THEODORE SIDER

TIME TRAVEL, COINCIDENCES AND COUNTERFACTUALS

(Received 3 May 2002)

ABSTRACT. In no possible world does a time traveler succeed in killing her earlier self before she ever enters a time machine. So if many, many time travelers went back in time trying to kill their unprotected former selves, the time travelers would fail in many strange, "coincidental" ways, slipping on banana peels, killing the wrong victim, and so on. Such cases produce doubts about time travel. How could "coincidences" be guaranteed to happen? And wouldn't the certainty of coincidental failure imply that time travelers are not *free* to kill their earlier selves? But if so, what would inhibit their freedom? Despite initial appearances, these and other doubts may be answered: the possibility of time travel survives yet another challenge.

1. FAILURES IN AUTOINFANTICIDE

Imagine you had a time machine. Nothing would stop you from going back in time and killing yourself as an infant, before you ever entered the time machine. But then a contradiction would ensue: you would never have entered any time machine since you were killed before doing so (let "killing" be understood throughout as implying permanent death), and yet you would have entered a time machine, in order to travel back in time to kill yourself. Some conclude that time travel is impossible, since it would lead to this contradiction.

There is nothing special about autoinfanticide: similar problems arise whenever a time traveler resolves to go back in time and do something that did not in fact occur. A time traveler who remembers owning a 1974 Plymouth Gold Duster could, it would seem, go back into the past and prevent herself from ever owning such a fine automobile; a time traveler could, it would seem, go back and prevent Lincoln from giving the Gettysburg address, and so on. But autoinfanticide is an especially vivid example.

As it stands, this argument is very weak. All it shows is that autoinfanticide is impossible, as are related scenarios, such as one

Philosophical Studies **110:** 115–138, 2002.

in which an address is given but in which someone travels back in time and prevents that address from being given. The impossibility of a certain kind of time-travel scenario does not impugn the possibility of time travel in general, any more than the existence of an impossible story about an empty box containing a figurine impugns the possibility of boxes.

We have admitted the possibility of time travel, though not the possibility of autoinfanticide. But these possible time travelers who do not kill their earlier selves: some have the desire as well as the means. What stops them?

No one thing. Some have a sudden change of heart. Some fear awful forces they think would be unleashed by a violation of the laws of logic. Some attempt the deed but fail for various reasons: non-lethal wounds, slips on banana peels, and the like. Others succeed in committing a murder, only to find they killed the wrong person. And there are possible worlds in which time travelers are shackled by gods, or are by other means prevented from doing mischief, though surely this is not required for time travel to occur.

But focus now on cases in which time travelers are not shackled in ways we do not take ourselves to be shackled. These time travelers would then have the *ability* to do the sorts of things we could do, in their circumstances. If I, who do not travel back in time, had a gun, had the evil desire to kill, and were suitably positioned near an unprotected victim, I would have the ability to kill that victim. So a time traveler relevantly like me could likewise kill her victim. But the time traveler's victim is her earlier self, and surely the time traveler cannot kill her earlier self, since contradictions would be true if she did. Thus, this argument concludes, unless time travelers are strangely shackled by gods or whatnot, time travel is impossible. An unshackled time traveler would both have and lack the ability to kill her earlier self.

Paul Horwich (1975) and David Lewis (1976) have defended time travel against this argument by objecting to the claim that the time traveler would not be able to kill her earlier self. Forget time travel for a moment and focus on ordinary cases of action. My having the ability to do A in a world, w, does not require that my doing A is consistent with every other fact about w. If I in fact will not do A, it is a fact about w that I will not do A (let us set aside

philosophies of time according to which the future is "open"); but no one other than a fatalist thinks this undermines my ability to do *A*. For Lewis, one has the ability to do *A* in *w* if one's doing *A* is consistent with the *relevant* facts about *w* (where what facts count as relevant varies according to the context of the speaker);[1] that the time traveler's victim is the time traveler herself is (typically) not a contextually relevant fact.

We have arrived at the following familiar position: despite the impossibility of autoinfanticide, time travel is possible. Moreover, though time travelers *do not* kill their earlier selves, they typically have the ability to do so. This position is, I believe, correct. But consider the following challenge. In an interesting thought experiment due to Horwich (1987, Chapter 7; 1995), time travelers *repeatedly* go back in time with the goal of killing their former selves. Imagine a futuristic Institute for Autoinfanticide sending out legions of assassins. (Perhaps these assassins have been emboldened by the failures of repeated attempts on their lives in their childhoods and fear nothing, not even the rumored cataclysmic destruction of the world that would result from a violation of the laws of logic.) Since autoinfanticide is impossible, each assassin fails. Some change their minds, others slip and fall on banana peels, yet others kill the wrong target, and so on. But there is something odd about the idea that such "coincidences" are *guaranteed* to happen, again and again!

There are, in fact, a few different arguments in the vicinity. While some have been made explicitly in the literature, I suspect others to lie below the surface of continuing inchoate resistance to time travel. Each is seductive, but none, I think, succeeds in undermining its possibility or likelihood. The purpose of this paper is to present and then rebut these arguments.

2. WHY BOTHER?

Before expending too much energy on the topic, it is worth thinking a bit about its point. Beyond the (perfectly legitimate) desire to set the record straight, is there any reason to care about time travel?

The most straightforward reason to care is that today's physics community cares. Whether the actual laws of nature permit time

travel is a live debate in contemporary physics journals (see Earman, 1995). Suppose the arguments to be discussed in this paper against the possibility of time travel (without shackles) succeeded. Given that many physicists tell us otherwise, that would be a problem! Whatever else metaphysicians must do, they should at least try to make metaphysical sense out of what physicists take seriously.

Secondly, time travel is tied up with larger issues in metaphysics and philosophy of science concerning the direction of time, causation, and so on. The possibility of time travel limits the space of acceptable theories in these areas.

Finally, time travel is connected with important issues in the philosophy of persistence. I have argued elsewhere (2001, Chapter 4, Section 7) that the possibility of time travel undermines "three-dimensionalism", the view that objects persist over time by being "wholly present" or "enduring", rather than by "perduring", i.e., persisting by means of temporal parts.

3. THE IMPROBABILITY OF COINCIDENCES

Having resolved to care about time travel, let us consider what arguments might be based on Horwich's thought experiment. There is first the argument that Horwich himself advances (1987, Chapter 7). Repeated attempts at autoinfanticide would lead to repeated "coincidental" failures – slips on banana peels, failures of nerve, etc. But we have empirical reason to think that such repeated coincidences do not occur. We do notice the odd slip on a banana peel *en route* to a murder, but such slips are rare indeed. Since we have strong inductive evidence against the existence of a rash of coincidences of this sort, we have reason to think that time travel into the recent past does not occur.

The argument establishes at most that we have *defeasible* reason to believe that time travel into the recent past does not *actually* occur. The argument concerns only the actual world because the evidence against coincidences is contingent; clearly strings of coincidences *might* have occurred. The argument provides only defeasible evidence because the evidence is inductive: the future existence of strings of coincidences is logically compatible with

our present evidence. The argument does not even establish that we have reason to think that time travel is prohibited by the actual laws of nature; a rash of coincidences would apparently not violate any law. Thus, the argument has little impact on much of the philosophical interest in time travel. For even if time travel is *unlikely* to occur in the actual world, if time travel is nevertheless possible – even, perhaps, physically possible – we have no right to ignore it while theorizing about the nature of persistence, time, causation, and so on. (This is no criticism of Horwich, who is not trying to undermine the possibility of time travel; indeed, he defends that possibility.)[2]

Moreover, the argument does not show that time travel *per se* is unlikely, for time travel might well occur without the formation of an Institute for Autoinfanticide. Large numbers of attempts at autoinfanticide and the like would result in large numbers of "coincidences"; but for all the argument shows, there might well be thousands of time travelers among us today, avoiding the banana peels and annoying pricks of conscience simply because they have no interest in "changing the past". Alternatively, perhaps the advent of time travel is so far in the future that time travelers consider us ancient history not worth bothering with.

Moreover, the kinds of coincidences envisioned here seem to require a *person* traveling back in time with the *intention* of doing things inconsistent with what she takes to have happened in the past (or at least a person sending some other object back in time, for example a message with instructions to kill her earlier self).[3] Quantities of unthinking time-traveling particles from the future going about their random business would not be particularly likely to exhibit "coincidental" patterns noticeable to us. The more cynical might accept the conditional "if time machines will one day be invented then there now exist numerous time-traveling assassins hunting down their ancestors", but an analogous conditional for electrons is implausible. Thus, Horwich's argument at best concerns the likelihood of future persons traveling back in time.

Horwich acknowledges some of these limitations of his argument, but, as Nicholas Smith has pointed out (1997, Section 4), there is another limitation he does not acknowledge. Our present evidence does not make it unlikely that in the future we will encounter time

travelers with attendant strings of "coincidences", since the present absence of coincidences does not seem to be projectible. To date, we have observed no rash of coincidences. But this observation does not warrant postulating any *law of nature*, only a certain *matter of particular fact*: that no legion of assassins has descended upon the present time (either because time travel will never be discovered, or because no future Institute for Autoinfanticide has attended to our time). There is no reason to expect such a particular matter of fact to continue to obtain (absent independent evidence against the existence of time machines); it is only observed lawlike patterns that are projectible into the future.

For all we have learned from the probabilistic argument, time travel might yet occur in our world. But let us leave the actual world to the physicists and return to the question of whether there are conceptual or metaphysical challenges to the very *possibility* of time travel. As noted, the probabilistic argument provides no such challenge. But further interesting challenges to time travel may be based on Horwich's thought experiment.

4. COUNTERFACTUALS OF COINCIDENCE

Imagine that time travel is indeed possible, and that The Corporate Board contemplates forming an Institute for Autoinfanticide. In fact they decide against its formation. But what would have happened had the Institute been formed? There would have been an incredible series of coincidences. Had the institute been formed, there would have been a long string of slips on banana peels, serendipitous changes of heart, and the like.

We are all familiar with *might*-conditionals of this sort: if I had gotten up from the couch today, I might have tripped on a banana peel. But few of us think that in normal cases, *would*-conditionals of this sort are ever true. Unless one's couch is surrounded by banana peels, garden rakes and hanging cymbals, counterfactuals of this sort seem false:

If I had gotten up from the couch, I *would* have encountered some "coincidental" disaster

At most what is true is:

If I had gotten up from the couch, I *might* have encountered some "coincidental" disaster

One does have the feeling that "something funny is going on" after hearing time travel defended in the face of Horwich's thought experiment. The following argument against the possibility of time travel makes precise one worry in the vicinity. If time travel were possible, then counterfactuals like the following would be true:

If many, many time travelers went back in time intending to kill their earlier selves, equipped with deadly weapons, hardened hearts and excellent information about their targets, there would be a long string of coincidences: slips on banana peels, sudden attacks of remorse, mistaken identities and so on.

But "would-counterfactuals of coincidence" like this are never true. At best, "might-counterfactuals of coincidence" are true. Coincidences are not things that *would* happen; they are things that *might* happen. Therefore, time travel is impossible.

5. COUNTERFACTUALS OF COINCIDENCE AND FREEDOM

Rather than basing an argument against time travel on the claim that would-counterfactuals of coincidence are never true, one might instead use those counterfactuals to undermine the alleged ability or freedom of the time traveler to kill her infant self.[4] Kadri Vihvelin (1996) in effect does just this. She argues that if S has the ability to do A, then it must be the case that if S were to try to do A, S would or at least might succeed. If S would fail – repeatedly! – were S to try to do A, then S does not in fact have the ability to do A. Given Vihvelin's principle, would-counterfactuals of coincidence would undermine the freedom of time travelers to kill their former selves, for if those time travelers attempted autoinfanticide, they would fail.

One might press the argument further, as an argument against the possibility of time travel, at least in cases where the time traveler has no strange shackles that restrict her activity. Absent any such strange shackles, a time-traveler with the means and inclination surely *could* kill her former self; what would be stopping her? Thus, such a time traveler both could and could not kill her former self. The only escape from this contradiction, so the argument runs, is to reject the possibility of time travel, or to argue that time travel essentially

requires strange shackles on the time traveler. Vihvelin herself does not draw this conclusion, but it is hard to see why. Once the inability of the time traveler to kill her former self is admitted, one wonders what prevents her from doing so. Vihvelin is willing to grant that a time traveler is free to do many things other than kill her earlier self, but what prevents the time traveler from doing *this* thing?

6. COINCIDENCES AND FREEDOM

A final argument would be that the repeated slips on banana peels are too predictable and regular to be coincidences. There must be some force or mechanism causing the slips, perhaps due in some way to the fact that the would-be assassins are time travelers. But if that's so, then again we have a challenge to the freedom of the time-travelers, for such a force would presumably be inconsistent with their freedom. Of course, this is no threat to the possibility of time travel itself, for one can always spin a time-travel yarn with a convenient guardian of logic ready to cause slips on banana peels when inconsistency threatens. But the argument nevertheless threatens those of us who think that time travel is possible without such shackles on time travelers. It moreover threatens the possibility of time travel in worlds like our own, in which time travelers would presumably be unshackled.

7. A CLOSER LOOK AT COUNTERFACTUALS OF COINCIDENCE

In fact, each of these arguments may be rebutted. I begin with the argument of Section 4, that would-counterfactuals of coincidence are never true. Let us leave time travel for the moment and consider a more mundane case. Suppose I tried to throw a heavy stone at a fragile window. Since I have good aim and a strong arm, the window would break. I *might*, I suppose, slip on a banana peel, or hit a bird passing by with the rock, or have my throw deflected by a great gust of wind, or have a sudden failure of aim despite my many years of training in stone-throwing. But at the very least, it surely is not the case that one of these strange coincidences *would* happen. The would-counterfactual of coincidence:

If I were to try to throw the stone at the window, I would slip on a banana peel or hit a passing bird or . . .

is false.

But now let us consider a different counterfactual:

(C) If I were to try to throw the stone at the window but the window did not subsequently break, then I would slip on a banana peel or hit a passing bird or . . .

Here I have built my failure into the antecedent; the counterfactual concerns what would have happened had I tried and failed. Here, I think, our sense is that the counterfactual is now true. Given the background facts, the only way for me to fail to hit the window would be for some strange coincidence to occur. Though most ordinary would-counterfactuals of coincidence are false, some are true, namely those whose antecedents describe circumstances that could only come about by an "unlikely coincidence". We can think of the antecedents of these conditionals as describing states of affairs that embed a certain "tension", states of affairs that are "difficult" to make true. To include such a state of affairs, a possible world must include some strange coincidence. (Or something even stranger, perhaps a lurking guardian of the window ready to spring out and intercept the rock. Since no such guardian is present in the actual world, surely no guardian would have existed had I thrown the rock.)

We should all agree that there are true would-counterfactuals of coincidence like (C) whose antecedents "embed tension" in this way. But in fact, the counterfactuals in the time travel case are similar. We ask what would have happened if *a time traveler had tried to kill her earlier self*. If the time traveler is in fact a capable assassin and has the appropriate resolve and information, then making this antecedent true is very "difficult"; it is hard to find a non-coincidental reason why the time traveler would fail. And yet there is no possible world in which the time traveler successfully kills her earlier self. Thus, the would-counterfactual of coincidence:

If a certain time traveler had tried to kill her earlier self, she would have slipped on a banana peel or had a sudden change of heart or . . .

looks a lot like (C). But we admitted that (C) is true. We should say the same thing about this counterfactual. Coincidences *would* have happened, in the right circumstances.

The same can be said for counterfactuals concerning repeated attempts at auto-infanticide:

If many, many time travelers went back in time intending to kill their earlier selves, equipped with deadly weapons, hardened hearts and excellent information about their targets, there would be a long string of coincidences: slips on banana peels, sudden attacks of remorse, mistaken identities, and so on.

These counterfactuals are true as well; their truth is no more remarkable than that of (C). Here the coincidences described in the consequent are even more "unlikely" than those in the case of an individual assassin. But this is what one should expect since the antecedent here describes a state of affairs that is more difficult to make true than the antecedent in the single-assassin case. This same phenomenon may be observed in uncontroversial cases. Consider "loading" the antecedent of (C) thus:

If I were to try to throw the stone at the window and the window did not break as a result, and there existed no banana peels in the entire world, and there existed no birds, and I were the deadliest rock-thrower in the world, then . . .

To get a true counterfactual, the consequent will need to become even wilder:

. . . a random quantum-event would have caused the rock to explode or I would have been struck by lightning or . . .

The antecedents of these counterfactuals concerning time travel in a sense "have their difficulty built-in". The first begins "If a certain time traveler had tried to kill her earlier self"; one can tell just from looking at this sentence that a world in which it is true will contain an odd, "coincidental" event. But we might instead consider counterfactuals like this:

(T) If Tina had tried to kill the little girl standing in front of her at *t*, she would have slipped on a banana peel or had a sudden change of heart or . . .

where Tina is the time traveler in question, and where the little girl standing in front of her at *t* is in fact her earlier self. (T) is not quite so parallel to (C), since the antecedent of (C) also "has its difficulty built-in" – it says explicitly that my attempt to break the window fails. Nevertheless, the truth of (T) may still be explained in the same way. Let us assume the Lewis-Stalnaker theory of counterfactual conditionals, according to which, roughly, ⌜If it had been

the case that P then it would have been the case that Q] is true at *w* iff Q is true in the possible world most similar to *w* in which P is true.[5] The relevant similarity relation is one determined in part by the conversational context of the speaker, and may weight some contextually salient factors more heavily than others. In the context in which (T) seems true, we are holding constant the fact that Tina is a time traveler, and the fact that the little girl standing in front of her at *t* is Tina herself. So in this context, the counterfactual is a lot like (C), in containing an antecedent that is very "difficult" to make true in worlds similar to the actual world, given that in all those worlds Tina is a time traveler attempting to kill her earlier self. (T)'s truth is therefore no more surprising than (C)'s.

8. SELECTIVE ATTENTION

The truth of certain would-counterfactuals of coincidence in cases of time travel has been defended. But what of their connection with freedom? How can the members of the Institute for Autoinfanticide be free of unusual constraints during time travel, when if they were to attempt to kill their former selves they would repeatedly fail?

Let us again examine an analogous case having nothing to do with time travel.[6] Suppose we define a *permanent bachelor* as a man who never gets married. When we survey the class of permanent bachelors across the space of all possible worlds, we find that they fail to get married for a variety of reasons. Some never have the inclination, others wish to be married but never find a suitable partner, others slip on banana peels and fatally injure themselves while walking down the aisle, and so on. No anti-nuptial force need be postulated to account for this: by our definition of "permanent bachelor" *we* selectively attend to a certain class of possible individuals when we ask for the class of permanent bachelors. Many of these permanent bachelors *could have* gotten married. No force stands in their way; had they gotten married, they would no longer have counted as permanent bachelors.

The example may be brought a step closer to relevance by considering certain counterfactuals.

For all x, if it had been the case that (x is a permanent bachelor who attempts to get married), it would have been the case that: (x slips on a banana peel and dies, or gets a bad case of cold feet, or . . .)

or even:

For any series of persons, x_1, x_2, . . ., if it had been the case that (x_1, x_2, . . . are permanent bachelors that attempt to get married), then it would have been the case that (each slips on a banana peel and dies, or gets a bad case of cold feet, or . . .)

Properly filled in, these (universally quantified) would-counterfactuals of coincidence may well be true: given the "tension" embedded in their antecedents, those antecedents could be true only if an odd coincidence or series of coincidences occurred.

But now let us ask whether these would-counterfactuals of coincidence undermine the freedom of the permanent bachelors. Clearly, they do not. Their truth simply issues from selective attention on our part. Given our definition of "permanent bachelor", we do not count as a permanent bachelor anyone who succeeds in marrying; we therefore ignore all possible individuals that marry. The class of individuals that remains under scrutiny then contains a disproportionate number of individuals to whom "coincidental" things occur. But these individuals need not be subject to extraordinary constraints. The "coincidences" that prevent their marriages might be just that – coincidences. And we do not count coincidences of this sort as undermining a person's freedom. If a person accidentally slips and falls on his way to the altar, we do not regard him as incapable of marriage.

Of course, after repeated mishaps, a permanent bachelor might begin to doubt that the mishaps are coincidental. (The thought that the obstacles he repeatedly encounters have some common cause, a cause that undermines his freedom to marry, might be particularly compelling after, say, the fifteenth lightning bolt eliminates yet another nervous bride.) Such doubts might even be reasonable. But they would be mistaken. Logical space does, after all, contain persons to whom repeated coincidences occur. By defining "permanent bachelor" as we did, we single out these unfortunates for attention. Of course, some permanent bachelors *are* in the grips of social, psychological, or supernatural pressures rendering them incapable of marriage. Others lack the desire. But those in a third group fail to get married through mishap or coincidence.

In thinking about the freedom of permanent bachelors, we ought to distinguish between two sorts of claim. First, consider a permanent bachelor who never attempts to get married. Let this person be a perfectly ordinary, actual, permanent bachelor with no extraordinary social or psychological impediments to marriage. We clearly want to say that this person could have married, despite the truth of counterfactuals like: "if this person had been a permanent bachelor, and had tried to get married, he would have slipped on a banana peel or . . .". The *actual* freedom to get married is claimed, and his *counterfactual* failure (under the description "permanent bachelor") is irrelevant to this claim. Contrast this with a claim of *counterfactual freedom*. One might claim *also* that, had this person been a permanent bachelor and tried to get married, he would have failed in one of a number of "coincidental ways", but would nevertheless have been free. Here one is claiming the permanent bachelor has counterfactual freedom, despite counterfactual failure. This claim of freedom would also be true, I think. As argued above, cases in which permanent bachelors try and fail to get married include cases of "coincidental failure" similar to cases with which we are familiar. Some people really do slip on banana peels on their way to the altar. Such slips involve bad luck, but no failure of freedom. The appearance to the contrary is due to neglecting the role of selective attention in the truth of would-counterfactuals of coincidence.

9. SELECTIVE ATTENTION, COINCIDENCES AND FREEDOM

Let us now return to time travel, beginning with the argument of Section 6. The argument was that the numerous mishaps faced by time travelers attempting autoinfanticide just couldn't be coincidences. There must be some force or mechanism preventing the killings, perhaps a force causally linked to the would-be assassins' status as time travelers. But then, those assassins are shackled in an implausible way, and would not be free.

In light of our remarks on selective attention, the argument loses its force. The many mishaps facing the class of permanent bachelors require no explanation beyond the fact that *we* delineated the class with our notion of a permanent bachelor. The class of possible

worlds containing time travelers who repeatedly attempt autoinfanticide is similar. We have placed two constraints on this class of worlds that are very difficult to satisfy jointly. The first requires the worlds to contain large numbers of persons who want to kill certain persons, and have the means and the desire to do so. The second constraint is that the would-be killers are time travelers and the would-be victims are their former selves. This in effect requires that these persons fail in their missions. We thereby selectively attend to a class of worlds that contains large numbers of "coincidental failures". This requires no force compelling failure. *We* have delineated the class of worlds so that it contains the failures; the failures can still be genuine coincidences. The freedom of the time travelers is not compromised.[7]

It may help to remember that logical space contains many worlds with segments that are qualitatively like cases in which time travelers confront their former selves. In some, the would-be murderers fail, but in others they succeed. These latter cases are cases in which the victims are not the former selves of the murderers. Whether a given segment is embedded in a world in which a would-be murderer counts as the later self of a victim determines the inclusion of the segment in the class of worlds to which we have selectively confined our attention, in discussion of time travel and autoinfanticide.

10. SELECTIVE ATTENTION AND VIHVELIN'S ARGUMENT

The other argument against the possibility of time travel (Section 5) was that the freedom of time travelers is undermined by the truth of would-counterfactuals of coincidence. The argument appealed to Vihvelin's principle that a person is genuinely free to do a certain thing only if it is true that if she had tried to do the thing, she would have or at least might have succeeded. This argument, too, is undermined by the phenomenon of selective attention, although the matter is delicate. The truth of counterfactuals like these:

(PB1) For all x, if it had been the case that (x is a permanent bachelor who attempts to get married), it would have been the case that: (x slips on a banana peel and dies, or gets a bad case of cold feet, or . . .)

(PB2) For any series of persons, $x_1, x_2, \ldots$, if it had been the case that ($x_1, x_2, \ldots$ are permanent bachelors that attempt to get married), then it would have been the case that (each slips on a banana peel and dies, or gets a bad case of cold feet, or . . .)

was argued to undermine neither the freedom of actual permanent bachelors, nor the counterfactual freedom of permanent bachelors in worlds in which they attempt and fail to get married. Likewise, I say that the truth of counterfactuals like:

If many, many time travelers went back in time intending to kill their earlier selves, equipped with deadly weapons, hardened hearts and excellent information about their targets, there would be a long string of coincidences: slips on banana peels, sudden attacks of remorse, mistaken identities and so on.

undermines neither the freedom of time travelers who do not attempt to kill their earlier selves, nor the freedom of those that do but fail for "coincidental" reasons. The slips on banana peels and other mishaps are genuine coincidences; the fact that they *would* occur, were attempts at autoinfanticide made, is simply due to our selective attention in the antecedent of the counterfactual to a certain class of worlds in which failure has been definitionally guaranteed.

There is, however, a disanalogy between the permanent bachelors and the time travelers. Counterfactuals like (PB1) and (PB2) do not directly undermine Vihvelin's principle. Vihvelin's principle says that a necessary condition for an agent's being free to do something is that the agent might have succeeded had she tried to do that thing. The necessary condition here is *not* that the agent might have succeeded *under a certain description*; it is simply that the agent might have succeeded. But all (PB1) and (PB2) show is that a permanent bachelor can be free to marry despite the fact that, *under the description "permanent bachelor"*, he would have failed had he tried to marry. The predicate "permanent bachelor" was (deliberately) placed inside the scope of the counterfactual conditional in (PB1) and (PB2). If those predicates are placed outside the scope of the counterfactual then the claims become false:

(PB1′) For any permanent bachelor, x, if it had been the case that (x attempts to get married), it would have been the case that: (x slips on a banana peel and dies, or gets a bad case of cold feet, or . . .)

(PB2′) For any series of permanent bachelors, $x_1, x_2, \ldots$, if it had been the case that ($x_1, x_2, \ldots$ attempt to get married), then it would have been the case

> that (each slips on a banana peel and dies, or gets a bad case of cold feet, or . . .)

(PB1′) and (PB2′) are needed to refute Vihvelin's principle, for it is (PB1′) and (PB2′) that concern the counterfactual success of the agents *simpliciter*; (PB1) and (PB2) concern the counterfactual success of the agents under the description "permanent bachelor".

Constraints on the similarity metric for the counterfactual conditional can accomplish what is accomplished by the explicit inclusion of a predicate inside the antecedent of such a conditional. Imagine a similarity metric that holds constant a person's status with respect to permanent bachelorhood, and therefore counts possible worlds in which an actual permanent bachelor gets married as being very distant from the actual world. In that case (PB1′) is true; nevertheless, actual permanent bachelors are still free to marry. The truth of (PB1′) is due to the same sort of selective attention as results in the truth of (PB1), only now the selective attention is accomplished by the similarity metric rather than by the presence of the predicate "permanent bachelor" inside the scope of the counterfactual. The moral is that Vihvelin's principle fails if the similarity metric for the counterfactual conditional accomplishes this sort of selective attention. Vihvelin's principle, for all we have said, may remain true when counterfactuals are *not* interpreted via these sorts of similarity metrics.

It is a hard question exactly which interpretations of counterfactuals vindicate Vihvelin's principle. Though I will not take up this question in any depth, I note that Vihvelin's principle seems to fail under similarity metrics that hold constant facts about agents that occur in the *future* of the actions in question. Someone's status as a permanent bachelor, for example, concerns in part what occurs in the future of an attempt to marry, and correspondingly cannot be held constant in evaluating counterfactuals if Vihvelin's principle is to be true.[8]

Vihvelin's use of her principle to undermine the freedom of time travelers can now be questioned. This is especially clear if we presuppose the metaphysics of temporal parts. Given this metaphysic, the confrontation of the earlier self by the later self involves a meeting between a temporal part we may name "Killer" and a temporal part we may name "Victim". In fact, Killer does not kill

Victim; the question is whether Killer was free to kill Victim. To answer this question, Vihvelin's principle instructs us to consider a certain counterfactual. To correctly apply Vihvelin's principle we must *not* consider:

> If Killer had been the temporal part of the same time-traveling continuant as Victim, and Killer had attempted to kill Victim, then Killer would have or at least might have succeeded

(for that concerns Killer's counterfactual doings under a description), but rather simply:

> (KV) If Killer had attempted to kill Victim, Killer would have or at least might have succeeded

Vihvelin's principle says that Killer is free to kill Victim only if (KV) is true. Is (KV) true? (KV) may well be false under a similarity metric that gives weight to what occurs after the confrontation of Victim by Killer. Such a metric might require Killer and Victim to be temporal parts of the same time-traveling continuant in nearby worlds, and in such worlds Killer's murderous attempt fails. But under such interpretations of counterfactuals, Vihvelin's principle is false. On the only interpretation of counterfactuals that vindicates Vihvelin's principle, what occurs after the confrontation does not affect similarity of worlds, and therefore whether worlds match the actual world with respect to whether Killer and Victim are temporal parts of the same continuant is irrelevant. Under this similarity metric, presumably Killer succeeds in killing Victim in the worlds most similar to actuality; in such worlds, Killer and Victim are no longer temporal parts of the same person. (KV) is therefore true, under this similarity metric. Vihvelin's principle is no obstacle to the temporal part, Killer, being free to kill the temporal part, Victim.[9]

On most temporal parts theories, the time traveler is a "spacetime worm" that contains Killer and Victim as proper temporal parts. Thus, showing that Killer is free to kill Victim is not the same as showing that the time traveler is free to kill her earlier self.[10] Nevertheless, it seems to me to show that the case contains as much freedom as we should want. Let us take a step back and recall the point of inquiring into the freedom of time travelers. I am trying to show that a time traveler would not be subject to constraints that a normal person in similar circumstances would not be subject to

(nor would she have her ordinary powers undermined). The point of doing so is to avoid the conclusion that the possibility of time travel could only be secured by postulating a kind of "force" or "guardian of logic" that shackles time travelers by ensuring that they perform certain tasks. Such guardians should be resisted, not because they are impossible (since surely they are *not* impossible), but rather because it is hard to believe that the sort of time travel the physicists consider an open possibility would require such exotic metaphysical add-ons, and moreover because time travel seems possible in worlds that lack such guardians. If temporal part Killer is free to kill temporal part Victim, that is sufficient for time travel being possible without metaphysical add-ons; it shows that the causal order in a world including time travel need not contain any such guardians. This is the bar for success I set for myself: showing that time travelers need not be shackled by metaphysical add-ons. It would not matter much if it turned out that what we ordinarily mean by "the time traveler can kill her earlier self" is false. (In fact I argue below that this does not turn out false.)

It is worth showing that Vihvelin's challenge to the freedom of time travelers can be answered without assuming the metaphysics of temporal parts. Pretend for a moment that objects "endure", i.e., persist "wholly present" through time. As we have learned, utilizing Vihvelin's principle requires consulting a counterfactual that does *not* contain the description "time traveler" in its scope. So let us name the time traveler "Katy", and consider the truth of the following counterfactual relative to a world, *w*, in which Katy travels back in time with the intention of killing her former self, but decides against doing so:

(K) If Katy had tried to kill Katy, she would have or at least might have succeeded

There may be similarity metrics under which (K) turns out false, but the question is whether (K) is true under a metric that does not hold constant facts about what happens after the attempted killing, for that is the only metric under which Vihvelin's principle is true. To answer this question we must compare the similarity of various worlds to the original world *w*. Those worlds include:

Class 1: Worlds in which there are "two Katys", the "larger" of which is the time traveling future self of the "smaller" one, and who tries to kill the smaller one but fails in one of the familiar "coincidental" ways.

Class 2: Worlds in which there is only one Katy, who simply commits suicide.

Class 3: Worlds in which there are two Katys, the larger of which is the time traveling future self of the smaller one, in which the larger one tries to kill the smaller one, but in which the smaller one in fact gets the drop on the larger one and kills her.[11]

Class 4: Worlds in which there are two Katys, there occurs no time travel, and the larger one kills the smaller one.

The locution "two Katys" must be explained. Recall that we are pretending away temporal parts for the sake of argument. Thus, the original world *w*, in which Katy the time traveler confronts her earlier self, does *not* involve two temporal parts of the same space-time worm; rather, we have a single thing, Katy, that wholly occupies two distinct person-shaped regions of space at a single moment. "There are Two Katys" means that Katy is spatially bi-located in this way; talk about what "the larger Katy" does is talk about what occurs in the larger region of space, and so on.[12]

Our question is which class of worlds is most similar to *w*. Class 2 we may set aside as being too distant from *w* since *w* contains two Katys. Class 3 is presumably also quite distant from *w* since in *w* the smaller Katy is only an infant, and is no match for the ruthless, powerful larger Katy. So the question is whether Class 1 or Class 4 contains the worlds closest to *w*. (K) turns out false if Class 1 contains the closer worlds, true if Class 4 contains the closer worlds. The correct answer, I believe, is that Class 4 contains worlds closer to *w*; therefore (K) turns out true. In the worlds in Class 4, Katy is spatially bi-located, just as she is in *w*. This spatial bi-location is *not* caused by Katy's future entrance into the time machine (as it is in *w*) since Katy never grows up to enter a time machine in these worlds; the smaller Katy is rather killed in her infancy by the larger Katy. These are admittedly bizarre worlds, since they contain persons that, for no causal reason, come to occupy two distinct person-shaped regions of space. Moreover, they differ from *w* since in *w* Katy's bi-location is caused. But this difference is exclusively about events that occur *after* the encounter between the two Katys, and so is irrelevant to the question of similarity between these worlds and *w*.

Moreover, the worlds in Class 4 lack the strange coincidences that occur in the worlds in Class 1. Therefore they seem closer to *w* than are the worlds in Class 1.

I have argued that since (KV) and (K) come out true under the only similarity metric that vindicates Vihvelin's principle, that principle presents no obstacle to the freedom of time travelers. But there is a more fundamental reason this is true, a reason that is independent of the question of the truth of these counterfactuals. Suppose for the sake of argument that, contrary to what I have argued, (KV) and (K) turn out false (under the relevant similarity metric). The freedom of time travelers would still not be undermined, for Vihvelin's principle may then be argued to be false.

The reason Vihvelin's principle has intuitive appeal is that counterfactuals in typical cases have a certain causal significance. Typically, if it is true that I would have failed repeatedly had I tried to do a certain thing, then this is a sign of either i) the presence of some causal force or mechanism that impedes my doing the thing in question, or ii) the absence of some causal mechanism that would otherwise enable me to do the thing in question (for example a good shooting eye or strong trigger finger). For in typical cases, if the possible world most similar to the actual world in which I try to do the thing is one in which I fail, that is because a disabling causal mechanism is present in that world, or because a typically enabling mechanism is absent in that world; but then because of the similarity between that world and the actual world, the presence or absence of this mechanism carries over to the actual world; the actual presence of absence of this mechanism then precludes my actual freedom to do the action. But now recall the moral of selective attention. As was argued at the beginning of this section, the repeated failures of time travelers to kill their former selves are merely coincidental. There is no one driving force that makes a time traveler fail; the failures need not be caused, for example, by the fact that they are time travelers. There are, of course, causes for the failures; "coincidental failures" are not uncaused failures. But the sorts of causes here – slips on banana peels and the like – are not the kinds of causes we take to undermine freedom; they are neither the presence of disabling mechanisms nor the absence of enabling mechanisms. So even if the closest possible worlds to a given world, *w*, in which a time traveler

attempts to kill her former self are worlds containing such coincidental failures, this does not imply the presence in *w* of disabling mechanisms or the absence in *w* of enabling mechanisms. Therefore, the falsity of counterfactuals like (K) and (KV) in *w* would not show that time travelers in *w* lack the ability to kill their former selves. Vihvelin's principle fails because in these cases the falsity of counterfactuals like (K) and (KV) lacks its usual import.

Remarks like those of the past few paragraphs also answer a challenge to the freedom of time travelers based on a principle weaker than Vihvelin's: that *freedom to do X requires the metaphysical possibility of doing X*, whereas it is metaphysically impossible for Katy to kill Katy in circumstances intrinsically like the case of time travel.[13] First, this principle allows that the temporal part Killer is free to kill the temporal part Victim, and that is freedom enough. Second, temporal parts aside, the worlds in Class 4 show that it *is* metaphysically possible that Katy kill Katy in circumstances intrinsically like those of the case of time travel. This is not to say that there is a possible world in which Katy kills Katy in this way but is also a time traveler in that world. But the case of the permanent bachelors shows that freedom at most requires that the agent possibly do the action *simpliciter*, not that the agent possibly do the action under some description. Finally, if the existence of these possibilities is rejected, this would only show that their existence is not required for freedom, for the nonexistence of the possibilities would not have its usual import; it would not indicate the presence of disabling causal mechanisms or the absence of enabling mechanisms in worlds in which time travel occurs.

11. CONCLUSION

Time travel escapes again, unscathed. If many time travelers attempted autoinfanticide, an apparently miraculous series of coincidences would ensue. But this fact is unremarkable, and in no way undermines time travelers' freedom. As comparison with uncontroversial cases has shown, it is the result of the description "time traveler who attempts autoinfanticide" focusing our attention on a certain class of possible worlds, a class of worlds that is guaran-

teed, by definition, to include large numbers of what we consider to be highly improbable coincidences.[14]

NOTES

[1] Lewis uses the contextual relativity of "can" to explain why the argument that a time traveler *cannot* kill her earlier self is so seductive.

[2] See Horwich (1975, 1987, Chapter 7).

[3] Thanks to Ioan Muntean for this point.

[4] I thank Europa Malynicz for drawing my attention to this argument.

[5] See Lewis (1973), Stalnaker (1968).

[6] I put this case to different use in Sider (2000).

[7] One caveat. I am only claiming that the truth of counterfactuals of coincidence does not *on its own* imply the existence of freedom-undermining facts about the world. In some cases, even cases without supernatural shackles, such facts may exist. Suppose I go back in time and contemplate shooting my former self, and that I travel on a closed timelike curve. Suppose further that the laws locally permit my firing the gun, in the sense that they are consistent with an intrinsic description D of my firing the gun. But finally suppose that the laws do *not* allow an event of type D to be embedded in any spacetime that contains a closed timelike curve like that in the actual world; they merely allow such an event to be embedded in a different spacetime structure. It seems to me an open question whether I can, in this case, kill my earlier self. One might argue that I cannot on the grounds that I cannot do anything inconsistent with the laws plus certain propositions P concerning the world's spacetime structure. Given the compatibilism I accept, this argument cannot be made for just any P, since given general relativity all actions affect the world's spacetime structure; but perhaps if P is chosen appropriately the argument goes through (P might say that the world contains a closed timelike curve of a certain sort). Even if this argument is correct, that would not undermine the possibility of time travel in these cases; it would just undermine the claim that I must be free since I am intrinsically like a non-time-traveler who *is* free. Freedom cannot be read off of local facts if global space-time structure is a potentially freedom-undermining fact.

[8] Compare Lewis (1976) on "can". For Lewis, to say that someone can do something is to say that her doing that thing is consistent with the relevant facts. What facts are relevant varies with context. In normal contexts, contexts in which fatalism is false, facts about what occurs after the time in question are irrelevant.

[9] Vihvelin also claims that worlds in which Killer is not part of the same continuant as Victim are worlds in which Killer and Victim miraculously have the same DNA, and are therefore more distant than worlds in which Killer fails in some coincidental way to kill Victim (1996, p. 326). I disagree: the DNA-match between Killer and Victim *from the point of view of what occurs before their encounter* is no more miraculous in the worlds in which Killer succeeds as it is in the worlds in which Killer fails. From the point of view of what occurs *after* the

encounter there is indeed a difference: the match remains unexplained in the world in which Killer succeeds whereas there is a causal link between Killer and Victim via the future time machine. But what happens *after* the encounter is irrelevant to the interpretation of the counterfactual, at least under a similarity metric relevant to questions of ability and freedom.

[10] I myself favor a version of the theory of temporal parts according to which continuants are instantaneous temporal stages, and according to which typical statements of the persistence of continuants are to be analyzed using temporal counterpart theory (Sider, 2001, Chapter 5, Section 8). Given this theory, since Killer is free to kill Victim it follows that the time traveler is free to kill her earlier self, for the time traveler *is* Killer and the earlier self *is* Victim.

[11] Thanks to Hud Hudson for pointing out this amusing possibility.

[12] In Sider (2001, Chapter 4, Section 7) I present objections to this conception of enduring time travelers, but this is the way defenders of endurance will need to understand the case.

[13] I thank Chris Heathwood and Hud Hudson for helpful comments here. Note that unlike me, Vihvelin does *not* stipulatively define "killing" as requiring permanent death; she allows the possibility of a time traveler killing her former self, but argues that worlds in which this occurs are typically too distant to be relevant to the counterfactuals in question.

[14] I would like to thank David Braun, Earl Conee, Tamar Szabó Gendler, Chris Heathwood, Hud Hudson, Europa Malynicz, Ioan Muntean, Daniel Nolan, Peter van Inwagen, Kadri Vihvelin, Peter Vranas, Brian Weatherson, and an anonymous referee for helpful comments.

REFERENCES

Earman, John (1995): 'Recent Work on Time Travel', in S. Savitt (ed.), *Time's Arrows Today: Recent Physical and Philosophical Work on the Direction of Time* (pp. 268–324), Cambridge: Cambridge University Press.

Horwich, Paul (1975): 'On Some Alleged Paradoxes of Time Travel', *Journal of Philosophy* 72, 432–444.

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

Horwich, Paul (1995): 'Closed Causal Chains', in S. Savitt (ed.), *Time's Arrows Today* (pp. 259–267), Cambridge: Cambridge University Press.

Lewis, David (1973): *Counterfactuals*, Cambridge, MA: Harvard University Press.

Lewis, David (1976): 'The Paradoxes of Time Travel', *American Philosophical Quarterly* 13, 145–152.

Sider, Theodore (2000): 'Simply Possible', *Philosophy and Phenomenological Research* 60, 585–590.

Sider, Theodore (2001): *Four-Dimensionalism: An Ontology of Persistence and Time*, Oxford: Clarendon Press.

Smith, Nicholas J.J. (1997): 'Bananas enough for Time Travel', *British Journal for the Philosophy of Science* 48, 363–389.
Stalnaker, Robert (1968): 'A Theory of Conditionals', in N. Rescher (ed.), *Studies in Logical Theory*, Oxford: Blackwell.
Vihvelin, Kadri (1996): 'What Time Travelers Cannot Do', *Philosophical Studies* 81, 315–330.

Rutgers University, NJ
USA
E-mail: trsider@syr.edu