, improbable outputs can be derived only on the assumption that equally improbable inputs occur.

6.1 Curiosity

We have now dealt with the objection according to which, when we turn our attention to attempts by the time traveller to kill *his younger self*—rather than *the child who lives at his old address*—we see that backward time travel does indeed entail a great many slips on banana peels and so on. Now we face a new objection, however: for someone might argue, in response to the point made in the previous paragraph, that there is an *obvious* motive for someone with time machines at her disposal to do things such as programme time travellers to try to kill their younger selves—namely, *curiosity*. The military regime *knows* that the soldiers will not succeed in killing their younger selves: it simply wants to see what stops them. In general, the new objection is as follows. Won't a great number of coincidences occur if time travellers attempt, not to *kill* their younger selves or their ancestors, but simply to see what stops the younger selves and the ancestors dying whenever the time travellers throw grenades at them and so on?[17] This sort of scenario does not involve time travellers in attempts to perform tasks at which they know they will fail, and seeing all the strange ways in which one's younger self survived would no doubt be rather enjoyable. There is, then, no reason to suppose that time travellers would not make a *great many* attempts of this sort. Hence, backward time travel *does* entail a great many slips on banana peels and so on after all.

There are two cases to distinguish here. First, suppose that our time traveller knows, *before* throwing any grenades herself, of grenade throwings involving her younger self (or her ancestors). In this case, the time traveller might intend to be the person who causes all the near-death experiences she knows about; or she might intend to *make sure* that the grenade throwings occur—that her beliefs about the past are correct; or again, she might not *intend* to do anything—her knowledge of what she is going to do might lead to quietism on her

[17] Several persons raised this objection; the first to do so was Amitavo Islam.

part (she throws the grenades, but her heart is not in it).[18] The first two of these options involve rather strange intentions, however: not ones which we may assume a great many time travellers would have; while the third option is not the one we want either: the reason we supposed many time travellers would throw grenades at their younger selves and ancestors was that it seemed like a fun thing to do. So, to get cases involving actions which it is fair to suppose many time travellers would try to perform, we need to assume that the time travellers in question do not know of the occurrence of any grenade throwings and so on involving their younger selves or their ancestors, *until* they have thrown the grenades or whatever themselves. (After all, *curiosity* goes hand in hand with ignorance, and as for the *fun*, that was supposed to be provided by the *surprise*.) But now, if a time traveller is going to throw grenades at herself, then the more grenades she throws, the more of a coincidence it is that she has forgotten the stupendously dangerous events of her childhood (improbable outputs can be derived only on the assumption that equally improbable inputs occur).[19] And if she is going to throw grenades, not at herself, but at one of her ancestors, then she faces a dilemma: either the time traveller knows where the ancestor in question was on some particular day—in which case it would be a great coincidence if she did not know about any near-death experiences on the part of the ancestor on that day (improbable outputs can be derived only on the assumption that equally improbable inputs occur); or else the time traveller does not know where the ancestor was on some particular day. In the latter case, it will be quite a coincidence if the time traveller locates the right person at all (improbable outputs can be derived only on the assumption that equally improbable inputs occur); and if she does *not* locate the right person, then no coincidences need ensue: the person will very likely die at the first grenade.

The case of the military regime is similar: it is extremely unlikely that the regime could find out enough about the whereabouts of its soldiers' younger selves on particular days to enable it to send the soldiers back to make killing attempts, *without* finding out about the spectacular, exciting and stupendously dangerous events in which the younger selves were involved on those days—but then, if the regime *does* know about these dangerous events before deploying its soldiers, we cannot make sense of that deployment in terms of either curiosity or fun-seeking: the deployment becomes a highly mysterious

[18] This third case involves a causal loop (see footnote 11 above): the time traveller throws the grenades because she knows she will, and feels she has no choice; but she only knows that she will throw the grenades *because* she will throw them. Beware the second-time-around fallacy here: it cannot be the case that the time traveller throws the grenades joyfully, and then—the second time around, as it were—knows what she will do, and so throws the things half-heartedly. She only gets one go—be it half-hearted, or be it joyful.

[19] Someone might say that the traumatic nature of the events could be expected to lead to a suppression of memory on the part of the time traveller. But equally, seeing a few of these events again (as the perpetrator rather than the victim) could be expected to bring back memories of all of them.

event, the occurrence of which is simply stipulated by the opponent of backward time travel. Once again, improbable outputs can be derived only on the assumption that equally improbable inputs occur.[20]

In sum, switching the case from time travellers trying to kill their younger selves, to time travellers seeing what stops their younger selves dying, does not lead to any new problems for backward time travel.

6.2 Memory failure

Someone might accept the foregoing arguments, and yet still claim that backward time travel really does entail a great many coincidences. True enough, the objection might run, there are worlds in which the time traveller succeeds in killing the child who lives at her old address, and in such worlds no slips on banana peels occur. But coincidences of a *different* type occur in these worlds: namely, coincidences which cause the time traveller to be mistaken about the identity of the child who lives at her old address. *Either* coincidences will cause her to fail to kill anyone, *or* coincidences of a different type—such as failures of memory concerning where her younger self was on her destination date—will cause her to kill someone who is not, in fact, her younger self.

Clearly, this objection can get under way only if it is assumed that the time traveller tries in the *de dicto* sense to kill her younger self. For suppose that a time traveller sets out to kill certain persons, but it is not her intention to kill her younger self: she merely happens, occasionally, to try to kill someone who is, unbeknownst to her, her younger self—in other words, she tries only in the *de re* sense to kill her younger self. It would indeed be very odd if every time this time traveller tried to kill someone, she ended up killing a Frangipani tree—but that is because she is trying in the *de dicto* sense to kill *persons*. No strange coincidence is involved, however, if she fails to confront her younger self: it is part of the story that she does so only sometimes, and then only by accident.

If we assume that time travellers try in the *de dicto* sense to kill their younger selves, then we do indeed get a lot of output coincidences occurring: if not slips on banana peels, then slips of memory. Such memory slips really do seem

[20] One might imagine, of course, that there is a part of the brain which keeps a record of where one has been, and *another* part which keeps a record of what has happened to one—and that the regime has a device which reads the first part of the brain, but not the second. The regime takes a brain reading, wipes the soldier's brain, then programs the soldier to go to some place and time where she will find her younger self, and attempt to kill the person she finds. But do the opponents of backward time travel really want to hang their objections on claims about the information storage structure of human brains, and on the existence of certain kinds of brain reading-and-wiping devices? Even if they do want to do this, to do so would be to beg the question: for the fact that a combination of time machines and a certain type of brain reading-and-wiping device (and a military regime with a certain sort of plan) leads to the occurrence of long strings of slips on banana peels and so on, no more shows that time machines are unlikely to exist as that *the type of brain reading-and-wiping device in question* is unlikely to exist.

highly improbable, especially in cases in which the time traveller travels only a little way into the past. But as we saw above, the case in which we assume that time travellers try in the *de dicto* sense to kill their younger selves is one in which improbable outputs can be derived only on the assumption that equally improbable inputs occur (the inputs being things such as a great many time travellers falling for the second-time-around fallacy)—and it makes not the slightest difference whether the output coincidences are slips on banana peels or slips of memory. In the normal case, in which no input coincidences are assumed to occur, time travellers simply do not try in the *de dicto* sense to kill their younger selves.

6.3 Forcing coincidences

I shall consider a final attempt to argue that backward time travel really does entail a great many coincidences. Suppose that you are visited by your future self. The future self recognizes the impossibility of younger-self-killing, and attempts to do you no harm: she merely has a chat, and then pops off again—but not before telling you exactly when and where her time trip began. There are no memory-slips or banana-peel-slips here—but it seems that there are coincidences of a different type. For suppose that you do not want to be a time traveller: you have other plans. No matter how hard you try to be somewhere other than the place your future self mentioned at the time she mentioned, coincidences will force you to enter a time machine (or wormhole, or whatever) at that time and place.[21]

This scenario is similar to the case in which objects have written upon them the dates on which they will cease to exist (the comments made above in relation to the latter case apply again here—*mutatis mutandis*—so at this point I shall be brief). Just as the existence of a date on an object does not *require* the failure of attempts to preserve the object beyond that date—the date being there *because* the attempts will, as it happens, fail—so your older self informing of you of your upcoming time trip does not *require* that you make the trip. It is not that, after being visited by your older self, you *have* to make the trip, on pain of a contradiction being true (namely, your both making the trip and not making it). Rather, it is only *because* you *will* make the trip—for whatever reason (forced or voluntary)—that your older self is there telling you about it in the first place. Your older self's tale is a report of what you will, as it happens, end up doing: the tale conforms to your future actions, rather than your actions having to conform to the tale. If you were *not* going to make the trip, then either you would have no visitor, or your visitor would be an imposter: she would not be your older self. As in the objects-with-dates case, the only way to see a

[21] This objection is due to Mat Ripley, as is the term 'forcing coincidence'.

problem here is to fall for the second-time-around fallacy: to suppose that your future self willingly hops in her time machine to pay her younger self a visit, but that then—the second time around, as it were—you are forced into the machine against your will.

6.4 Backward time travel without trying?

I have made use of the fact that it makes no sense to speak of someone trying in the *de dicto* sense to perform a task at which he feels certain he will fail. At this point, someone might assert that it is possible to construct an objection to backward time travel that makes no reference to *trying*—and in the face of such an objection, the type of argument employed earlier in the present section would be powerless. Consider a well-known case, due to John Earman: so far from involving trying, it does not even involve any creatures *capable* of trying to do things. Earman writes:

> consider a rocket ship which at some space-time point x can fire a probe which will travel into the past lobe of the null cone at x. Suppose that the rocket is programmed to fire the probe unless a safety switch is on and that the safety switch is turned on if and only if the 'return' of the probe is detected by a sensing device with which the rocket is equipped. Is the probe fired? We find that the answer is that it is fired if and only if it is not fired, which is a contradiction if standard logic holds (Earman [1972], pp. 231–2).

As it stands, this situation is easily accommodated in the style of Lewis. Either, by coincidence, the safety switch fails, and the probe fires even though its return has already been detected by the sensor; or, by coincidence, the sensor fails, and the probe fires after returning undetected; or, by coincidence, the sensor fails, and registers a return of the probe, even though the probe is not fired and did not return; or, by coincidence, the probe goes off course and fails to return to the rocket; and so on. One might wish to press the objection further, in the style of Horwich. In this case, in order to get problematic numbers of coincidences occurring, one would have to suppose that after each coincidental failure of the Earman system, things were set up again for another go. How would that happen? The system might just pop into existence each time—but in this case, the old argument would apply: improbable outputs can be derived only on the assumption that equally improbable inputs occur. The only other way to get the system set up again after each coincidental failure is to reintroduce *trying*—in which case the type of defence of backward time travel employed earlier would again become effective. It makes no sense to try in the *de dicto* sense to build and operate a device, such as Earman's, which one knows to be self-defeating. (No one who tries to build a perpetual motion machine thinks that such machines are impossible.) So the persons involved

must be trying, in the *de dicto* sense, to build and operate a system which is *not* self-defeating—so it is no coincidence that the systems they come up with frequently exhibit behaviour which would count as failure if exhibited by a self-defeating system (were there any such thing): that is, non-self-defeating behaviour. Occasionally the persons involved might try in the *de re* sense to do something such as load a probe into an otherwise fully-functional self-defeating device—but not often enough for the ensuing coincidences to even appear problematic.

7 Conclusion

The auto-infanticide objection to backward time travel should be permanently laid to rest. Lewis rebutted the classic version, and although Horwich did the best that could be done by way of resurrection of the objection, the most that Horwich's argument could show is that local backward time travel *will not* occur within the next few generations. Even to reach that weak conclusion, it would have to be established that counterfactual **A** is true and counterfactual **B** false. Now the essential difference between **A** and **B** is that **A** says that if the time traveller were to kill the child who lives at his old address (at time *t*), then *after t*, things would be different; while **B** says that if the time traveller were to kill the child who lives at his old address (at time *t*), then things would be different *before t* as well: the time traveller and the child would have distinct births, for a start. Given this difference between **A** and **B**, it is difficult to see how one could establish the truth of **A** without either appealing to, or else in the process establishing, a general thesis to the effect that there is something suspect about counterfactuals which say that an earlier time would be different if a later time were different; and if such a general thesis *were* available, then one could dismiss backward time travel at the outset, on the grounds that backward time travel involves backward causation, which the general thesis would (presumably) show to be inherently suspect, along with counterfactuals such as **B**. In this case one would not need to appeal to the auto-infanticide objection. Thus the latter objection cannot play a useful role in discussions of backward time travel. If counterfactual **A** is true then the objection shows something very limited which we already know, and is in any case redundant. If counterfactual **B** is true then the objection shows nothing.

Acknowledgements

I am very grateful to Amitavo Islam, David Lewis, Michael McDermott, Huw Price, Mat Ripley, Sybille Smith, Catherine Vidler, an audience at the 1996 Annual Conference of the Australasian Association of Philosophy (held in

Brisbane in July), and two anonymous referees for this journal for extremely helpful discussions and criticisms of earlier drafts of this paper.

Department of Philosophy
Princeton University
Princeton, NJ 08544–1006
USA

References

Brier, B. [1973]: 'Magicians, Alarm Clocks, and Backward Causation', *Southern Journal of Philosophy*, **11**, pp. 359–64.

Dwyer, L. [1975]: 'Time Travel and Changing the Past', *Philosophical Studies*, **27**, pp. 341–50.

Dwyer, L. [1977]: 'How To Affect, But Not Change, the Past', *Southern Journal of Philosophy*, **15**, pp. 383–5.

Earman, J. [1972]: 'Implications of Causal Propagation Outside the Null Cone', *Australasian Journal of Philosophy*, **50**, pp. 222–37.

Fulmer, G. [1980]: 'Understanding Time Travel', *Southwestern Journal of Philosophy*, **11**, pp. 151–6.

Gödel, K. [1949]: 'A Remark About the Relationship Between Relativity Theory and Idealistic Philosophy', in P.A. Schilpp (*ed.*), *Albert Einstein: Philosopher-Scientist, Vol. II*, (*Library of Living Philosophers, Vol. VII*), Evanston, IL, The Library of Living Philosophers Inc., pp. 555–62.

Horwich, P. [1975]: 'On Some Alleged Paradoxes of Time Travel', *Journal of Philosophy*, **72**, pp. 432–44.

Horwich, P. [1987]: *Asymmetries In Time: Problems In the Philosophy of Science*, Cambridge, MA, MIT Press.

Hospers, J. [1967]: *An Introduction to Philosophical Analysis*, 2nd edn, Englewood Cliffs, NJ, Prentice-Hall.

Kripke, S. A. [1980]: *Naming and Necessity*, Oxford, Blackwell.

Levin, M. R. [1980]: 'Swords' Points', *Analysis*, **40**, pp. 69–70.

Lewis, D. K. [1976]: 'The Paradoxes of Time Travel', *American Philosophical Quarterly*, **13**, pp. 145–52.

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

Price, H. [1996]: *Time's Arrow and Archimedes' Point*, New York, Oxford University Press.