I'd Do Anything to Change the Past (But I Can't Do 'That')

Nicholas J.J. Smith

21 December 2015

Abstract

This paper addresses a worry about backwards time travel. The worry is that there is something mysteriously inexplicable about the combination of commonplace events that will inevitably conspire to prevent the time traveller doing something impossible such as killing her younger self. The worry is first distinguished from other problems for backwards time travel, concerning its alleged impossibility or improbability. It is then shown that the worry is misplaced: there is in fact no real problem here. Yet the worry has been widely expressed—so a suggestion is also made as to why it is so easy to get into the position of thinking that there is a genuine problem here, when in fact there is not. Finally, in light of the resolution of the inexplicability worry, a new way of dealing with the other two problems for backwards time travel—concerning its alleged impossibility and improbability—is proposed.

1 Introduction

Backwards time travel gets a bad rap. It is often thought to be logically *impossible*. Failing that, it is often thought to be extremely *improbable*. Failing that, it is often thought to involve something mysteriously *inexplicable*. The main purpose of this paper is to address the third of these concerns—but in light of the resolution of this concern we can then also see a new way of dealing with the first two concerns.

Section 2 distinguishes the main argument to be addressed here—the inexplicability argument—from other arguments against backwards time travel: those according to which it is impossible or improbable. Section 3 addresses the inexplicability argument and concludes that backwards time travel does not involve any mysterious or puzzling kind of inexplicability. In light of this

discussion, section 4 then looks back at the other two kinds of argument—the impossibility and improbability arguments.

2 Isolating the Argument

2.1 Preliminaries

The problems we wish to discuss involve backwards, not forwards time travel—and they do not arise in models involving hypertime, multiple temporal dimensions, or parallel universes: the problems arise in models in which there is a single past. Therefore we shall confine our discussion to backwards time travel in a model in which there is a single past. Thus, for example, a time traveller who journeys to some event that she recalls from her childhood visits *the very same past* that she recalls: not some very similar (but numerically distinct) sequence of events in (for example) a parallel universe. The time traveller will be involved in *the very events* of which she has memories—not some duplicate set of events that differ from the events she recalls by the addition of her time travelling self.

In the popular imagination, backwards time travel would allow one to change the past: to right the wrongs of history, to prevent one's younger self doing things one later regretted, and so on. In a model with a single past, however, this idea is incoherent: the very description of the case involves a contradiction (e.g. the time traveller burns all her diaries at midnight on her fortieth birthday in 1976, and does not burn all her diaries at midnight on her fortieth birthday in 1976). It is not as if there are two versions of the past: the original one, without the time traveller present, and then a second version, with the time traveller playing a role. There is just one past—and two *perspectives* on it: the perspective of the younger self, and the perspective of the older time travelling self. If these perspectives are inconsistent (e.g. an event occurs in one but not the other) then the time travel scenario is incoherent.

This means that time travellers can do less than we might have hoped: they cannot right the wrongs of history. But this does *not* mean that time travellers must be entirely powerless in the past: they *can* (in principle) do anything that *did* happen. Time travellers cannot *change* the past: they cannot make it different from the way it was—but they can (for all we have said) *participate* in it: they can be amongst the people who did make the past the way it was.[1]

[1]For early statements of this crucial distinction between changing the past and participating in (aka affecting or influencing) the past see e.g. Brier [1973, 361], Horwich [1975, 435–6] and Dwyer [1977]. For further discussion see Smith [2013, §1.2].

2.2 Bananas

The most important objection to the logical possibility of backwards time travel is the so-called Grandfather paradox. This paradox has actually convinced people that backwards time travel is impossible:

> The dead giveaway that true time-travel is flatly impossible arises from the well-known "paradoxes" it entails. The classic example is "What if you go back into the past and kill your grandfather when he was still a little boy?"... So complex and hopeless are the paradoxes... that the easiest way out of the irrational chaos that results is to suppose that true time-travel is, and forever will be, impossible. [Asimov, 2003, 276–7]

The paradox comes in different forms. Here's one version:

> If time travel was logically possible then the time traveller could return to the past and in a suicidal rage destroy his time machine before it was completed and murder his younger self. But if this was so a necessary condition for the time trip to have occurred at all is removed, and we should then conclude that the time trip did not occur. Hence if the time trip did occur, then it did not occur. Hence it did not occur, and it is *necessary* that it did not occur. To reply, as it is standardly done, that our time traveller cannot change the past in this way, is a *petitio principii*. Why is it that the time traveller is constrained in this way? What mysterious force stills his sudden suicidal rage? [Smith, 1985, 58]

The idea is that backwards time travel is impossible because if it occurred, time travellers would attempt to do things such as kill their younger selves (or their grandfathers etc.). We know that doing these things—indeed, changing the past in any way—is impossible (recall section 2.1). But were there time travel, there would then be nothing left to stop these things happening. If we let things get to the stage where the time traveller is facing Grandfather with a loaded weapon, then there is nothing left to prevent the impossible from occurring. So we must draw the line earlier: it must be impossible for someone to get into this situation at all; that is, backwards time travel must be impossible.

In order to defend the possibility of time travel in the face of this argument we need to show that time travel is not a sure route to doing the impossible. So, given that a time traveller has gone to the past and is facing Grandfather, what could stop him killing Grandfather? Some science fiction authors resort to the idea of chaperones or time guardians who prevent time

travellers from changing the past—or to mysterious forces of logic. But it is hard to take these ideas seriously—and more importantly, it is hard to make them work in detail when we remember that changing the past is impossible.[2] Fortunately there is a better response—also to be found in the science fiction literature, and introduced into the philosophical literature by Lewis [1976]. What would stop the time traveller doing the impossible? She would fail "for some commonplace reason", as Lewis [1976, 150] puts it. Her gun might jam, a noise might distract her, she might slip on a banana peel, and so on. Nothing more than such ordinary occurrences is required to stop the time traveller killing Grandfather. Hence backwards time travel does not entail the occurrence of impossible events—and so the above objection is defused.

2.3 Improbability

So the Grandfather paradox can be resolved by the 'commonplace reasons' response. But from its ashes a new problem arises.[3] This new objection—due to Horwich [1987]—is not to the possibility but to the *probability* of backwards time travel.

Think about correlated events in general. Whenever we see two things frequently occurring together, this is because one of them causes the other, or some third thing causes both. Horwich calls this the Principle of V-Correlation:

> 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, 97–8]

For example, suppose that two students turn up to class wearing the same outfits. That could just be a coincidence (i.e. there is no common cause, and no direct causal link between the two events). If it happens every week for the whole semester, it is *possible* that it is a coincidence, but this is *extremely unlikely*. Normally, we see this sort of extensive correlation only if either there

[2]The chaperone is acting to ensure that the past remains as it was—but the only reason it ever was that way is because of his very actions. For further discussion of the kinds of issue that arise here see Smith [2005].

[3]A different problem that also remains after the Grandfather paradox has been resolved in the way indicated in the previous section is the 'can and cannot' paradox. For discussion of how this problem differs from the Grandfather paradox—and how it may be resolved—see Smith [2013, §2.1].

is a common cause (e.g. both students have product endorsement deals with the same clothing company, or both slavishly copy the same fashion blog) or a direct causal link (e.g. one student is copying the other).

Now consider the time traveller setting off to kill her younger self. As discussed, no contradiction need ensue—this is prevented not by chaperones or mysterious forces, but by a run of ordinary occurrences in which the trigger falls off the time traveller's gun, a gust of wind pushes her bullet off course, she slips on a banana peel, and so on. But now consider this run of ordinary occurrences. Whenever the time traveller contemplates autoinfanticide, someone nearby will drop a banana peel ready for her to slip on, or a bird will begin to fly so that it will be in the path of the time traveller's bullet by the time she fires, and so on. In general, there will be a correlation between autoinfanticide attempts and foiling occurrences such as the presence of banana peels—and this correlation will be of the type that does not involve a direct causal connection between the correlated events or a common cause of both. But extensive correlations of this sort are, as we saw, extremely rare—so backwards time travel will happen about as often as you will see two people wear the same outfits to class every day of semester, *without* there being any causal connection between what one wears and what the other wears.

We can set out Horwich's argument this way:

1. If time travel were ever to occur, we should see extensive uncaused correlations.

2. It is extremely unlikely that we should ever see extensive uncaused correlations.

3. Therefore time travel is extremely unlikely to occur.

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

I have elsewhere attacked both premisses of Horwich's argument.[4] I shall not discuss these responses here, however, for two reasons. First, my purpose at this point is not to resolve Horwich's problem but to distinguish it from the inexplicability objection to time travel that I wish to consider in detail in the present paper. In particular, the objection to be considered does

[4]See Smith [1997]. For further discussion of and references to the literature arising from Horwich's argument, see Smith [2013, §2.2.].

not hold that time travel scenarios are *unlikely* or *improbable*: it holds that they are problematically inexplicable. Second, recall that Horwich's problem arose out of the 'commonplace reasons' response to the Grandfather paradox. Later in this paper (section 4)—in light of the discussion of the inexplicability objection in section 3—I'll propose a better way of responding to the Grandfather paradox: one that never gives rise to Horwich's problem in the first place.

2.4 Initial Conditions

Return to the 'commonplace reasons' resolution of the Grandfather paradox (section 2.2) and consider another objection to time travel that arises from it—again, in order to distinguish that objection from the inexplicability worry about time travel to be considered in this paper. The time traveller fails to kill his younger self because he slips on a banana peel, and then a bird intercepts his bullet, and so on. Where do the peel, bird and so on come from? The peel was discarded by a passing chimp. Why was the chimp there? Because it had escaped from the zoo. How was it able do that? Because one of the bars on its cage was badly welded. Why was that? Because the welder knew a monkey would be needed to prevent a time traveller killing his younger self? Hardly. The welder was simply distracted by a passing airship. Why was it there? And so on. Eventually, tracing back the causes of the banana peel being there, the bird being there, and so on, we will get back to some specific initial conditions at or near the Big Bang. Now isn't it utterly mysterious that things were just like *that*? Won't the relevant initial data look utterly ad hoc, gerrymandered, inexplicable and/or improbable?

The issue of the (in)explicability and/or (im)probability of initial conditions is a general one that transcends considerations of time travel. For example, there is a debate on whether observed thermodynamic asymmetries require that the universe was initially in a very special state—the 'Past State'—and whether such an initial state is inexplicable and/or improbable.[5] Again, this kind of issue needs to be distinguished from the objection to time travel that I wish to consider in the present paper. In particular, the objection to be considered does not hold that time travel scenarios require initial conditions that are improbable and/or inexplicable: the problem is meant to arise in the time travel scenario itself—not just in the initial conditions that we reach if we trace back from the time travel scenario to the Big Bang.

[5]See e.g. Price [2002], Callender [2004a], Price [2004] and Callender [2004b]. Cf. also Horwich [1987, 120].

2.5 Inexplicability

We are now in a position to set out the problem for backwards time travel to be addressed in this paper. Here are a few quotations to give the flavour of the problem (presented in order of date of publication):

> Suppose I meet my former self... I draw a pistol. Can I shoot my old self? It seems not, for to do so would entail my having died before I did so. And this is absurd... What explanation is possible for my failure if I try?... [¶] we are faced with the problem of explaining why it is that I cannot fire the gun or, if I can, why it is that I can fire only in certain directions. Either the gun is not behaving as the normal physical object we take it to be, or the notion of voluntary action does not apply in the usual way. [Gorovitz, 1964, 366–7]

> there should remain a lingering doubt about whether the autofanticide problem, and its variants, have been adequately disposed of.[6] Such misgivings arise, I think, when we focus our attention on cases of *attempted* autofanticide—cases where someone *tries* to realize a self-defeating causal chain. We know of course that success is impossible. Yet repeated failure is nevertheless surprising and disturbing... [¶] Now we know, on the basis of purely semantic reasons, that attempts at [bringing about some past event that did not occur] will invariably fail. But we recognize that there is considerable strangeness in this—something ad hoc and unsatisfying about explaining the repeated failures in terms of changes of mind, guns misfiring, and so forth... it is implausible that such mishaps would occur so faithfully over and over again [Horwich, 1987, 119–21][7]

> Lewis gives Tim only one chance to murder his grandfather. However, Tim could attempt to confirm his opinion that he will suc-

[6] Horwich refers to the act of killing one's younger self as 'autofanticide'. The standard term in the literature is 'autoinfanticide'.

[7] Horwich here points to doubts and misgivings that (according to Horwich) we do and should feel when we think about attempts to change the past—even once we accept that changing the past is impossible. He goes on to present his argument for the *improbability* of time travel as a crystallisation of these misgivings. However, it is clear from my discussions with them—for example at the Workshop on Probability and Time Travel held at the University of Sydney in November 2014—that many philosophers who have thought carefully about these issues and who *accept* that Horwich's improbability argument is flawed nevertheless *still* feel the force of the initial worry as spelt out by Horwich in these quotations (i.e. *before* explicit considerations of *improbability* are introduced). In this section I am trying to bring out this residual worry about time travel.

> ceed eventually by arranging additional opportunities to shoot at his grandfather. If Tim had continued to try to kill his grandfather, it is clear that he must fail all attempts or a contradiction will ensue. Yet, failure on one occasion need not entail failure on other occasions. What's more, the reason Lewis offered for Tim's sole failure need not be applicable to any other, if repeated attempts are made. In other words, Lewis's account may do for a once only attempt, but is untenable as a general explanation of Tim's continual lack of success if he keeps on trying. [Riggs, 1997, 52]
>
> Considered individually, there will be nothing anomalous in the explanations [of why attempts to implement self-defeating causal chains fail]; they will have precisely the same form and status as explanation of failures in topologically ordinary cases. [¶] It is almost irresistible to suppose, however, that there is something anomalous in the cases considered *collectively*, i.e., in our *unfailing* lack of success [Ismael, 2003, 308]
>
> . . . the common concern that the real threat to the possibility of one-dimensional backward time travel is not that we could change the past, but that there would be surprising and inexplicable constraints on what a time traveler can do. [Carroll, 2010, 86]

The worry is as follows. The would-be autoinfanticidal time traveller is attempting to do something impossible (render herself permanently dead from an age younger than her age at the time of the attempts). We accept that she will not succeed. We also accept that what will stop her succeeding is a succession of commonplace occurrences (slips on banana peels and so on—as opposed to chaperones, time lords or mysterious forces of logic). We are bracketing any worry that such a succession is *improbable*. Also, we are not tracing back from this succession of occurrences to the Big Bang and reflecting on what the initial state of the universe must have been like in order for this succession to occur: we are just focussing on the time at which the failed autoinfanticide attempts are made. Yet still there is a problem (so the worry goes). The problem is that the exclusion of the time traveller from successfully committing autoinfanticide seems mysteriously inexplicable. Each particular event that foils the time traveller is explicable in a perfectly ordinary way; but the *inevitable combination* of these events amounts to a ring-fencing of the forbidden zone of autoinfanticide—and this ring-fencing is mystifying. It's like a grand conspiracy to stop the time traveller doing what she wants to do—and yet there are no conspirators: no time lords, no magical forces of logic. This is profoundly perplexing.

I'll argue in the next section that there is no mystery here at all.

3 Responding to the Argument

3.1 Other Cases

In order to cast light on the case of time travel, let's begin by considering some other kinds of case.

Terrace 1. You and your friend live in neighbouring terrace houses that have exactly the same floor plan. You cannot enter your kitchen—when you try to do so it is as if you run into a glass wall—but she can enter hers without any problem. In this case there is something mysterious afoot in your house that demands explanation—and we will not be satisfied until we find an explanation. This is not to say that there has to be a 'scientific explanation' rather than an explanation in terms of ghosts: the point is just that the phenomena cry out for *some* kind of systematic explanation.

Terrace 2. You and your friend live in neighbouring terrace houses that once had exactly the same floor plan—however she has extended her house to create an attic room and you have not extended yours (i.e. you have no attic room). Now if someone (who did not know the layout of your house and incorrectly assumed that you had an attic) requested you to go up to your attic and throw confetti from the window, you would not be able to comply: you *cannot* enter 'your attic room'. But your inability to do so is very different from your inability in Terrace 1 to enter your kitchen. There just *is no* attic room and so although (a) you *cannot* enter 'it' and (b) there is no systematic scientific or supernatural *explanation*, of the kind we sought in Terrace 1, of your inability (no ghosts, no unusual magnetic fields, etc.)—still there is no *mystery* here. The mere fact that you *have no* attic is enough to ensure that you cannot enter 'it': no (further) explanation of your failure is required. (It is not that there is some further explanation that could potentially be given, but we do not require it in this case for some reason—say, we do not care enough: the point is that there just isn't anything else relevant that could be said about your failure.)

Terrace 3. As in Terrace 2, you and your friend live in neighbouring terrace houses that once had exactly the same floor plan—however she has extended her house to create an attic room and you have not extended yours. Unlike in Terrace 2, however, suppose that you have been living in your friend's house for a year while she has been away and have just returned to your own house. You have become accustomed to having an attic room and have completely forgotten that your house lacks one. In fact, you have become convinced that you *do* have an attic room—just the same as your

neighbour's. Yet—from your point of view—whenever you try to enter it, something always stops you: either something distracts you from your attempt (e.g. the telephone rings or someone knocks at the front door), or you slip on a banana peel as you climb the stairs, or a wall appears where (according to you) there should be a door and you run into it, or you cannot recall exactly where the door to the attic is supposed to be, and so on. Now let's suppose you find this very perplexing and mysterious. When you try to enter your kitchen, bathroom or any other room there is no problem—and yet when you try to enter the attic you always fail. We—considering the case from the outside—can furthermore see that you will *invariably* fail: you will *never* succeed in entering the attic (because you have no attic to enter). To you, it seems as if there is a conspiracy to prevent you entering the attic—indeed a very powerful conspiracy, that *always* prevails—and yet a conspiracy of which you can find absolutely no evidence: you call in exorcists but they find no ghosts; you place security cameras around your house and then examine the footage, but it shows no-one moving doors or walls or placing banana peels; and so on. You find this deeply perplexing.

Now it is clear why you are puzzled in Terrace 3—and what we should do to remove the puzzlement. You are deluded: you think that you have an attic when in fact you do not. To help you, we need to remind you that you *have no* attic. Once you realise this, recent occurrences in your house will no longer seem in the slightest bit mysterious. The situation is thus very different from Terrace 1. What you need in Terrace 3 is not an explanation of why you can't get into the attic of a kind analogous to the explanation required in Terrace 1 of why you can't get into the kitchen: that is, the discovery of a systematic obstacle standing in the way of your doing what you want to do (whether the obstacle be physical forces or malicious gremlins). What you need is simply to be reminded that you *have no* attic room. Once you realise this, all the mystery vanishes. Recent occurrences no longer cry out for explanation: they now seem entirely ordinary and exactly the kinds of thing one might have expected, given that there just *is no* attic for you to enter.

Fountain 1. There is a fountain of youth located in Florida, but a deity—or the army, or whatever—does not want anyone to discover it. They scramble satellite navigation signals in the area, publish false maps, create roadblocks, sabotage the vehicles of explorers, disseminate frightening stories about giant creatures that live in the area, and so on. You are searching for the fountain of youth. Something always prevents you finding it: your vehicle breaks down; you spend the day walking in a large circle when you thought you were travelling in a straight line; and so on. You never—or only very rarely—encounter these kinds of issue on other expeditions: but

your attempts to find the fountain of youth are plagued by problems. From your point of view it is as if there is a conspiracy to prevent you finding the fountain of youth. In fact you are exactly right—there *is* a conspiracy—and if you were ever to discover it, your repeated failures would be explained and the air of mystery dispelled.

Fountain 2. Consider Juan Ponce de León, searching (what we now call) Florida for the fountain of youth. He did not find 'it'. But this was not because of a conspiracy to place obstacles in his way: it was because there simply *was no* fountain of youth there to be found. Once we realise this, there is no need for—and indeed no possibility of—any (further) explanation of his failure.

Shaving 1. A barber sets out to shave every man in Newcastle in one week. He fails. There must be some reason why he fails: equipment malfunction; the number of men in Newcastle combined with limitations on how fast he can shave; men refusing to allow themselves to be shaved; and so on.

Shaving 2. A Newcastle barber sets out to shave all and only Newcastle barbers who do not shave themselves (i.e. he wishes to shave every Newcastle barber who does not shave himself and to shave no Newcastle barber who does shave himself). He fails. This time, there is no question of *what stopped him* succeeding. Consider the following description:

(B) A Newcastle barber shaves all and only Newcastle barbers who do not shave themselves.

No possible scenario whatsoever fits this description. *Whatever* happens, the outcome will not be 'that'. And so we do not need an explanation of why 'that sort of outcome' was prevented. *Nothing prevents* such an outcome: the point is that there just *are no* such outcomes.

3.2 Lessons from the Cases

There are two points that arise from these cases on which I wish to focus.

First, there can be pairs of cases that are more or less similar in various ways, in each of which someone invariably fails to do something, but in only one of which the failure requires explanation—is perplexing in the absence of such explanation. For example, in Terrace 1, your inability to enter your kitchen requires explanation—and is perplexing in the absence of such explanation. In Terrace 2, your inability to enter 'your attic' demands no explanation and is not in the least perplexing. In Terrace 3, your inability to enter your attic perplexes *you* because you do not realise you lack an attic—but the mystery is dispelled not by giving an explanation of the kind sought in Terrace 1, but just by reminding you that you have no attic. Once

it is clear that there is no attic to enter, there is no need for an explanation as to why you fail to enter 'it' and nothing perplexing about the fact that you invariably fail to do so.

Second, the key difference between the two types of case is as follows. In both types, one invariably fails to enact a scenario satisfying a certain description (e.g. 'entering the attic', 'finding the fountain of youth', 'shaving all and only those who shave themselves'). In one type of case—the one where failure demands an explanation—there are contextually relevant possible scenarios satisfying the description. In the other type of case there are not. The clearest examples of the latter sort are where there are no possible scenarios at all satisfying the description (not just no contextually relevant ones). For example, in Terrace 2, there are *possible* scenarios (although they are not contextually relevant) in which you *build* an attic and then enter it. So I suppose one *might* ask for an explanation of why you do not do *that*. But in Shaving 2, there is *no possible scenario at all* satisfying the description (B). Hence there is no question of explaining why 'such scenarios' are not actualised. Beyond pointing out that there simply *are no* such scenarios—so *whatever happens*, one of 'them' will not be actualised—there is nothing more to say by way of explanation of the barber's failure.

3.3 Clarifications

At certain points I have used certain phrases. For example, I said that beyond pointing out that there are no scenarios satisfying the description (B) there is nothing more to say by way of explanation of the barber's failure—note the 'nothing more'—and I said that the mere fact that you have no attic is enough to ensure that you cannot enter 'it': no (further) explanation of your failure is required—note the 'further' in parentheses. My aim was to leave open two possibilities:

1. In cases like Terrace 2, Fountain 2 and Shaving 2, *no explanation* of failure (to enter 'the attic', to find 'the fountain of youth' etc.) is required.

2. In cases like Terrace 2, Fountain 2 and Shaving 2, the *only explanation* of failure (to enter 'the attic', to find 'the fountain of youth' etc.) that is required is pointing out that there *is no* attic, fountain etc.

In cases like Terrace 2, Fountain 2 and Shaving 2, there is certainly something that it may be helpful to point out: that there is no attic, that there is no fountain, that there is no scenario whatsoever that fits the description (B). (Terrace 3 is a case in which it is clearly helpful to point out this sort of

thing.) The question that I am leaving open is whether pointing out this sort of thing counts as an *explanation* of failure. One's answer to this will depend on one's theory of explanation: on some accounts of what an explanation is, pointing out that there is no attic does constitute an explanation of your failure to enter 'it' (and similarly for the other cases); on other accounts of what an explanation is, it doesn't. An example of the former is a pragmatic theory according to which an explanation is a satisfying answer to a 'Why?' question—for it can certainly be a satisfying answer to the question 'Why can't Bob enter his attic?' to be told that Bob has no attic. (This is the kind of thing that happens in Terrace 3.) Another example of the former is a theory according to which to explain something is to show that it should have been expected—that it should not have been surprising. For once we realise that Bob has no attic, we should certainly expect that he will not enter 'it' and should not be surprised when he fails to do so. An example of the latter is a theory according to which to explain an event is to say what caused it—for the *absence* of an attic is not part of any causal process leading up to any of the events that occur in Bob's (attic-free) house. However, for purposes of this paper there is no need to engage with the large and vexed literature on explanation: I can simply leave open whether I am making the first claim above (that *no* explanation of failure is required in certain kinds of case) or the second (that no *further* explanation is required, beyond pointing out that there is no attic, no fountain, no scenario that fits the description (B), and so on). Nothing of substance will depend on which way one goes here: only questions of wording—which is why I have chosen non-committal formulations such as 'no (further) explanation'.

At this point I should ward off a potential confusion. Consider the observation that no scenario whatsoever fits the description (B). I have allowed that we might want to view this observation as an explanation of the barber's failure in Shaving 2—if, for example, we adopt a certain kind of pragmatic theory of explanation—and that we might not want to view it as an explanation—if, for example, we adopt a certain kind of causal theory of explanation. But we must not confuse these two views of explanation (or, if one prefers to put it so, these two kinds of explanation). I certainly do *not* want to allow that we might view the fact that no scenario fits the description (B) as a *causal* explanation of the barber's failure. There is a mistaken picture that can be brought to mind by formulations such as 'The barber fails because it is impossible for a Newcastle barber to shave all and only Newcastle barbers who do not shave themselves'. The picture is this: the impossibility of success explains the barber's failure in Shaving 2 in just the same kind of way that (say) a certain kind of force field explains your inability to enter your kitchen in Terrace 1. This picture is confused. In Terrace 1, there *is* a

kitchen there, and something—a force field, let us say—stops you entering it. If the force field weren't there, you would enter. In Shaving 2, on the other hand, there just *are no* scenarios satisfying the description (B). Hence there is no question of *what stops you* enacting 'such a scenario'. The point is that *whatever happens*, it won't be 'that' (i.e. a Newcastle barber shaving all and only Newcastle barbers who do not shave themselves)—because there *is no* 'that'. Thus the *right* way of understanding the claim that 'such scenarios are impossible' is as pointing out that there just *aren't any* 'such scenarios' (the descriptions of 'them' involve contradictions and so describe nothing at all). The *wrong* way of understanding the claim is as implying that there are such scenarios but they are surrounded by a force field—the force of the law of noncontradiction, as it were—and if only the force field weren't there, the barber *would* enact such a scenario.

Thus, while I am leaving it open whether pointing out that there are no scenarios that fit the description (B) counts as an *explanation* of the barber's failure in Shaving 2, I am *not* leaving it open whether, if it *is* an explanation, it might be the kind of causal explanation that we seek in Terrace 1. If it is an explanation at all, it is a different kind of beast: one that consists not in saying how we are kept out of a certain zone (the kitchen, or the zone of possibilities in which such-and-such occurs)—but in pointing out that there just *is no* such zone at all (there is no attic, there are no scenarios satisfying a certain description).

3.4 Time Travel

Now recall the worry about time travel isolated in section 2.5. The exclusion of the time traveller from successfully committing autoinfanticide seems mysteriously inexplicable. Each particular event that foils the time traveller is explicable in a perfectly ordinary way; but the inevitable *combination* of these events amounts to a ring-fencing of the forbidden zone of autoinfanticide—and this ring-fencing is mystifying. It's like a grand conspiracy to stop the time traveller doing what he wants to do—and yet there are no conspirators: no time lords, no magical forces of logic. This is profoundly perplexing.

It should now be clear what to say about this worry: it isn't a sound one! This is a case of the type in which no (further) explanation of failure is required. There are *no scenarios at all*—no points in logical space—satisfying the description 'a time traveller commits autoinfanticide'. There *is no* forbidden zone and hence no need or even possibility of an explanation of why the time traveller does not enter 'it'. *Whatever happens*, it won't be autoinfanticide because *no scenario at all* satisfies that description. The reason for this is that the description is self-contradictory (e.g. it involves the time traveller

permanently dying at 20 and also being alive at 40). So the crucial point here is that there *is no* forbidden zone. This is completely different from saying that there is one, but 'laws of logic' prevent us entering it. That would be to think of the time travel case as being like Terrace 1—in which failure *is* mysterious. The key point is that the purported descriptions of scenarios in the forbidden zone (involving autoinfanticide—or indeed *any* successful changing of the past) are all self-contradictory and describe *no sequence of events at all*. Hence there just is no forbidden zone and thus no mysterious ring-fencing of 'it': there is *nothing to ring-fence*.

Consider the following quotation from Horwich:

> time travel into the local past engenders closed causal chains that verge on being self-defeating, and this possibility can be precluded only by means of appropriate coincidences. [Horwich, 1987, 125]

This perfectly illustrates the mistaken line of thought: that there is a kind of possibility (this is the term Horwich uses in the passage just quoted)—an area of logical space—that we are kept away from only by coincidences. That *would* be mysterious! It would be like Terrace 1. The resolution of the mystery lies in seeing that there just is no such forbidden zone: there are no such scenarios. And so there is no mystery as to what keeps us out of 'it'. The case of time travel is thus like Terrace 3, not Terrace 1. Or more precisely, it is like Shaving 2: not only are there no contextually relevant possible scenarios in which the time traveller commits autoinfanticide: there just are no such scenarios *at all*.

3.5 Why the Worry?

So the time travel case is like the case of the Newcastle barber who tries to shave all and only Newcastle barbers. Failure is not perplexing or mysterious. There is literally *no way* not to fail: the description of success is self-contradictory and describes *nothing at all*. Hence there is no need for—or even possibility of—an explanation (beyond pointing out that there is literally no way not to fail) of why failure invariably occurs. Yet the time travel case *feels* more puzzling—and indeed has been a persistent puzzle in the literature (recall the quotations in section 2.5). Why is this? If we cannot explain why the time travel case has in fact puzzled so many people (even though, for the reasons given, it should not have)—while the barber case has not—then we haven't really got to the bottom of the issue.

The reason for the puzzlement, I think, is that it *seems so easy* to envisage successful autoinfanticide (whereas we have no picture at all of what it would be like for the barber to succeed: that would quite evidently involve him

shaving himself and not shaving himself). And once we do *that*—imagine there to be a space of scenarios in which autoinfanticide occurs—then we are in the situation where this space of scenarios is ring-fenced and the mystery begins. The key to dissolving the mystery is to remember that there just *are no* such scenarios. So why do we think we can envisage 'them'? Well, we do it bit by bit. First we picture the time traveller visiting her younger self. No problem there. Then we picture the time traveller shooting the person before her. Again, no impossibility there (there are possible scenarios of which this image forms a part—for example, possible scenarios in which the person killed is not the time traveller's younger self, or possible scenarios in which the person killed is the time traveller's younger self but in which she is later resurrected—i.e. in which she is not rendered *permanently* dead by the killing). The *delusion* lies in thinking that the two pictures put together depict a coherent autoinfanticide scenario. They do no such thing. Either they depict a scenario that does not involve changing the past (e.g. the time traveller kills her younger self but is later resurrected) or they depict nothing at all—because the picture that we get by combining the two individual pictures is *self-contradictory*: it shows the time traveller dying (permanently) at age 20 *and also* being alive beyond the age of 20 (growing up to become a time traveller). This no more depicts any particular scenario than does any other description that includes both some claim and its negation.

So there are no successful autoinfanticide scenarios—hence no mysterious ring-fencing of them and no need for any explanation of why 'they' are not actualised. The reason this fact is hard to keep in plain sight—the reason the worry about time travel persists—is that it is so easy to delude ourselves into thinking that we have pictured scenarios in the forbidden zone. Plenty of films and books seem to picture such scenarios very clearly! But in fact, while *pieces* of these pictures depict parts of possible scenarios, the whole picture—obtained by stitching the pieces together—contradicts itself and depicts no particular scenario at all.

Consider Escher's famous picture 'Ascending and Descending'.[8] It shows a staircase in four straight sections which join to form a square. Looking at each section individually, there is no problem: we see two lines of men passing one another, one lot ascending and one lot descending, each man keeping to the left side of the stairs (from his point of view). Looking at each corner individually, there is again no problem: the ascending men turn right and continue their ascent and the descending men turn left and continue their descent (each man still keeping to the left side of the stairs, from his

[8]John Bigelow also discussed this picture, in his paper at the Workshop on Probability and Time Travel mentioned in n.7.

point of view). As a whole, however, the picture does not represent any possible situation: there is a closed loop of ascending men, each of whom is below the man in front of him, and a closed loop of descending men, each of whom is above the man in front of him. Now suppose we give this picture to some master builders and ask them to use it as a plan to build such a set of stairs. They will invariably fail. But this is not mysterious and does not require explanation. The picture depicts *no possible object at all*, and hence *whatever they build*, it will not be 'that'.

The time travel case is precisely analogous. Our descriptions or films of autoinfanticide scenarios (or of less dramatic cases of changing the past) are like Escher's picture. A time traveller who tries to realise such a description invariably fails. There is no mystery here: the description as a whole contradicts itself and depicts no possible scenario at all, hence *whatever happens—whatever* the time traveller does—it will not be 'that'. The only obstacle to seeing this is the failure to realise that the time travel stories we know and love from film and fiction (or at least the ones that involve changing the past)[9] are self-contradictory (for the reasons discussed in section 2.1, where we noted that changing the past is impossible) and hence do not depict possible scenarios at all. It seems so easy for the time traveller to commit autoinfanticide! He'd just have to *pull the trigger*—a tiny movement of the finger, that's all! The problem is that, *in conjunction with other pieces of the story* (which specify that he is facing his younger self, and so on) this final piece creates a whole story that is self-contradictory—and hence depicts no possible scenario at all.

4 Solving the Grandfather Paradox: You're Doing it Wrong

In light of the discussion in section 3, we are now in a position to see that the entire discussion of the Grandfather paradox got off on the wrong foot.[10] Had the original 'commonplace reasons' response to the paradox been phrased better, a host of subsequent problems would simply never have arisen. The key mistake was to speak as if there *are* such things as autoinfanticide scenarios and then ask why such scenarios do not get actualised: it was to ask a question along the lines of 'What stops him?'—that is, to ask what prevents the time traveller killing his younger self (or succeeding in other attempts to change the past). The question Lewis asks, in introducing the 'common-

[9]Not all do. For example, the film *12 Monkeys* does not—and so it *does* depict a possible scenario.

[10]I don't just mean my own discussion in this paper: I mean the discussion in the literature. The way that I set out the problems earlier in this paper reflects the course that the literature has taken.

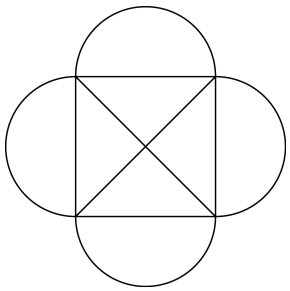
place reasons' response, is 'Why not?'—i.e. why does the time traveller not kill Grandfather? His answer is "For some commonplace reason. Perhaps some noise distracts him at the last moment, perhaps he misses despite all his target practice, perhaps his nerve fails, perhaps he even feels a pang of unaccustomed mercy" [Lewis, 1976, 150]. But questions such as 'Why not?' and 'What stops him?' are like the infamous 'Have you stopped beating your wife?' They have a false presupposition—in the time travel case, that there is such a thing as success and hence an explanation of failure is required. However, the point of the preceding discussion has been that because there is no way for the time traveller to succeed, no explanation of failure is called for. Because these questions have a false presupposition, no entirely satisfactory answer can be given. Suppose we answer 'What stops him are chaperones or time lords'. This is clearly unsatisfactory. So suppose we answer, as Lewis does: what stops the time traveller are commonplace occurrences such as slips on banana peels and gun jammings. Then we engender both Horwich's improbability objection, and the inexplicability objection that we have been discussing in this paper. For both objections begin by picking on the commonplace reasons answer to the 'What stops him?' question: one says that the answer involves something improbable and the other says that it involves something mysteriously inexplicable. Finally, if we give as our answer 'Nothing stops him' then it sounds as though we are allowing that the time traveller *can* change the past—which is no good either, because that is impossible.

A better way to proceed is to reject the question outright. We do not have to answer the question 'What stops him?' because the description of what he is trying to do—commit autoinfanticide or change the past in some other way—is self-contradictory and describes nothing at all. Hence *whatever happens*, it won't be 'that'. The key point is that there *is no* 'that'—and hence no question of what stops 'it' occurring. When we ask 'what stops the time traveller committing autoinfanticide?' we give legitimacy to the thought that there *is* such a thing as autoinfanticide—that is, that there are scenarios fitting this description. For we are asking: what keeps the time traveller from enacting such a scenario. We are asking for something such that, had it *not* happened, the time traveller *would* have committed autoinfanticide! But there *are no* scenarios that satisfy the description 'the time traveller commits autoinfanticide'. Once we are clear on that, the question of what stops him enacting 'such a scenario' evidently cannot be answered. The right response is not 'commonplace reasons' or 'nothing': it is to point out that there simply are no such scenarios.

In general, the wrong way to think about attempts to perform impossible tasks is to think that there is this thing that one cannot do. The right way is to think that *nothing one can do* counts as completing the task:

whatever one does it will not be 'that' because there *is no* 'that'. It might be helpful here to consider another kind of case where someone tries to do the impossible.[11] It is impossible to draw Figure 1 without lifting pen from paper and without retracing a line one has already drawn. The impossibility

Figure 1:



can be proven mathematically: considered as a graph, the figure is 'non-Eulerian'.[12] However, many people do not realise that the task is impossible and will spend hours trying to succeed. Now suppose Bob is one such person: he has a pen and a stack of paper and he is trying again and again to draw the figure without lifting pen from paper and without retracing a line he has already drawn. Suppose we ask 'What stops him succeeding?' One answer we'll certainly not want to give is that chaperones or math lords stop him. What about the following two kinds of answer?

- commonplace occurrences: his pen runs out of ink, he falls asleep, he runs out of paper, he is distracted by a phone call...
- nothing

There's an element of truth in both of them—but they are also both apt to engender a mistaken picture of the situation. The 'commonplace occurrences' answer can make it seem as though had these commonplace things not happened, Bob would or might have succeeded—and yet we know that success is impossible, and so then the occurrence of the commonplace events can start to seem strangely improbable or inexplicable. After all, when Bob tries

[11]Here—in talking of doing 'the impossible'—I myself slip into the familiar mode of talking as if there is an impossible thing that the person is trying to do—whereas my point is that the key to thinking clearly about this kind of case is to realise that the description of what the person is trying to do actually describes nothing at all.

[12]For an explanation of what this means and a proof of the impossibility, consult an introductory textbook on graph theory such as Wilson [1996].

to draw other kinds of picture, he never falls asleep or runs out of paper! The 'nothing' answer can make it sound as though Bob succeeds!—but of course success is impossible, so that is no good. And yet, there is something right about both answers. It is true that only quite commonplace things will happen: no math lords will appear on the scene. It is also true that nothing stops Bob: but that is because nothing is *required* to stop him. The right way to think of things is this. Bob can do whatever he likes. He can move his pen left or right, up or down, forwards or backwards, at whatever speed he likes and pressing as hard or as soft as he likes. He can think, he can pray, he can meditate. He can eat or drink whatever he likes. And so on. The key point is that *whatever he does*, the result will not satisfy that statement of the problem (i.e. it will not be true that he has drawn the figure without lifting pen from paper and without retracing a line he has already drawn). This is because there *is no solution* to the problem: there simply *is no way* of satisfying the statement; there *does not exist* a certain kind of route through the figure (a 'trail'—this is what the mathematical proof establishes; cf. n.12). So it is misleading to ask: What stops him succeeding? This is misleading because it presupposes that there *is* such a thing as 'a way of succeeding' and we need an answer as to why Bob doesn't find it or stumble upon it—we need to know what keeps him from it. A better way to put things is to say: we need no explanation of Bob's failure because there simply is *no such thing* as the solution he takes himself to be seeking.

The same point applies in the time travel case. It is misleading to ask 'What stops the time traveller?' or 'Why does he fail?' because it tends to encourage the thought that there *is* such a thing as 'a way of succeeding' and we need an answer as to what keeps the time traveller away from it. A better way to put things is to say that the time traveller can do *whatever he likes*: fire guns, drop bombs, and so on. Still, *whatever happens*, it won't be the case that he commits autoinfanticide: because the description of autoinfanticide is self-contradictory and describes no scenario whatsoever. There *is no scenario* that satisfies the description 'the time traveller commits autoinfanticide' (and similarly for other descriptions that involve the time traveller changing the past). So 'nothing' and 'commonplace occurrences' are both, in a way, correct answers to the question 'What stops him?'—but a much less misleading response to the question is to point out that the question itself is out of place, that it really should not be asked at all. There simply *is nothing* for him to succeed at—there is no such thing as a scenario that satisfies the description 'autoinfanticide'—and so there is no question as to why he fails to do 'that'.

I should make one clarification. I am suggesting that, rather than try to answer it, we should reject the question 'Why did X not occur?' ('Why does

the time traveller fail?', 'What stops him?' etc.) as out of place when the X in question is autoinfanticide or some other example of changing the past. I am not suggesting that we should always reject such questions, whatever the X in question is. Sometimes the question is a perfectly good one—but in such cases it will also have a perfectly good answer. For example, suppose the time traveller tries to fire his gun and it jams. We can ask 'Why did the gun not fire?' The answer will be: because it jammed. It also makes sense to ask why the gun jammed. The answer will be (say): because it was dirty. We can then ask how it got to be so dirty—and so on. These are all perfectly reasonable questions and will have perfectly good answers. In general, where Y is something that did happen but might not have, we can ask 'Why did Y happen?' and where X is something that did not happen but might have we can ask 'Why didn't X happen?' I am only suggesting that we should reject such questions when, respectively, it is impossible for Y not to occur or impossible for X to occur. In that case, the key point is that there *is no scenario* satisfying the description 'Y does not occur' or 'X occurs' and so there just *is no question* of why 'such a scenario' fails to occur. When the time traveller attempts autoinfanticide and someone asks 'Why did it not occur?' or 'What stopped it occurring?' the correct response is:

> What do you mean why did 'it' not occur? There is no such 'it'! Descriptions of 'it' involve contradictions and really describe nothing at all. There is no question why what happened was not 'that' (i.e. autoinfanticide): there just is no such thing as 'that'.

The question presupposes that there are 'success scenarios' and calls for an explanation of why one of them was not actualised. But there are no such scenarios—and so the question is out of place.

In sum, instead of answering the 'What stops him?' question with 'commonplace reasons', we should reject the question altogether as being out of place. And if we do that, then the further problems engendered by the commonplace reasons answer to the question—for example, Horwich's improbability objection, and the inexplicability objection considered in this paper—simply never arise.

5 Conclusion

I have isolated a particular worry about backwards time travel in models with a single past, distinguishing it from other (alleged) problems such as that backwards time travel is impossible or improbable. The worry begins with the acceptance that a time traveller who tries to kill her younger self

will fail and that it will be commonplace reasons (not time lords or magical 'forces of logic') that stop her. The worry is that the invariable occurrence of such foiling events is mysteriously inexplicable. I have argued that there is in fact no mystery here. The key is to realise that there is no forbidden zone—containing scenarios involving autoinfanticide or other successful attempts to change the past—and hence no question as to what keeps the time traveller out of such a zone. Descriptions of autoinfanticide—and of changing the past more generally—involve (sometimes well hidden) self contradictions. So whatever happens, it won't be 'that': it won't be 'the time traveller killing her younger self'—because that description is ultimately self-contradictory and hence describes no scenario at all. Thus there *is no* forbidden zone, and so the time traveller's failure to enter 'it' requires no explanation and is entirely lacking in mystery.

So why does this worry about time travel persist? Because we can put together descriptions of autoinfanticide (or changing the past more generally) piece by piece—where each piece is coherent. We are thereby deluded into thinking that we can picture a certain kind of scenario: a kind which then belongs in the forbidden zone—and then the mystery arises of how this zone gets to be ring-fenced, given that there are no time lords or magical forces of logic. We are like the character in Terrace 3, who is deluded into thinking there is an attic (and hence wonders what prevents it from being entered). We think: all the time traveller would have to do to commit autoinfanticide is make a tiny movement!—pulling the trigger, or pressing the button. And then it looks as though there is a mysterious conspiracy to prevent him doing so: a perplexing ring-fencing. The key to dissolving the worry is to realise that the complete descriptions of autoinfanticide (and other cases of changing the past)—when all the individually coherent parts are stitched together—*contradict themselves* and hence describe nothing. So there is no question as to why 'what they describe' does not occur. They don't describe anything—and it is entirely obvious and unremarkable that no time traveller (or anyone else) can do 'that'.

Acknowledgments

Thanks to the anonymous referees for very helpful comments.

Department of Philosophy
University of Sydney
NSW 2006 Australia

References

Isaac Asimov. *Gold: The Final Science Fiction Collection*. Harper Voyager, 2003.

Bob Brier. Magicians, alarm clocks, and backward causation. *Southern Journal of Philosophy*, 11:359–64, 1973.

Craig Callender. Measures, explanations and the past: Should 'special' initial conditions be explained? *British Journal for the Philosophy of Science*, 55 (2):195–217, 2004a.

Craig Callender. There is no puzzle about the low-entropy past. In Christopher Hitchcock, editor, *Contemporary Debates in Philosophy of Science*, pages 240–55. Blackwell, 2004b.

John W. Carroll. Context, conditionals, fatalism, time travel, and freedom. In Joseph Keim Campbell, Michael O'Rourke, and Harry S. Silverstein, editors, *Time and Identity*, pages 79–93. MIT Press, Cambridge MA, 2010.

Larry Dwyer. How to affect, but not change, the past. *Southern Journal of Philosophy*, 15:383–5, 1977.

Samuel Gorovitz. Leaving the past alone. *Philosophical Review*, 73(3):360–371, 1964.

Paul Horwich. On some alleged paradoxes of time travel. *Journal of Philosophy*, 72:432–44, 1975.

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

J. Ismael. Closed causal loops and the bilking argument. *Synthese*, 136: 305–20, 2003.

David Lewis. The paradoxes of time travel. *American Philosophical Quarterly*, 13:145–52, 1976.

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

Huw Price. Boltzmann's time bomb. *British Journal for the Philosophy of Science*, 53(1):83–119, 2002.

Huw Price. On the origins of the arrow of time: Why there is still a puzzle about the low-entropy past. In Christopher Hitchcock, editor, *Contemporary Debates in Philosophy of Science*, pages 219–39. Blackwell, 2004.

Peter J. Riggs. The principal paradox of time travel. *Ratio*, 10:48–64, 1997.

Joseph Wayne Smith. Time travel and backward causation. *Cogito*, 3:57–67, 1985.

Nicholas J.J. Smith. Bananas enough for time travel? *British Journal for the Philosophy of Science*, 48:363–89, 1997.

Nicholas J.J. Smith. Why would time travellers try to kill their younger selves? *Monist* (Special Issue on Time Travel), 88:388–95, 2005.

Nicholas J.J. Smith. Time travel. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Winter 2013 edition, 2013.

Robin J. Wilson. *Introduction to Graph Theory*. Longman, fourth edition, 1996.