*philosophies*

MDPI

*Article*

# Autoinfanticide Is No Biggie: The Reinstatement Reply to Vihvelin

**Richard Mark Hanley**

Department of Philosophy, University of Delaware, Newark, DE 19716, USA; hanley@udel.edu

**Abstract:** David Lewis's attempt to defuse grandfather paradoxes consistently without special restrictions on the ability of time travelers to act in the past is controversial. Kadri Vihvelin uses the case of possible autoinfanticide—killing one's infant self—to argue on Lewisian grounds that Lewis is wrong, since all counterfactual attempts at autoinfanticide would fail. I present a new defense of Lewis against Vihvelin premised on the possibility of personal *reinstatement*, where a person who dies prematurely is replicated from information collected from a previous live scan. I argue on Lewisian grounds that in a Vihvelin case where Suzy does not attempt to kill Baby Suzy, Vihvelin has not shown that Suzy would have failed had she tried to kill Baby Suzy. For, Baby Suzy might have been reinstated. Hence, even granting Vihvelin's own assumptions, a Lewisian can assert that Suzy can kill Baby Suzy. Reinstatement does not require a "big" miracle; so autoinfanticide is no biggie.

**Keywords:** time travel; reverse causation; fatalism; ability; autoinfanticide; counterfactual dependence; possible worlds; teleportation; personal identity; personal fission; Newcomb Problem



**Citation:** Hanley, R.M. Autoinfanticide Is No Biggie: The Reinstatement Reply to Vihvelin. *Philosophies* **2021**, *6*, 87. https://doi.org/10.3390/philosophies6040087

Academic Editor: Alasdair Richmond

Received: 15 September 2021
Accepted: 10 October 2021
Published: 18 October 2021



## 1. Introduction

David Lewis [1] argues that even in progenitor or *retro-killing* cases—the most notorious being "grandfather paradox" cases—time travelers have more or less the same abilities as anyone else. In a series of pieces of which [2–4] are representative, Kadri Vihvelin argues that although time traveling to the past and retro-killing is logically possible, in a typical progenitor case the time traveler lacks the ordinary ability to do the deed. In the ordinary sense of "can", a time traveler cannot retro-kill. Ryan Wasserman [5] presents a vigorous recent defense of Vihvelin's view against a range of objections. I take no issue with that defense here, and instead present a new argument to a limited conclusion: that Lewis need not change his own position in response to Vihvelin's arguments. Vihvelin's strategy is to argue against Lewis assuming many of his own views: his account of single timeline time travel, his account of counterfactual dependence, his temporal parts or "worm" theory of persistence, and—though not mentioned explicitly—his account of truth in fiction. But I shall show that Vihvelin has not brought the full suite of Lewisian views to bear on the issue. Considered in the broader light of Lewis's view that teleportation is survivable, that transworld identity is a matter of modal counterpart theory, and that *de re* modality is inconstant, Vihvelin's case is uncompelling, even granting the *Counterfactual Possibility Principle* she proposes to analyze the "can" of ability.

Lewis describes the case of trained assassin Tim who time travels to the past in a single timeline, and who wants his paternal grandfather dead. Can Tim kill Grandfather in 1921, before Tim's father is conceived? Lewis answers Yes, and No. Tim can, in the *ability* sense of "can", kill Grandfather. But Tim will suffer a temporary lack of luck and fail to kill Grandfather, in spite of his ability, because he did after all fail. So in the *luck* sense of "can", Tim cannot kill Grandfather [1]. The crucial assertion Lewis makes is that no systematic explanation of Tim's failure is required—no "boring" temporal censor of the sort others are tempted to invoke [1] (p. 149). Some ordinary occurrence—the world failing to cooperate fully with one's plans—is sufficient. It will be handy to refer to such an occurrence as a

*banana peel*. Time traveler or not, even the best trained assassin is liable to be foiled by a gun jam or a wind gust or a stray bird or a literal banana peel.

Consider that there are ways for Tim to succeed that are compossible with the killing taking place in 1921, before Tim's father is conceived. Suppose Tim tries to kill Grandfather and succeeds. Unexpected! But it turns out that Grandfather had an arrangement with a sperm bank, and Grandmother was artificially inseminated after 1921. Does ruling out such devices help? Not entirely. Suppose conception happened the old-fashioned way, and by just the man Tim thinks is responsible. Tim nevertheless succeeds in killing him, but it turns out that Grandfather is a time traveler, too. Tim kills him in 1921, before the conception in external time, but not before the conception in Grandfather's personal time. Grandfather had been to the future, visited Grandmother, and . . . you get the picture.

The point is that in Lewis's treatment of the grandfather paradox, when he says that Tim cannot kill Grandfather, the progenitor aspect is not after all central to the problem. The real issue is whether or not a single timeline time traveler can change the past—in this case, by causing something to happen on the timeline that never happened on the timeline. That is a contradiction [2]. To eliminate sperm banks and other devices, we need to state the fatalist-sounding view more precisely: Tim cannot kill Grandfather in 1921, *given* that Tim does not kill Grandfather in 1921. But the schema *X cannot do Y at Z given that X does not do Y at Z* is perfectly general, and we should not—on pain of global fatalism—thereby conclude that no one can ever do anything other than what they actually do.

It is no surprise that some philosophers think Lewis has defused the grandfather paradox, while others think he has merely dodged it. Whereas Lewis takes his opponent to be a kind of global fatalist, Vihvelin thinks that Lewis is wrong only about cases sufficiently like retro-killing. Just how far the cases extend beyond retro-killing is an interesting question. Individual human existence seems modally lucky: just about any small change to a range of events preceding your conception would have resulted in your non-existence. Call those events that had to happen just-so for you to exist *fragile*. Where the conception of any of time traveling Tim's progenitors is concerned, not only did Tim not mess with any fragile event, he *could not* have. Thus, Vihvelin's view might have far-reaching consequences for the abilities of time travelers; the fatalism is not global, but quite extensive, going far beyond retro-killings. The stakes are high.

Like Lewis, Vihvelin does not think you will succeed in retro-killing anyone that you did not actually retro-kill. Her distinctive claim is that you could not have succeeded; more precisely, that in a case where you did not try to retro-kill a progenitor on a particular occasion, it is true that had you tried, you would have failed. This raises the further question: what would have stopped you? Her arguments lead Vihvelin to a position that departs significantly from Lewis's. On the one hand, Vihvelin asserts—à la Lewis—that any counterfactual attempt to retro-kill would be foiled by a banana peel; on the other hand, she asserts—*pace* Lewis—that thanks to the nomological impossibility of processes like resurrection, it is the laws of nature that prevent retro-killings.

The structure of this paper is as follows. Section 2 presents Vihvelin's basic argument, which focuses on the case of possible autoinfanticide, and rests upon her Counterfactual Possibility Principle analysis of the "can" of "wide" ability. Section 3 introduces two logical possibilities, personal relocation and personal reinstatement. I show that Lewis believes both are cases of personal survival, and that reinstatement allows for successful autoinfanticide. Section 4 argues that reinstatement is nomologically possible and therefore arguably counterfactually relevant to autoinfanticide cases. Section 5 shows how Vihvelin might reassert the counterfactual irrelevance of reinstatement, by using Lewis's own metric of overall similarity of worlds to argue against the considerations of Section 4. Section 6 rebuts that argument in turn by pointing out that time travelers have counterfactual opportunities to manipulate the past that non time travelers lack. Section 7 shows how to use Lewis's metric to understand his own judgments about counterfactuals in a time travel version of a Newcomb Problem, and applies this understanding to the case of autoinfanticide with reinstatement. I argue that the closest success world is closer by Lewis's metric than any of

Vihvelin's banana-peel failure worlds. Section 8 offers a diagnosis: that the evaluation of these cases is made more difficult by Lewis's own somewhat misleading description of the metric, since "small" miracles need not be small, "big" miracles are not a matter of absolute size, and whether or not a miracle is big or small is highly context dependent. Section 9 uses these considerations to argue that Vihvelin's own position mentioned above—that counterfactual failure to retro-kill would be both effected by a banana peel and forced by the actual laws—is untenable under Lewis's metric. I conclude that Lewisians should continue to say that time travelers can retro-kill in the same sense that non time travelers can kill.

## 2. Vihvelin's Argument

Can Tim kill Grandfather? Like Lewis, Vihvelin answers Yes and No. Vihvelin allows that it is logically possible to retro-kill. For instance, there are worlds where Tim kills Grandfather and Grandfather is then resurrected [2] (p. 317). So Tim can kill Grandfather. And whereas Lewis says unequivocally Yes, Tim is able to kill Grandfather, Vihvelin answers Yes and No. Tim has the *narrow* ability, in that nothing in Tim's intrinsic properties precludes his killing Grandfather; but lacks the *wide* ability, in that something about Tim's extrinsic properties does preclude his killing Grandfather [3] (pp. 318–319). That makes two senses of "can" for which Vihvelin's answer is Yes, but the dispute is not merely verbal. Vihvelin claims that the wide ability sense is the *ordinary* sense of "can", and moreover that the wide ability sense is not the same as the luck sense Lewis identifies, so that global fatalism does not follow. Non time travelers will often fail to do things in the luck sense, all the while being widely able to do those things, and time travelers will often fail to do things in the luck sense, all the while being widely able to do those things. But time travelers have the distinction of sometimes failing to do a thing because they lack the wide ability to do it, *because* they are time travelers [3].

To distinguish between the luck or logical possibility senses and the wide ability sense, Vihvelin employs a principle not found in Lewis's work. Vihvelin's most comprehensive statement of it is as follows [3] (p. 319):

> *Counterfactual Possibility Principle*: S has, at time *t*, the wide ability to A only if it's *not* true, at t, that if S tried (again) to A, S would fail.
>
> [emphasis original]

This is equivalent to saying that S can do A only if were S to try (again) to A, S might succeed. The point of "again" is to allow for temporary lack of luck. If Tim were indeed widely able to kill Grandfather, then if Tim at first failed to shoot Grandfather dead because Grandfather's cigarette case deflected the bullet meant for his heart, it would be true that were Tim to try again, he might succeed. But, Vihvelin thinks, Tim would fail no matter how many times he tried. So Lewis [1] (p. 150) is wrong to say that Tim's failure even once is due to temporary lack of luck rather than any lack of ability.

The Principle also distinguishes between wide ability and logical possibility. By Lewis's own analysis, a counterfactual (or more generally, a subjunctive conditional) of the form "If it were that $p$, then it would have been that $q$" is actually true if and only if a world where $p$ is true and $q$ is true is, on balance, closer to the actual world than any world where $p$ is true and $q$ is not true. Vihvelin applies this schema to the example of autofanticide, which cannot be dodged by introducing a sperm bank or making the progenitor a time traveler. She imagines adult Suzy, who is sent back in time to visit her infant self and fails five times to kill Baby Suzy. Would Suzy have succeeded in a sixth attempt? No. Like Tim, Suzy would always have failed, no matter how often she tried. Therefore, by the Counterfactual Possibility Principle, Suzy is unable to kill Baby Suzy. Why would she always fail? Vihvelin writes [3] (p. 322):

> The worlds where Suzy tries and succeeds are worlds which either have resurrection from the dead or some sort of system of ontological understudies.

Call any "ontological understudy" case a *replacement*, where the identity between Suzy and Baby Suzy is broken in the world where Suzy's attempt succeeds. Vihvelin then gives an argument by cases. A world in which Suzy succeeds and in which Baby Suzy is resurrected is logically possible but nomologically impossible, and such a world is too distant to underpin counterfactual success. A replacement world is either not a world in which *Suzy* kills *Baby Suzy*, and so is utterly irrelevant, or if it somehow does count as a success world it involves breaking the causal dependence of adult Suzy on Baby Suzy, and hence such a world is too distant to underpin counterfactual success.

I am going to simplify things a little. Focus on what I shall call *Good Suzy World*, where 30-year-old time traveler Suzy is very healthy and strong and quite reasonably makes no attempt at all on Baby Suzy's life. She visits Baby Suzy and is alone with her for 10 min (their parents are downstairs). Baby Suzy is asleep in an open crib, and having locked the door and window, Suzy leans in at time $t$ and gently, silently chucks Baby Suzy under the chin. No banana peels are present to get in Suzy's way. While Baby Suzy continues sleeping, Suzy unlocks things and leaves, and then time travels back to the future, never to return. If Vihvelin is right about Good Suzy World, Suzy not only does not but cannot kill Baby Suzy at $t$. Assume there is one closest possible world where Suzy tries to kill Baby Suzy by crushing her windpipe instead of chucking her under the chin; call this *Bad Suzy World*. According to Vihvelin, it is a world where Suzy's attempt fails.

## 3. A Lewisian Counter: Personal Reinstatement

A series of metaphysical objections to Vihvelin have defended Lewis by appeal to cases of replacement, some of them quite exotic [4]. For instance, Peter Vranas argues that Vivhelin's argument is refuted by attending to a relevant metaphysical possibility [7] (pp. 118–119):

> [C]onsider a world—to simplify, and without loss of generality, say it is the actual world—at which Baby Suzy has an identical twin, Twin Baby Suzy, and at which Suzy sets off a bomb in a room where Baby Suzy and Twin Baby Suzy are asleep, intending to kill them both, but the bomb happens to kill only Twin Baby Suzy. Consider also a world $w$ which is qualitatively identical to the actual world, but at which (i) the bomb happens to kill only Baby Suzy, and (ii) Suzy is a later stage of Twin Baby Suzy, not of Baby Suzy . . . . Then $w$ is a world at which Suzy tries to kill Baby Suzy and succeeds, and at which Suzy is a later stage of some baby-stage (namely Twin Baby Suzy) whose DNA matches the DNA of Baby Suzy not by some miracle or improbable coincidence, but rather because the two baby-stages are identical twins. Since $w$ is qualitatively identical to the actual world, $w$ is at least as close to the actual world as any world at which Suzy tries to kill Baby Suzy but fails.

This is a difficult case to understand. It seems to appeal to *haecceitism*; and, if so, Lewis would reject it [8] (pp. 220–235). Lewis allows that you can coherently contemplate the possibility of being someone else, exactly as they actually are, but that is not contemplating some distinct possible world [8] (pp. 231–232); and it is not a possibility where you are both you and them. But let us suppose that Lewis can be persuaded that $w$ is a possible world distinct from the actual world. It is still not clear that this possibility would refute Vihvelin. For, $w$ would have to be as close to the actual world as the actual world is, since the nearest world where Suzy tries to kill Baby Suzy but fails is by hypothesis the actual world. Perhaps a difference in who is who does not matter to how close a world is to the actual world when assessing subjunctive conditionals, and Vranas can truly say that Suzy *might* have succeeded. (He cannot say she *would* have). But, on balance, I think Lewisians should prefer a more convincing rebuttal.

Fortunately, there is for Lewis a much more promising metaphysical possibility. Vihvelin allows that there are resurrection worlds where Suzy succeeds: they are worlds at which Baby Suzy dies and is buried, but is later resurrected from the dead and grows up to be the adult Suzy who travels back through time and kills her baby self [2] (p. 321).

Note what would not count as success. Had Suzy tried to drown Baby Suzy by throwing her into a frozen lake, it might be a case where the heart stops through hypothermia, and yet the victim can be revived by (carefully) warming them up again. In such a case Baby Suzy has not actually died, so this is not a success world. Nor is a world where Suzy tried to kill Baby Suzy by placing her in suspended animation, with Baby Suzy subsequently being reanimated. Nor is any *Princess Bride* world, where people can be "only mostly" dead. Vihvelin is instead thinking of more drastic, even colder cases where the body's individual cells have died and have decomposed and are somehow brought back to life through a reversal of that decomposition. Call this *corpse-resurrection*. I grant Vihvelin's claim that corpse-resurrection worlds are nomologically impossible and too distant for Suzy's counterfactual success.

But imagine a different kind of world: a teleportation world or *T-world*, where humans employ scanning and replication technology, in the first instance to travel. Scanning at the departure point is instantaneously followed by deliberate, total bodily destruction—which surely does count as death—and then single replication at the destination. Call this *relocation*. Is relocation resurrection of the same person, or is it merely replacement by a doppelgänger? For Lewis, it is resurrection. He writes concerning the question of what matters in personal survival [9] (p. 17):

> I answer, along with many others: what matters in survival is mental continuity and connectedness . . . . My total present mental state should be but one momentary stage in a continuing succession of mental states. These successive states should be interconnected in two ways. First, by bonds of similarity. Change should be gradual rather than sudden . . . Second, by bonds of lawful causal dependence . . . . [E]ach succeeding mental state causally depends for its character on the states immediately before it.

Lewis believes in person stages which are proper temporal parts of persons. When the two kinds of bond are present between two stages, they are R-related. In relocation a T-world traveler is temporally gappy, but if the first stage of the replica is R-related to the last stage of the scanned subject, then the gap is no obstacle to survival, and the traveler is one person. Lewis explicitly endorses teleportation as survival [10] (pp. 192–193):

> Consider our opinions about teletransportation, an imaginary process that works as follows: the scanner here will take apart one's brain and body, while recording the exact state of all one's cells. It will then transmit this information by radio. Traveling at the speed of light, the message will reach the replicator. This will then build, out of new matter, a brain and body exactly like the one that was scanned. Some philosophical positions on personal identity imply that one survives teletransportation (unless it malfunctions). Others imply that teletransportation is certain death. Now, imagine that a philosopher is caught on the seventeenth story of a burning building. He has some hope, but no certainty, of the ordinary sort of rescue. Then he is offered escape by teletransportation, provided he accepts the invitation right away. At that point, I think his philosophical opinion may very well guide his decision. *If he thinks what I do, he will accept teletransportation* even if he reckons his chance of ordinary rescue to be quite high.
>
> [footnotes omitted, emphasis added]

Suppose that in the T-world humans are also regularly scanned as insurance against unforeseen death (e.g., by murder or bad accident); call any consequent replication *reinstatement*. The apparent difference between relocation and reinstatement is that in the latter, the scanned subject has two continuers: the short-lived stage whose death prompts the replication; and the longer-lived replica. Is reinstatement resurrection or replacement? To begin, here is Lewis in his final publication [11] (p. 12):

> Suppose you are about to be beamed up, and you know that the signal will be received both on the starship *Enterprise* and on the starship *Potemkin*. Let's assume that beaming up works not by transmission of matter, but by transmission

> of structural information. That guarantees causal continuity in all bodily and mental respects. You will survive twice over. (What does it matter that you will be made of different atoms afterward? Atoms are the ultimate interchangeable parts, and most of them will be replaced within a few years anyway). Should you expect to find yourself aboard the *Enterprise* or aboard the *Potemkin*? Both. One of your future selves will be aboard one and another will be aboard the other . . . .
>
> Suppose you're about to be beamed up, with the signal received both on the *Potemkin* and on the *Enterprise*. At the last moment you find out that the receiver on the *Enterprise* is malfunctioning: anyone transported there will be dead on arrival, or very soon after. What to expect? No worries, you'll be safe and sound aboard the *Potemkin*. Your death branch should not figure in your expectations.

Here Lewis seems to be drawing upon his treatment of fission in [9,12] where he argues that the R-relation is near enough to identity. The I-relation holds between two temporal parts of one and the same person, and Lewis argues that the R-relation is the I-relation. Since strict identity and its cognate I-relation are each symmetric, Lewis posits a symmetric R-relation. Lewis writes [9] (pp. 23–24):

> If a stage $S_2$ is mentally connected to a previous stage $S_1$, $S_1$ is available in [quasi-] memory to $S_2$, and $S_2$ is under the [quasi-] intentional control of $S_1$ to some extent—not the other way around. We can say that $S_1$ is R-related *forward* to $S_2$, whereas $S_2$ is R-related *backward* to $S_1$ . . . . $S_1$ and $S_2$ are R-related *simpliciter* if and only if $S_1$ is R-related either forward or backward to $S_2$ . . . .
>
> In a case of fission, for instance, we have a prefission stage that is R-related forward to two different, simultaneous postfission stages that are not R-related either forward or backward to each other.
>
> [emphasis original]

Hence, the R-relation is not transitive. To illustrate this, suppose that Yuri is beamed from the *Hood* and is the one who knows that he will have two long-lived continuers on the *Enterprise* and *Potemkin*. Yeva is beamed from the *Hood* and knows that she will have one long-lived continuer on the *Potemkin* and that her replica on the *Enterprise* will arrive alive and awake but die soon thereafter [5]. In both cases, suppose that Yuri, Yeva and their continuers never undergo any other fissions or fusions. Since a person is a maximal aggregate of R-related person-stages, Lewis would say that the Yuri and Yeva cases are fissions each involving exactly two persons. In each case, a person-part exists up until the scan: call them $\text{Yuri}_\text{H}$ and $\text{Yeva}_\text{H}$. After the replications, in each case, there are two distinct stages on board each ship; call them $\text{Yuri}_\text{E}$ and $\text{Yuri}_\text{P}$, who are not R-related to each other, and $\text{Yeva}_\text{E}$ and $\text{Yeva}_\text{P}$, likewise not R-related to each other. The Yuri case has two persons, ($\text{Yuri}_\text{H}$ + $\text{Yuri}_\text{E}$) and ($\text{Yuri}_\text{H}$ + $\text{Yuri}_\text{P}$); the Yeva case also has two persons, ($\text{Yeva}_\text{H}$ + $\text{Yeva}_\text{E}$) and ($\text{Yeva}_\text{H}$ + $\text{Yeva}_\text{P}$). On Lewis's view, the pre-fission $\text{Yuri}_\text{H}$ was a common part shared by two persons; ditto for $\text{Yeva}_\text{H}$. Each person sharing $\text{Yuri}_\text{H}$ wants to survive beaming, and their desire to survive must include a plural desire; of the *strong* form *let all of us survive* or of the *weak* form *let at least one of us survive*. In the Yuri case, both the strong and the weak forms would be satisfied. But Yeva is a case of survival, too. According to Lewis, there is a weak ordinary desire to survive that is satisfied in both cases [12] (pp. 75–76).

Back on our T-world, suppose that Stan has never relocated, but is scanned regularly. At the age of 25, Stan is murdered, dying instantly exactly one day after his most recent scan. That scan is used to reinstate Stan exactly one day after the murder. Call the 25-year-old worm that exists up until the scan $Stan_1$, the one-day stage between the scan and the murder $Stan_2$, and the replica $Stan_3$; and for simplicity suppose that $\text{Stan}_3$ is never relocated or reinstated. The Stan case seems more like the Yeva case than the Yuri case, so it seems we should interpret it as a fission case with a death branch. That seems to be Lewis's view when discussing an analogous case [12] (p. 75):

> [C]onsider a system of survival insurance . . . . From time to time your mind is recorded; should a fatal accident befall you, the latest recording is played back into the blank brain of a fresh body . . . . [T]he fission occurs at the time of recording . . . . This system satisfies the weak desire for survival, but not the strong desire.

Aggregating R-related stages we count two overlapping persons. $Stan_1$ wants to survive, but as a shared stage has the weak plural desire *let* ($Stan_1$ + $Stan_2$) *or* ($Stan_1$ + $Stan_3$) *survive*. By day two after the murder, only the latter survives, but that is good enough for Lewis. With this in place, suppose that Good Suzy World is a T-world, and that like Stan, Suzy is not a relocator. Suzy does not try to kill Baby Suzy, and Baby Suzy was regularly scanned but never reinstated. If Suzy had tried and succeeded in killing Baby Suzy, then no biggie: like Stan, Baby Suzy would have been reinstated from the most recent scan exactly one day before $t$ (the time of death), and the replica who began exactly one day after $t$ would have grown up into (Bad) Suzy. So Suzy can kill Baby Suzy.

But there is a problem. Stipulate that Baby Suzy already counts as a person, and already has whatever counts as the ordinary desire to survive. In Good Suzy World, Baby Suzy is one and the same person as Suzy, so that Suzy killing Baby Suzy would have to count as autoinfanticide. Bad Suzy in Bad Suzy World just described is closely analogous to Stan in Good Suzy World; but we might be troubled by her relevance to the *Good Suzy* story. By hypothesis, Good Suzy has no shared stages, but she nevertheless has the weak plural desire for survival (since according to Lewis that is a desire that ordinary folk have). In Bad Suzy World, there are two persons sharing the Baby Suzy stage, $BS_1$. There is the one-day stage $BS_2$, and there is the replica $BS_3$. On the fission reading, the two persons are ($BS_1$ + $BS_2$) and ($BS_1$ + $BS_3$). It seems plausible by Lewis's account that adult Bad Suzy is adult Suzy. But who kills who? $BS_3$ kills $BS_2$, and they are (parts of) two different persons! Hence, on the fission hypothesis, Bad Suzy World seems not to be a world in which Suzy kills Baby Suzy—indeed, it is not an autoinfanticide world at all. It is a replacement world.

Two responses are open to Lewis. The first is to appeal as Lewis does to a general inconstancy in *de re* modal judgments— such as saying *this* thing could have killed *that* thing—claiming that the same actual thing can be multiply represented at another possible world. For Lewis, cross-world judgments of personal identity are analyzed in terms of a modal counterpart relation mediated by resemblance, and the counterpart relation does not always behave like the identity relation. Which way we represent *de re* is heavily affected by context, governed by a Rule of Accommodation according to which there is a presumption that utterances be interpreted as true [8] (pp. 248–263). But whatever success this inconstancy reply enjoys, as long as we interpret the Bad Suzy case as a fission, it remains true that Bad Suzy does not commit autoinfanticide. Hence, I shall pursue a different strategy.

*Reinstatement Restated*

The strategy is to argue that Lewis can maintain the relevance of Bad Suzy World without invoking inconstancy, by denying that the Bad Suzy case is a fission. It is time to reconsider how we view a reinstatement like Stan's. Lewis has in effect described four cases: the Yuri case with two life branches; the Yeva case with one life branch and one death branch; the relocation case; and (in a slightly different version) the Stan case. The Yuri case is definitely a fission; the relocation case is definitely a nonfission. Lewis does not explicitly say that the Yeva case is a fission, but we should presume he thinks so since he explicitly says that a case like Stan's is a fission.

In discussing fissions Lewis is mainly concerned with the forward-looking attitudes of the prefission subjects. Consider instead some backwards-looking attitudes. Suppose that Stan at 50 has two life-long friends. Dan, who is also 50, has never relocated or been reinstated. Dan contemplates what would have happened had he been killed at 25. He knows the facts about the timing of his scans, and judges that had *he* been killed at a certain moment at 25, *he* would have been reinstated from the most recent scan, taken 24 h earlier.

Is he wrong? After all, Dan reasons, that is what happened to Stan: *he* died at 25 and *he* was reinstated. On the other hand, their friend Ivan had not been scanned since he was 25, but had no need of relocation or reinstatement until just last week, when killed at the age of 50. 24 hours later, a replica was produced from the 25-year-old scan, and that replica is now chatting with Dan and Stan. The replica says that *he* was killed last week, and *he* was reinstated. Is he wrong? Yes, Dan reasons. The replica looks 25 and does not even remember the last 25 years of their friendship. Whoever he is, *he* was not killed last week. Dan remembers how grateful he was to get his friend Stan back after reinstatement. But this replica is no Ivan. In fact, he's only making things worse.

I think a Lewisian can make sense of Dan's judgments. The Stan case is quite unlike the Yuri case. It shares with the Yeva case the existence of one life branch and one death branch, but the death branch does not co-exist at the same external time as the life branch. In that respect the Stan case is more like relocation. Moreover, there is no causal dependence of the life of $\text{Yeva}_{\text{H}}$ on the death of $\text{Yeva}_{\text{E}}$. By contrast, $\text{Stan}_3$ only exists because $\text{Stan}_2$ dies. The last stage of $\text{Stan}_2$ and the later first stage of $\text{Stan}_3$ are very, very similar. The similarity is no coincidence, and very strong bonds of counterfactual dependence are in place: had $\text{Stan}_2$ been mentally very different then so would $\text{Stan}_3$ have been mentally very different. So the death of $\text{Stan}_2$ causes the existence of a stage whose mental states are counterfactually highly dependent upon $\text{Stan}_2$'s mental states, and the counterfactual dependence is normal in direction, with states later in external time counterfactually dependent upon earlier states. The Lewisian instinct is after all to analyze causal dependence in terms of counterfactual dependence [13]; so for Lewis, Stan's reinstatement is quite close to relocation, and quite close to ordinary survival.

The Ivan case is different again. Call the worm that lives until 25 $\text{Ivan}_1$, the worm from 25 to 50 $\text{Ivan}_2$, and the replica $\text{Ivan}_3$. $\text{Ivan}_2$ does not coexist at the same external time as $\text{Ivan}_3$, so his case is in that respect unlike the Yuri and Yeva cases, and more like the Stan case. It is also like the case of Stan in that the death of $\text{Ivan}_2$ causes the existence of $\text{Ivan}_3$. But the Ivan case is quite unlike the Stan case in that the bonds of similarity and counterfactual dependence between $\text{Ivan}_2$ and $\text{Ivan}_3$ are very much weaker. The lesson seems to be that very short-lived death branches are unproblematic, but the longer they last, the more problematic they become [6].

So I believe that the Lewisian can say that reinstated Stan is one person consisting of $\text{Stan}_1$, $\text{Stan}_2$, and $\text{Stan}_3$, with a one-day gap in his existence. And if that is true of Stan, then Bad Suzy is a single person consisting of $\text{BS}_1$, $\text{BS}_2$, and $\text{BS}_3$, and Bad Suzy World is straightforwardly a world where Bad Suzy commits autoinfanticide. So *she* died and *she* was reinstated. To summarize, I have argued that there are two potential ways for Lewis to endorse reinstatement as a means of personal survival, and to hold that there are possible worlds where Suzy kills Baby Suzy and Baby Suzy is reinstated. Now, I must make such worlds relevant to the assessment of Vihvelin's counterfactuals.

## 4. The Relative Closeness of Reinstatement

Suppose that Vihvelin grants that relocation or reinstatement at a T-world is survival. She might yet claim that T-worlds are still not relevant to the counterfactuals, even for Lewis, since Good Suzy World is not a T-world. In one version, this response might assert that relocation and reinstatement though logically possible are nomologically impossible. But Lewis is not bound to agree. Reinstatement is at most *technologically* impossible, and even that is doubtful. We already know how to kill, so we just need advanced enough scanners, and advanced enough 3-D printers [7]. Whereas Vihvelin thinks that time travel worlds are more like ours than any resurrection worlds are, for Lewis that judgment if anything seems to be reversed. Vihvelin [2] (p. 323) writes "I think that time travel is possible at worlds very much like ours, maybe exactly like ours"; whereas according to Lewis [1] (p. 145), "a possible world where time travel took place would be a most strange world, different in fundamental ways from the world we think is ours" [8].

Suppose Vihvelin grants that relocation and reinstatement are nomologically possible. And note two things. Good Suzy World is described in a fictional story—*Good Suzy*, told above—and according to Lewis's own account [14], what is true in a fictional story is what would have been true had the story been instead told as known fact [9]. Applying Lewis's own treatment of counterfactuals to that analysis, if the actual world is not a T-world, then Good Suzy World is not a T-world either, since adding relocation or reinstatement would be a gratuitous change. Moreover, if Good Suzy World is not a T-world, then neither is Bad Suzy World. Hence, Bad Suzy would fail to kill Baby Suzy.

If Good Suzy World is not a T-world, then neither is Bad Suzy World; merely attempting to kill someone will not license adding relocation or reinstatement technology to the world. But the other two steps in the argument just given on behalf of Vihvelin are dubious. Lewis does not think that time travel to the past occurs in the actual world, but he might think that our world is a T-world simply awaiting more advanced technology. And if so, then Good Suzy world—where time travel technology has been developed—likely is a T-world, too. But suppose ours is not a T-world. Vihvelin claims [2] (p. 323):

> But in any case there is no reason to suppose that there is any connection between time travel-permitting laws and resurrection-permitting laws.

This is plausible for the corpse-resurrection Vihvelin has in mind, but less so for relocation or reinstatement. Lewis leaves room for cases of *instantaneous* time travel, where a journey through external time takes no personal time at all [1] (p. 146). Lewis does not describe any such cases. He describes only devilish *Fred-cum-Sam* cases, which might appear to be time travel but are not [1] (p. 148). Consider instead DerfMas: Mas is born and lives a normal life until he is vaporized by a time-traveling demon who remembers his entire final qualitative state, and then uses that knowledge to produce Derf in the past. Derf appears as if in the midst of life, and lives normally from then until an ordinary death, before Mas is born. The demon ensures that Derf's initial qualitative state exactly (or as much as is physically possible) resembles Mas's final state. DerfMas is one person, and time travels instantaneously to the past, by Lewis's account.

Derfmas is nomologically impossible, thanks to the demon. But as long as the laws permit information to be sent into the past, there is no need for the supernatural; relocation can be used to send time travelers to the past. If ours is not a T-world, and granting Lewis that it is not a time travel world either, then Good Suzy World would be a T-world if the closest time travel worlds employ instantaneous relocation time travel. Nothing in Vihvelin's argument rules this out. I conclude that Vihvelin so far fails to show that Bad Suzy *would* fail to kill Baby Suzy.

## 5. The Metric of Overall Similarity of Worlds

Vihvelin [4] employs Lewis's own account [15] of the metric of overall similarity of worlds designed to be plugged into his "Analysis 2" of counterfactuals. Lewis supposes—as I shall in what follows—that the laws of nature are deterministic, but that nomological possibility nevertheless allows that on many occasions things could have happened differently. Pick one such occasion, at time $o$. Lewis describes four different kinds of possible world where the antecedent of a standard counterfactual about what happens at $o$ is true (in each case, I will assume that the centered world is our actual world). The one that counts as closest, $w_1$—I will call it a *first-rate* world—is in external time exactly like the actual world up until just before $o$, when there is one "small miracle"—a departure from our laws—and divergent thereafter (divergent with respect to the pattern of events, not with respect to the laws). A *second-rate* world $w_2$ matches our deterministic laws exactly, but differs in what happens at $o$, and so is never exactly like ours in matters of particular fact. A *third-rate* world $w_3$ is exactly like ours until just before $o$, when there is a small miracle, and then approximately but not exactly like ours thereafter, in virtue of a second small miracle that prevents the more drastic consequences of the first small miracle. A *fourth-rate* world $w_4$ is exactly like ours until just before $o$, when there is a small divergence

miracle, followed immediately by a big convergence miracle which in effect undoes the small miracle, and so is exactly like ours again thereafter. Lewis writes [15] (p. 472):

> Under the similarity relation we seek, $w_1$ must count as closer to [the centered world] than any of $w_2$, $w_3$, and $w_4$. That means that a similarity relation that combines with Analysis 2 to give the correct truth conditions for counterfactuals such as the one we have considered, taken under the standard resolution of vagueness, must be governed by the following system of weights or priorities.

(1) It is of the first importance to avoid big, widespread, diverse violations of law.
(2) It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.
(3) It is of the third importance to avoid even small, localized, simple violations of law.
(4) It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly.

At step (1) we eliminate fourth-rate worlds, at step (2) we eliminate second-rate worlds, and at step (3) we eliminate third-rate worlds, leaving a first-rate world as the closest, on balance, to the actual world. It is crucial to Lewis's account that we balance both similarity in the pattern of events *and* similarity in the laws. This is to block the "future similarity objection" that a metric of overall similarity must favor third-rate worlds where a second small miracle prevents a more drastic future difference in the pattern of events. Anticipating a little, call such third-rate worlds *banana peel* worlds. Grant that if the centered world is the actual world, had Nixon pressed the wrong button, then there would have been a nuclear holocaust—the system was reliably set up that way—and grant also that there never will be an actual nuclear holocaust. In a banana peel world, Nixon presses the button, but a banana peel that does not actually exist *would have* existed to prevent the holocaust. If we count only the pattern of events, then given the difference that a nuclear holocaust would make, a banana peel world arguably comes out closest, such that the counterfactual we began with wrongly comes out false. However, once we include the laws in the metric, then given the difference to the laws that a second small miracle makes, banana peel worlds lose, given condition (3) and buttressed by condition (4).

Vihvelin [4] summarizes her employment of Lewis's metric as follows (emphases original):

> The closest worlds where Suzy's attempt to kill the baby succeeds are worlds with *one small* and *one big miracle*, whereas the closest worlds where Suzy's attempt fails are worlds with, at most, two *small miracles*. Since Lewis's theory says that worlds with one or even two small miracles are closer than worlds with one big miracle, his theory says that worlds where Suzy's attempt fails are closer than worlds where her attempt succeeds [10].

Well, hold on. Vihvelin is right that there are possible worlds where a baddish Suzy tries to kill Baby Suzy and fails with no need of a second small miracle. But such worlds are first-rate only with respect to centered worlds that have built-in fail-safes—booby traps, motion-detectors, and the like—that would have stopped almost anyone from killing Baby Suzy. That includes an intrinsic duplicate of Suzy; yet Vihvelin grants that such a duplicate in Suzy's place could have killed Baby Suzy [2] (p. 327). Hence, such fail-safe worlds are irrelevant to Vihvelin's argument, since the centered goodish Suzy world where we evaluate the counterfactuals will also have those fail-safes. By contrast, I stipulated that Good Suzy World has no fail-safes, so Bad Suzy World has none, either. Hence, Vihvelin is committed to Bad Suzy failing because of a banana peel that does not exist at Good Suzy World.

On the other side of the ledger, Vihvelin need not deny the nomological possibility of relocation or reinstatement. Good Suzy World does not serve Vihvelin's argument if it is a T-world such that Baby Suzy has been recently scanned. That gets Lewis a first-rate world too easily—his own built-in fail-safe—and Suzy can kill Baby Suzy. This seems to at least narrow the scope of Vihvelin's conclusion, but to be fair, let's make things as bad as we can for Lewis. Suppose that Good Suzy World is a T-world, but stipulate that Baby

Suzy's parents are fanatical members of the van Inwagen Society and have refused to ever have Baby Suzy scanned. Now, there is trouble, for Vihvelin will claim that any success worlds will have to be big miracle worlds. (They will not be quite fourth-rate, since they do not require the entire world to reconverge on the centered world, but they are very, very distant from the centered world). Vihvelin would claim in such a case that Bad Suzy World is therefore a world in which Suzy fails. But what stops Bad Suzy? A second small miracle. So Bad Suzy World is a third-rate, banana peel world.

Or is it? A gap remains in Vihvelin's argument, for there is another important class of worlds it has not yet taken into account.

**6. Bad-but-Smart Suzy**

In the movie *Bill and Ted's Excellent Adventure*, the eponymous teens when faced with an uncooperative world repeatedly take advantage of the fact that they can time travel to produce desired outcomes. At one point they need Ted's father's keys to the police station, so they decide at time $k$ to later use their time machine to travel back to two days before $k$ and borrow the keys, which they can then leave behind a nearby sign to be found at $k$. They look behind the sign and retrieve the keys. The world must after all cooperate, and part of the setup is that the keys in fact have at $k$ been missing for two days. And of course Bill and Ted or someone else must follow through with the future time travel journey, and get the job done in the past.

Consider a possible world where *Smart Suzy* behaves like Bill and Ted. Smart Suzy World follows the *Good Suzy* story up until she chucks Baby Suzy under the chin, but then a villain breaks into the room, kills Baby Suzy, and escapes. Unexpected! As far as Smart Suzy knows her parents never had her scanned; but she must have been scanned, anyway. Someone must have arranged it, but who, and why? Smart Suzy leaves discreetly and visits a nearby scan bank. Sure enough, they have a recent scan of Baby Suzy on file and Suzy orders reinstatement. Suzy then uses her time machine to travel to a time earlier than her first visit, briefly kidnaps Baby Suzy and gets her scanned. (The order of these events in Suzy's personal time could be altered: she can go back to get the scan made before she orders the replica made). Now, apply this reasoning to the counterfactuals true in the *Good Suzy* story. Big and strong as she is, Bad Suzy crushes Baby Suzy's windpipe; her attempt to retro-kill succeeds. But Bad Suzy is also smart, and thereafter she uses time travel and reinstatement to ensure her own survival. She sneaks the replica back into the crib, thereby relieving her none-the-wiser parents of their grief, and leaving them believing that something like corpse-resurrection has occurred [11].

The question is how a Bad-but-smart Suzy World fits into Lewis's metric when it is centered on Good Suzy World. I shall argue that a Bad-but-Smart Suzy World is first-rate, and so beats out any banana peel world. So it is true in *Good Suzy* that had Suzy tried to kill Baby Suzy, someone would have had to have time traveled to a previous time and arranged a scan for Baby Suzy. (I am here assuming a competent attempt). Even if that is false, had Suzy tried to kill Baby Suzy, she *might* have succeeded, and so it might have been that someone would have had to have time traveled to a previous time and arranged a scan for Baby Suzy. Given the Counterfactual Possibility Principle, by Lewis's own account, it is true in *Good Suzy* that Suzy can kill Baby Suzy. No big miracle needed; autoinfanticide is no biggie.

Or so say I. Lewis never tells us explicitly how time travel counterfactuals are to be handled. But he does give a relevant judgment concerning a case of foreknowledge that is a version of the Newcomb Problem [16] (pp. 126–127):

> If we put a human predictor in place of God, and we ask again what would have been the case if I had declined the $1000, the answer will depend on the predictor's modus operandi. First case: the predictor is a time traveler. He saw me accept the $1000, then departed to the past taking his knowledge with him. His foreknowledge is causally downstream from its object. Then I want to hold

> fixed that the time traveler has foreknowledge, and say that if I had declined, the time traveler would have known that I was going to decline . . . .
>
> Second case: the predictor is an expert psychologist, who knows past conditions and regularities of cause and effect. His foreknowledge and its object are separate effects of common causes. Then I want to hold the past fixed, and say that if I had declined, I would have violated some one of the regularities the psychologist relied on.

For Lewis, the crucial difference between the time traveler and the expert psychologist is that the former employs reverse causation to make his pronouncement. Although you should two-box in a Newcomb Problem even with a 100% accurate predictor, in which case you receive $1000, if time travel enables the foreknowledge then you should one-box and receive $1 million. I need to show how this works.

## 7. A Forking Miracle

Suppose that in the actual world, @ Lewis one-boxes in the time travel Newcomb game. The counterfactual judgment in this case involves a serious back-tracking argument, but the time traveler's "prediction"—unlike the psychologist's—is caused by Lewis's choice [12]. Call the time of return from the future *t2*, and the time of Lewis's one-boxing choice *t3*; then the prediction occurs between *t2* and *t3* (Figure 1).

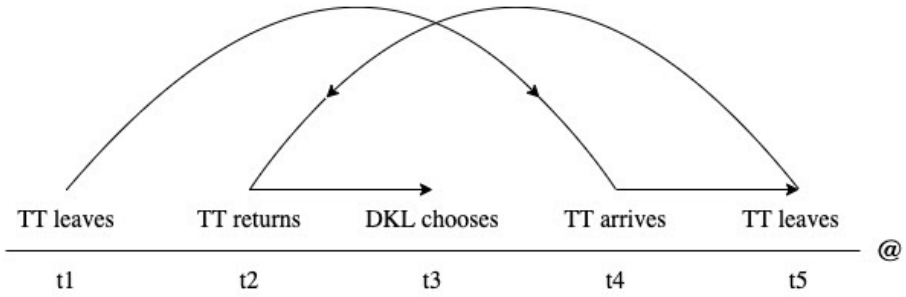


**Figure 1.** The time-traveling Newcomb game predictor.

The one-boxing Lewis will defend his choice by counterfactual reasoning: *If I had two-boxed at t3, then the predictor would have to have predicted that I would two-box at t3, and there would have to have been no $1 million placed in Box B. So if I had two-boxed at t3, I would have gotten only $1000.*

In the nearest two-boxing world—call it *TB*—it cannot be true that there is exact match of particular fact with @ until just before *t3*, since TB is already different from @ at *t2*, when the time traveler returns from the future with different beliefs. TB need not be different from @ at *t1*, so I shall assume that TB exactly matches @ until just before *t2*, when there is a small miracle.

Now, to the point. In TB, the time traveler's prediction does not cause Lewis's choice of two boxes. Is a second small miracle required, just before *t3*, to produce Lewis's different decision? Suppose that is true. Then by Lewis's metric, TB is a third-rate world, roughly as close to @ as a banana peel world is. By analogy then, Bad-but-Smart Suzy World contains two small miracles: the first just before Bad Suzy's arrival from the future at *t** (say, one day before *t*), and the second just before Bad Suzy's attempt on Baby Suzy's life. But that makes Bad Suzy World only *almost* as close to Good Suzy World as Vihvelin's banana peel world is. Since Lewis at step 2 tells us to *maximize* the region of exact match of particular fact, the banana peel world still apparently wins (with a drastically smaller margin of victory than Vihvelin claims).

But Lewis can do better. Notice that the argument just given for a second small miracle in TB rests on the fact that the prediction does not cause Lewis's choice; but it ignores the existence of reverse causation—the fact that Lewis's choice causes the earlier prediction.

For a centered deterministic world without reverse causation, a standard counterfactual judgment postulates one small miracle, and all the salient divergence from the centered world is causally traceable to the time of that miracle (any other divergence would be gratuitous). But it is misleading to think that the miracle *causes* the divergence, for the miracle is a difference in the laws, not a difference in the pattern of events. Better to say that the small miracle *permits* the divergence in the pattern of events. Then we are free to ask what the pattern of causal relations is between the events in the permitted divergence.

Thanks to the reverse causation, in the nearest world to TB where Lewis two-boxes, one small miracle permits a double divergence in causation. The first salient difference is the difference in the predictor's beliefs at *t2*, and this causes a different prediction, which in turn causes a different prize to be awarded. That and its consequences are one part of the divergence, all of which lie in the external future. In the second part of the divergence, Lewis chooses differently and that has its causal consequences, most of which lie in its external future, but some of which lie in its external past. If it helps, think of the divergence in causation as being doubly present thanks to the reverse causation, by analogy with the way a time traveler can be doubly present by traveling into their own past. When one small miracle permits a double causal divergence, call it a *forking miracle*.

Now, I can give my final judgment on what is true in *Good Suzy*. Had Suzy tried to kill Baby Suzy, she would or might have succeeded, and if she had succeeded someone would have had to have time traveled to past time $t^*$ to arrange the scanning of Baby Suzy, ensuring Baby Suzy's reinstatement after $t$. Thanks to a forking miracle just before $t^*$, the laws permit the appearance of a time traveler from the future, caused by the (later) killing of Baby Suzy at $t$.

## 8. It Is Not the Size of the Miracle, It Is What You Do with It

But wait—is the appearance at $t^*$ of a time traveler from out of nowhere not kind of a biggie? Not really, and I do not believe Vihvelin would think so. Lewis's account of his metric should not be read too narrowly. What Lewis calls a *small* miracle Lewis also calls "localized", but it could as well be called a one-off miracle. When making counterfactual judgments we move as far as we need to from the centered world to make the antecedent true, but no further. Any further change would be gratuitous.

In describing the Nixon case, Lewis describes the small miracle that facilitates Nixon's counterfactual pressing of the button as follows [15] (p. 468):

> The deterministic laws of [the centered world] are violated at $w_1$ in some simple, localized, inconspicuous way. A tiny miracle takes place. Perhaps a few extra neurons fire in some corner of Nixon's brain.

This is potentially misleading, since small miracles don't have to be as small as that one; rather they must be as small as possible to avoid gratuitousness. And if time travel to the past is logically possible, small miracles must permit time travelers to appear in the past in a first-rate counterfactual world at a point earlier than they appeared in the centered world. Suppose that in TB the predictor reliably comes back 15 min earlier with a two-box prediction than it does with a one-box prediction. Then had Lewis two-boxed, the predictor would have had to have arrived from the future 15 min before *t2*. The miracle required to permit that difference does not seem "simple" or "tiny" or especially "inconspicuous".

Lewis's distinction between big and small miracles tempts us to think that a banana peel world is quite like a first-rate world, and quite unlike a fourth-rate world. I see it differently. The metric rules out gratuitous law changes, and in that respect a banana peel world is quite like a fourth-rate world, and quite unlike a first-rate world. The big/small distinction is misleading, because it is not absolute size that matters. Vihvelin's own judgments reflect this. Concerning a centered world such as Good Suzy World, Vihvelin [4] invites us to compare:

(e) If Suzy tried to kill Baby Suzy, she failed.

(f) If Suzy had tried to kill Baby Suzy, she would have failed.

. . . It's not just that (e)—the indicative conditional—is true. (f) also seems to be true.

Stipulate that Vihvelin is correct about (f)—say, because the success would require corpse-resurrection. Corpse resurrection is eliminated at the first step, since it occurs in the set of big miracle worlds. As Vihvelin [4] puts it, these are:

[W]orlds where the baby dies but is subsequently resurrected from the dead and grows up to be the adult Suzy; (These are worlds where, in addition to the small divergence miracle that enables Suzy's attempt, there is a big miracle).

But now consider another pair of conditionals, also evaluated at Good Suzy World:

(e*) If Suzy killed Baby Suzy, Baby Suzy was corpse-resurrected.

(f*) If Suzy had killed Baby Suzy, Baby Suzy would have been corpse-resurrected.

By Vihvelin's account, indicative conditional (e*) is true if corpse-resurrection is required for success. Counterfactual conditional (f*) is true, too, since corpse-resurrection is logically possible. But then, corpse-resurrection must not be a big miracle, else that world would be eliminated at the first step in Lewis's metric. So when evaluating counterfactuals from Good Suzy World, the very same corpse-resurrection world contains a big miracle with respect to (f) but not with respect to (f*). By the same token—as in my treatment of Suzy's counterfactual attempt at autoinfanticide requiring a divergence beginning with a time traveler appearing in the past—even if such appearances would be big miracles in other contexts, that does not show it is a big miracle with respect to *Good Suzy*.

The point about miracle size arises with respect to banana peel worlds as well. Vihvelin postulates a preventative second small miracle, but she does not tell us what it involves. Does the signal from Bad Suzy's brain disappear en route to her arm and hand muscles? That seems, on balance, not a case of an attempt to kill. Given an attempt really carried out—for instance, suppose that Bad Suzy closes her strong grip on Baby Suzy's windpipe, an action that would ordinarily crush it beyond repair—what ordinary occurrence stops her from succeeding? The miracle in question might then need to be quite sizeable, but once again, absolute size is strictly irrelevant to the Lewisian view. What makes a miracle a small miracle is a matter of its not being gratuitous, and here Vihvelin loses the argument.

## 9. Strange Shackles Indeed

Here is another way to see the same point. Although Vihvelin does not tell us what in particular foils Bad Suzy's attempt, she does say the attempt fails because of the laws of nature [3] (p. 324):

> My arguments support the claim that any time-travel world where any person succeeds in killing the person that is her younger self is a world which includes events that are miraculous by the standards of our laws . . . . So I think we should conclude that the killing of one's younger self is *nomologically impossible* and that is why no one has the narrow ability to do such a thing.
>
> [emphasis original]

Vihvelin then rebuts an objection from Ted Sider [17] that her view after all requires a sort of temporal censor, and that such strange metaphysical "shackles" are better avoided [3] (pp. 324–325):

> I agree with Sider [on the desideratum]. But I deny that my argument commits me to any strange shackles or "exotic metaphysical add-ons".
>
> A time traveler trying to kill her infant self is like a person trying to build a perpetual motion machine . . . . If you try to build a perpetual motion machine you will fail . . . .
>
> And it's not just that anyone's actual attempts have happened to fail, every *counterfactual* attempt *would have failed* as well. If anyone, be they Edison or Elon Musk, had tried to build a perpetual motion machine, they *would have failed*. Not because of exotic metaphysical "forces" or "guardians" or "shackles" but because the creation of such a machine would contradict the laws of thermodynamics . . . .

> Just as the laws of nature entail the impossibility of perpetual motion machines, they also entail that people killed in infancy do not go on to become murderous adults. To suppose that the time traveler could succeed in killing her baby self is to suppose that these laws are false.
>
> [emphases original]

We are now in a position to use Lewis's metric of overall similarity of worlds to give a more nuanced version of Sider's complaint, and show that Sider is correct after all. Vihvelin's argument does postulate metaphysical shackles, and they are strange shackles indeed. Suppose Musk never tries to build a perpetual motion machine. It is true that had he tried, then he would have failed. Assuming determinism, the nearest relevant world is first-rate, and has different laws to ours, but the difference is the one small miracle required to permit the difference in the pattern of events that includes the attempt and its different causal consequences. We should think of a first-rate world as one in which the laws are the same as in the centered world from just after the small miracle. In the general schema, the small miracle occurs just before $o$. So a possible world which is all and only a duplicate of just the part of $w_1$ from $o$ onwards has the same deterministic laws as a possible world which is all and only a duplicate of just the part of the centered world from $o$ onwards. Same deterministic laws, different starting conditions, and so a different pattern of events.

The future similarity objection to Lewis's analysis of counterfactuals claims that Lewis's metric of overall similarity must favor a third-rate world which includes a second small miracle: a banana peel that foils Nixon's holocaust attempt, for instance. Notice what we should not even try to say. We should not say that *our* laws require any second small miracle. Quite the reverse: *our* laws *rule out* such a miracle—that is the point of Lewis's reply to the objection. The closest world is a first-rate world because that is a world with the same laws as the centered world, excepting only the small miracle required to permit the antecedent to be true. So Vihvelin is right about Musk's counterfactual attempt at perpetual motion. In a first-rate world he tries and fails, since *our* laws require the failure. He does not fail because of a sui generis banana peel. Musk fails because our laws have the fail-safes built in; they would foil anyone.

But this is manifestly not true of Vihvelin's postulated banana peel world. As we saw in Section 5, fail-safe worlds which would foil almost anyone are not relevant to Vihvelin's argument, and I therefore stipulated that Good Suzy World has no such fail-safes. So it is misleading at best to say as Vihvelin does that our laws—or more precisely, the laws of Good Suzy World—require a second small miracle that foils Bad Suzy's attempt. Instead, Vihvelin must postulate that Bad Suzy World has different laws, laws that foil in Bad Suzy World in a way that the laws of Good Suzy World do not. The laws of Good Suzy World require that there be no such extra miracle. So the extra miracle is indeed an exotic metaphysical add-on, a strange shackle of the sort Vihvelin agrees we should avoid if possible.

## 10. Conclusions

My response to Vihvelin has been long and involved. That is the nature of the beast. In the final analysis, though, Lewis's position is not only defensible but sensible. Time travel fiction is full of folks who really ought to know better trying to do things that just will not happen. Good Suzy is different, and better, and never even tries to kill her Baby self. Bill and Ted are also different, not trying to change things, but wisely using time travel to bring about desirable results. Bad Suzy is different again, because she by her own unwise actions is forced to emulate Bill and Ted. Like Good Suzy, Bad Suzy should simply have left well enough alone. Like Good Suzy, Bad Suzy should have reasoned: *I **could** kill Baby Suzy, but why would I? Why would I even try?*

Ultimately, no big miracle is needed for Suzy to succeed in killing Baby Suzy; autoinfanticide is no biggie, at least for Lewis. But there is more work to be done. To give a more general defense, I must defend the Lewisian account of survival against its many detractors, but that is a task for another time. Second, I have defended Lewis only given determinism.

I myself am a determinist, but Lewis is not, so a complete defense of Lewis would have to accommodate his application in [18] of the metric to indeterminism. Finally, both Vihvelin and Wasserman [5] independently object to Lewis's metric of overall similarity of worlds even for deterministic laws. I believe Lewis's metric needs some revision, but predict that the revisions will not undermine Lewis's position on what time travelers can do. Again, that is a topic for another time.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The author declares no conflict of interest.

## Notes

1 It's unclear whether Lewis here is using "ability" stipulatively or trying to capture the platitudes surrounding its use. Against the latter, it seems we often say we were "unable" to do something we could not do in the "luck" sense.

2 Lewis gives a second reason for the logical impossibility of changing the past in a single timeline, an argument which appeals to atomism about temporal duration, using "moment" as a technical term for a temporal atom [1] (p150). Though I am a fellow atomist, I prefer the more general appeal to flat contradiction.

3 Vihvelin shows that her view generalizes to some other cases involving reverse causation, where non time travelers are similarly restricted [3] (pp. 322-323). I shall not examine those extra cases here.

4 Some defenses even extend to replacing timelines, such as the branching time versions of the Suzy case suggested by John Carroll [6].

5 If Yeva's were the dead-on-arrival case there would only be one continuer that is a person stage; hence it would not be a fission case.

6 It's not simply the amount of time elapsed, however. It's that given normal changes, more time leads to weaker bonds.

7 One often hears that the scanning and replication technology is nomologically impossible because it violates Heisenberg Uncertainty. At most, this shows that it's impossible to *exactly* replicate. But since ordinary survival does not require that successive stages be intrinsic duplicates, neither does relocation or reinstatement. Whatever is near enough for ordinary survival will be good enough for relocation or reinstatement.

8 To be fair, Lewis at one point describes a relocation scenario as "far-fetched" [10] (p. 192, n. 3), but that is a very different case in which the scanning is somehow done remotely, à la *Star Trek* "beaming".

9 At least, according to Lewis's *Analysis 1*, which assumes that the background truths in fiction are supplied by actual world facts. I shall for simplicity ignore his alternative analysis which appeals instead to a background of mutual shared belief. (Analysis 1 is far more plausible, in any case).

10 Note well that neither Lewis nor Vihvelin is postulating worlds where the laws of *that world* are broken. A miracle big or small is instead a difference between the laws of different worlds.

11 There are other ways the story could go, all dependent upon the fact that Bad Suzy World is a time travel world. Perhaps one of Bad Suzy's parents would not have maintained their anti-reinstatement stance when confronted with Baby Suzy's corpse, especially given Bad Suzy's presence; and hence might be their own daughter's kidnapper. What is certain is that *somebody* does what is needed.

12 Given determinism *every* time-indexed standard counterfactual resolution involves a back-tracking argument, since there had to be a small miracle before the time in question; call these back-tracking arguments *minor*. A serious back-tracking argument takes us back before a different, earlier time index. The presence of a serious back-tracking argument is not sufficient for a backtracking counterfactual resolution; the resolution delivered must also differ in truth value from the "standard" resolution [15] (p. 457). I think Lewis should say that since there is reverse causation the one-boxer's serious back-tracking argument *delivers* the standard resolution.

## References

1. Lewis, D. The Paradoxes of Time Travel. *Am. Philos. Q.* **1976**, *13*, 145–152.
2. Vihvelin, K. What Time Travelers Cannot Do. *Philos. Stud. Int. J. Philos. Anal. Tradit.* **1996**, *81*, 315–330. [CrossRef]
3. Vihvelin, K. Killing Time Again. *Monist* **2020**, *103*, 312–327. [CrossRef]
4. Vihvelin, K. Counterfactuals, Indicatives, and What Time Travelers Can't Do. Available online: www.vihvelin.com (accessed on 8 September 2021).
5. Wasserman, R. *Paradoxes of Time Travel*, Illustrated ed.; Oxford University Press: Oxford, UK, 2018.
6. Carroll, J.W. Ways to Commit Autoinfanticide. *J. Am. Philos. Assoc.* **2016**, *2*, 180–191. [CrossRef]
7. Vranas, P.B.M. What Time Travelers May Be Able to Do. *Philos. Stud. Int. J. Philos. Anal. Tradit.* **2010**, *150*, 115–121. [CrossRef]
8. Lewis, D. *On the Plurality of Worlds*; Blackwell: Oxford, UK, 1986.

9. Lewis, D. Survival and Identity. In *The Identities of Persons*, Revised ed.; Rorty, A.O., Ed.; University of California Press: Berkeley, CA, USA, 1976; pp. 17–40.
10. Lewis, D. Academic Appointments: Why Ignore the Advantage of Being Right? Chapter. In *Papers in Ethics and Social Philosophy*; Cambridge Studies in Philosophy; Cambridge University Press: Cambridge, UK, 1999; Volume 3, pp. 187–200. [CrossRef]
11. Lewis, D. How Many Lives Has Schrödinger's Cat? *Australas. J. Philos.* **2004**, *82*, 3–22. [CrossRef]
12. Lewis, D. Postscript A to "Survival and Identity". In *Philosophical Papers Volume I*; Oxford University Press: Cambridge, UK, 1983; pp. 73–76. [CrossRef]
13. Lewis, D. Causation. *J. Philos.* **1973**, *70*, 556–567. [CrossRef]
14. Lewis, D. Truth in Fiction. *Am. Philos. Q.* **1978**, *15*, 37–46.
15. Lewis, D. Counterfactual Dependence and Time's Arrow. *Noûs* **1979**, *13*, 455–476. [CrossRef]
16. Lewis, D. Evil for Freedom's Sake? *Philos. Pap.* **1993**, *22*, 149–172. [CrossRef]
17. Sider, T. Time Travel, Coincidences and Counterfactuals. *Philos. Stud. Int. J. Philos. Anal. Tradit.* **2002**, *110*, 115–138.
18. Lewis, D. Postscript D to "Counterfactual Dependence and Time's Arrow". In *Philosophical Papers Volume II*; Oxford University Press: Cambridge, UK, 1987; pp. 58–65. [CrossRef]