

## What time travelers may be able to do

Peter B. M. Vranas

Published online: 3 April 2009  
© Springer Science+Business Media B.V. 2009

**Abstract** Kadri Vihvelin, in “What time travelers cannot do” (Philos Stud 81:315–330, 1996), argued that “no time traveler can kill the baby who in fact is her younger self”, because (V1) “if someone would fail to do something, no matter how hard or how many times she tried, then she cannot do it”, and (V2) if a time traveler tried to kill her baby self, she would always fail. Theodore Sider (Philos Stud 110:115–138, 2002) criticized Vihvelin’s argument, and Ira Kiourti (Philos Stud 139:343–352, 2008) criticized both Vihvelin’s argument and Sider’s critique. I present a critique of Vihvelin’s argument different from both Sider’s and Kiourti’s critiques: I argue in a novel way that both V1 and V2 are false. Since Vihvelin’s argument might be understood as providing a challenge to the possibility of time travel, if my critique succeeds then time travel survives such a challenge unscathed.

**Keywords** Time travel · Vihvelin

### 1 Introduction

In “What time travelers cannot do” (1996), Kadri Vihvelin imagines a time traveler (‘Suzy’) who tries to kill the baby who is her younger self (‘Baby Suzy’), and argues that the time traveler cannot succeed. More generally, Vihvelin claims that “no time traveler can kill the baby who in fact is her younger self, given what we *ordinarily* mean by ‘can’” (1996, pp. 316–317). And, according to Vihvelin, “what we ordinarily mean when we say that someone can do something is that she has both the ability and the opportunity to do it. More precisely, we mean that she has the ability to do an act of the relevant kind and that nothing prevents her from exercising this ability” (1996, p. 318).

---

P. B. M. Vranas (✉)  
University of Wisconsin-Madison, Madison, WI 53706, USA  
e-mail: vranas@wisc.edu

Vihvelin reasons as follows. First, “everyone should agree that if someone would fail to do something, no matter how hard or how many times she tried, then she cannot do it” (1996, p. 318). Second, “because she wouldn’t succeed no matter how often or how hard she tried, I don’t think the time traveler can kill her baby self” (1996, p. 319). So Vihvelin’s argument can be formulated as follows:

- (V1) If someone would always fail to do something if she tried to do it, then she cannot do it.<sup>1</sup>
- (V2) A time traveler would always fail if she tried to kill her baby self.

Thus: (C) A time traveler cannot kill her baby self.<sup>2</sup>

Although Vihvelin does not take her conclusion (that a time traveler cannot kill her baby self) to provide an argument against the possibility of time travel, she says that some of her readers have done so (1996, p. 317; cf. Sider 2002, pp. 121–122). This may be why Vihvelin’s argument has generated some interest: it has been criticized by both Sider (2002) and Kiourti (2008). In what follows I present a critique of Vihvelin’s argument different from both Sider’s and Kiourti’s critiques: I argue in a novel way that both V1 (Sect. 2) and V2 (Sect. 3) are false. If my critique succeeds, then time travel survives a potential challenge unscathed.

## 2 Against V1

V1 is false because in some cases the very act of trying would make one self-conscious and thus would make one fail. For example, suppose that if I tried to win the award for best singer I would become so nervous and I would sing so poorly that I would fail; but suppose further that in fact I sing without trying to win (I don’t even know that I am being considered for the award), and thanks to my ability I sing so well that I do win. Then I *can* win the award (since I *do* win thanks to my ability), although if I tried to win I would always fail. Similarly, suppose that if I tried to type faster than 90 words per minute I would become so self-conscious that I would fail; but suppose further that in fact I regularly type without trying to exceed that speed (I don’t care at all how fast I type), and I do exceed that speed. Then I *can* (since I

<sup>1</sup> Cf. Brown (1992, p. 434); Kenny (1975, p. 129; 1976, p. 229). By contraposition, V1 amounts to the claim that a certain condition is always satisfied when ability and opportunity are present, namely the condition that trying *might* lead to success (on the standard understanding of “might” as “not-would-not”: Lewis 1973, pp. 21–24, 1973/1986a, pp. 8–9, 1986b, pp. 63–64). This condition is weaker than the discredited condition that trying *would* lead to success (cf. Austin 1956/1979, p. 218; Kane 1996, pp. 53–54; Lewis 1976, p. 150; contrast Cross 1986, pp. 58–61).

<sup>2</sup> Assuming that Baby Suzy is the baby who grows up to become Suzy, I take Vihvelin to address primarily the question (1) of whether Suzy can kill Baby Suzy, not the question (2) of whether Suzy can kill the baby who grows up to become Suzy (cf. Smith 1997, p. 379). To see the distinction, assume that Suzy dies in 2050, and compare the question (1\*) of whether Suzy can kill herself before 2050 with the question (2\*) of whether Suzy can kill herself before she dies. If resurrection is a miracle and Suzy cannot do anything that guarantees a miracle, then the correct answers to questions (2) and (2\*) are negative, but I will argue that the correct answer to question (1) can still be positive. [One might argue that it is a negative answer to question (2), not a negative answer to question (1), which amounts to the claim that a time traveler cannot perform “autoinfanticide” and which provides a challenge to the possibility of time travel. I address these issues in another paper (Vranas 2009).]

regularly *do*) type faster than 90 words per minute, although if I tried to do so I would always fail. There are thus counterexamples to V1.<sup>3</sup>

In response Vihvelin might grant that V1 is false but might argue that the singing and typing examples are irrelevant to the time travel case because a time traveler would fail to kill her baby self even if she did *not* try to kill the baby.<sup>4</sup> But even if this response establishes a disanalogy between the time travel case and the singing and typing examples, it misses the point of the examples. The point is not to argue by analogy that a time traveler can kill her baby self; the point is rather to show that Vihvelin's argument is unsound because its first premise is false. And the singing and typing examples do show this. If Vihvelin wants to appeal to the extra premise—call it V3—that a time traveler would always fail to kill her baby self if she did *not* try to kill the baby, then Vihvelin needs to modify her argument, for example as follows:

- (V1') If, no matter what someone did, she would always fail to do something, then she cannot do it.
- (V2') No matter what a time traveler did, she would always fail to kill her baby self. [V2' combines V2 and V3.]

Thus: (C) A time traveler cannot kill her baby self.

It is not clear whether this argument is an improvement on the argument that Vihvelin in fact gives. On the one hand, V1' is weaker than V1 and is immune to the singing and typing examples. On the other hand, V2' is stronger than V2, so to defend V2' Vihvelin would need to do some extra work. Vihvelin might respond that her defense of V2 carries over straightforwardly to V2'. I am not so sure,<sup>5</sup> but in any case I don't think that Vihvelin can successfully defend V2': I argue next that even the weaker premise V2 is false.

<sup>3</sup> Consider also a case proposed by Fara: “although the key in my pocket is the right one, still if I were to use the key right now then the lock would jam, since an insect happens to be crawling through the lock mechanism. This is a case in which I have the ability to open the door, and I have the opportunity to exercise that ability, yet if I were to try to exercise it then I would fail” (2008, p. 847). Another kind of case in which arguably I can do something although I would fail if I tried to do it involves *finkish* abilities: abilities that I would lose if I tried to exercise them (cf. Lewis 1997; Martin 1994). Modifying an example proposed by Clarke (2008, p. 513), imagine a strong man who has the power to move (and who often accidentally moves) heavy objects, but who also has a property possession of which saps his strength when (and only when) he tries to move heavy objects.

<sup>4</sup> Alternatively, Vihvelin might grant that V1 is false but might retreat to a restriction of V1 to cases in which it is not the very act of trying that would make one fail, and might claim that such cases include those in which a time traveler tries to kill her baby self. I reply that the claim that it is not the very act of trying that would make a time traveler fail to kill her baby self seems closely related to the claim that a time traveler would fail to kill her baby self even if she did *not* try to kill the baby, and I address the latter claim in the text.

<sup>5</sup> To defend V2, Vihvelin considers the closest worlds at which Suzy tries to kill Baby Suzy. To defend V2', Vihvelin would also need to consider the closest worlds at which Suzy (does something but) does not try to kill Baby Suzy.

### 3 Against V2

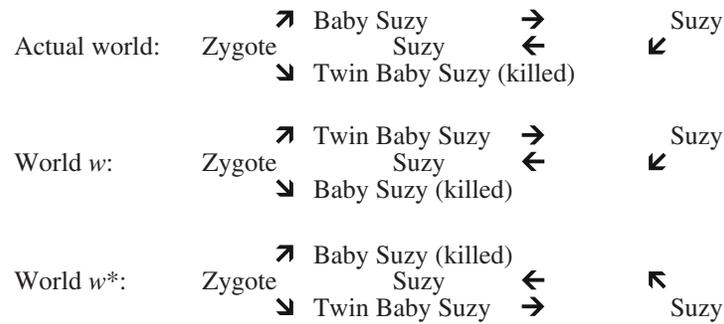
To defend V2, Vihvelin appeals to a standard way of evaluating counterfactuals: “‘If P, it would be the case that Q’ is true just in case Q is true at *all* the closest P-worlds” (1996, p. 319).<sup>6</sup> Using this way of evaluating counterfactuals, Vihvelin argues that, for example, the counterfactual “If Suzy had tried to kill Baby Suzy, she would have failed” is true “at any world remotely like ours” because “*all* the closest worlds at which Suzy tries to kill Baby Suzy are worlds at which she fails” (1996, p. 320). To support the last claim, Vihvelin argues in effect that all worlds at which Suzy tries to kill Baby Suzy and succeeds include either miracles or improbable coincidences. For example, they are worlds at which Baby Suzy is resurrected, or—if Suzy and Baby Suzy are (as from now on I take them to be) person-stages rather than persons—worlds at which Suzy is not a later stage of Baby Suzy and either Suzy “miraculously come[s] into existence out of thin air” (1996, p. 326) or Suzy is a later stage of some other baby-stage with DNA that (by some miracle or improbable coincidence) matches the DNA of Baby Suzy.<sup>7</sup> And, according to Vihvelin, “worlds like these are far less like ours than worlds where the gun jams, or the bullet misses, or ... and that adult stage fails to kill this baby stage” (1996, p. 326).

To see why V2 is false, consider 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.<sup>8</sup> (Later on I address objections to the claim that such a world exists. To ensure that the two worlds are qualitatively identical, assume that the locations which Baby Suzy successively occupies at the actual world are successively occupied at  $w$  by Twin Baby Suzy and vice versa, so that at each world the explosion kills the baby-stage which is located, for example, to the *south* of the bomb; see Fig. 1.) 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

<sup>6</sup> If the “Limit Assumption” that there is at least one closest P-world is false, replace “Q is true at *all* the closest P-worlds” with “some P-world at which Q is true is closer than every P-world at which Q is false” (cf. Lewis 1973/1986a, pp. 9–10).

<sup>7</sup> “[T]he DNA of some actually existing person would miraculously have changed so that it matches the DNA of that adult stage, or the world would have included another baby stage with DNA that matches the DNA of that adult stage and ... the appropriate causal connections would have linked this extra baby stage to that adult stage” (Vihvelin 1996, p. 326).

<sup>8</sup> One might ask: how do we know that at  $w$  Twin Baby Suzy grows up to become Suzy rather than, say, Twin Suzy? I reply that possible worlds “are *stipulated*, not *discovered* by powerful telescopes” (Kripke 1980, p. 44).



**Fig. 1** Three worlds

tries to kill Baby Suzy but fails.<sup>9</sup> (I am not claiming that  $w$  is *the* closest world at which Suzy tries to kill Baby Suzy: maybe  $w$  is just as close to the actual world as some world at which Suzy tries to kill Baby Suzy but fails.) It follows that V2 is false: some time travelers *might* kill their baby selves if they tried.<sup>10</sup> Next I examine four responses that Vihvelin (or other people) might make.

- (1) Vihvelin might reject my implicit assumption that there is a world at which Suzy is not a later stage of Baby Suzy: although for the sake of argument Vihvelin grants this assumption (1996, p. 325), she expresses misgivings about it, saying that it seems to conflict with Kripke’s “necessity of origin” thesis, which seems to imply that “there is no possible world where Suzy grows from anything other than Suzy’s zygote” (1996, p. 324). I reply that at  $w$  Suzy does grow from the zygote from which she grows at the actual world: Suzy, Baby Suzy, and Twin Baby Suzy grow from the same zygote at both worlds. So my assumption is compatible with the above necessity of origin thesis.<sup>11</sup>
- (2) Vihvelin might reject my implicit assumption that numerically distinct worlds can be qualitatively identical, and thus might claim that there is no such world as  $w$ . In reply consider a world  $w^*$  which is like what  $w$  was supposed to be—i.e.,  $w^*$  satisfies conditions (i) and (ii) above—but is *not* qualitatively identical to the actual world: at  $w^*$  the locations which Baby Suzy and Twin Baby Suzy successively occupy are the same as at the actual world, so at  $w^*$  (in contrast to

<sup>9</sup> One might object that this “may well be false under a similarity metric that gives weight to what occurs after the confrontation” between Suzy and Baby Suzy: “Such a metric might require [Suzy] and [Baby Suzy] to be temporal parts of the same time-traveling continuant in nearby worlds” (Sider 2002, p. 131). Sider, however, was presumably assuming that if at some world Suzy and Baby Suzy are not “temporal parts of the same time-traveling continuant”, then there is a qualitative difference between that world and the actual world in “what occurs after the confrontation”; I doubt that Sider would consider a similarity metric to be reasonable if it took qualitatively identical worlds to be dissimilar.

<sup>10</sup> One might object that at  $w$  Suzy does not kill her baby self: although at  $w$  Suzy kills Baby Suzy, at  $w$  Suzy is not a later stage of Baby Suzy. I reply that at  $w$  Suzy does kill her baby self, in the sense of (1) killing the baby who *in fact* (i.e., at the actual world) grows up to become Suzy, though not in the sense of (2) killing the baby who at  $w$  grows up to become Suzy. As I explained in footnote 2, it is (1) rather than (2) which is primarily relevant to Vihvelin’s argument and to my concerns in this paper.

<sup>11</sup> Kripke’s (1980, pp. 110–115) “necessity of origin” thesis, namely that a person who originates from a given sperm and egg could not have originated from a totally different sperm and egg, is about persons rather than person-stages, but from my reasoning in the text it follows that my assumption is compatible even with a variant of Kripke’s thesis which is about person-stages.

- the actual world) the explosion kills the baby-stage which is located to the *north* of the bomb (see Fig. 1). Then  $w^*$  is qualitatively *almost* identical to the actual world and, *mutatis mutandis*, my argument goes through.
- (3) Vihvelin claims: “It is false that if that adult stage had pulled the trigger, then the adult stage and the baby stage would have been stages of two different persons” (1996, p. 326). I reply that I need at most a *might*—not a *would*—counterfactual for V2 to be false.<sup>12</sup> Still, Vihvelin might resist even a “might” counterfactual: she might deny that if Suzy had set off the bomb then Suzy might have been a later stage of Twin Baby Suzy (and not of Baby Suzy). Resistance to accepting this counterfactual may be due to understanding it as “back-tracking”. Such resistance should dissolve once it is noticed that the past of the explosion at  $w^*$  is the *same* as (not just qualitatively identical to) the past of the explosion at the actual world: at both worlds, Suzy, Baby Suzy, and Twin Baby Suzy exist, occupy the same locations, and do the same things up to the explosion. The facts that at  $w^*$  Suzy is a later stage of Twin Baby Suzy and that at the actual world Suzy is a later stage of Baby Suzy are facts about the *future* of the explosion.<sup>13</sup> One can also see that the above counterfactual is not back-tracking by noticing that it is equivalent to the following: if Suzy had set off the bomb, then it is Twin Baby Suzy rather than Baby Suzy who might have grown up to become Suzy.
- (4) Finally, Vihvelin might grant that V2 is false but might retreat to the weaker premise that a time traveler *who has no identical twin* would always fail if she tried to kill her baby self. I have two points in reply. First, this response seems *ad hoc*, intended to close a loophole. How can Vihvelin ensure that other loopholes in her defense of V2 will not be found? Second, my argument against V2 may work, *mutatis mutandis*, against the weaker premise as well: if Suzy at the actual world has no identical twin, then the world  $w^*$  I considered is not qualitatively almost identical to the actual world, but  $w^*$  may be assumed to include no miracle or improbable coincidence and so may still count as at least as close to the actual world as any world at which Suzy tries to kill Baby Suzy but fails.<sup>14</sup>

<sup>12</sup> Some people (e.g., DeRose 1999) contest the standard understanding of “might” as “not-would-not” (see footnote 1), and thus might contest the inference from a “might” counterfactual to the falsity of V2. I reply that I don’t need to take a stand on the proper analysis of “might” counterfactuals because I don’t need the above inference: the falsity of V2 follows (on the standard understanding of “would” counterfactuals; see footnote 6) from my claim that  $w$  is at least as close to the actual world as any world at which Suzy tries to kill Baby Suzy but fails.

<sup>13</sup> I am not saying that future dissimilarities between  $w^*$  and the actual world are irrelevant to the evaluation of counterfactuals; I am rather saying that future dissimilarities between the two worlds are irrelevant to my claim that the past of the explosion is the same at the two worlds.

<sup>14</sup> In response Vihvelin might retreat to the even weaker premise that a time traveler who *cannot* have an identical twin (for example, a non-human time traveler whose reproductive system precludes genetic duplication) would always fail if she tried to kill her baby self. I reply that the impossibility of having an identical twin would presumably be biological, and would be compatible with metaphysical possibility (which is what I need).

## 4 Conclusion

My task in this paper might appear purely negative: I argued that Vihvelin's argument is unsound because both of its premises are false. But on an alternative reading this paper has the positive task of defending the possibility of time travel against a powerful challenge. On yet another reading, this paper has the positive task of enabling one to defend a particular solution to the "autoinfanticide paradox", namely a paradox which arises from the existence of plausible considerations for the conflicting claims that a time traveler can and cannot kill her baby self. Some people try to resolve this paradox by arguing that, contrary to appearances, a time traveler cannot kill her baby self. By rebutting Vihvelin's argument, this paper paves the way for resolving the paradox by arguing instead that a time traveler *can* kill her baby self after all.

**Acknowledgments** I am grateful to Martin Barrett, Joshua DiPaolo, Molly Gardner, Alan Hájek, Aviv Hoffmann, Ira Kiourti, Emily Pritzkow, Kelly Robbins, Jordan Rogers, Carolina Sartorio, and three anonymous reviewers for helpful comments, and to my mother for typing the bulk of the paper.

## References

- Austin, J. L. (1979). Ifs and cans. In J. L. Austin, *Philosophical papers* (3rd ed., pp. 205–232). New York: Oxford University Press (originally published 1956).
- Brown, B. (1992). Defending backwards causation. *Canadian Journal of Philosophy*, 22, 429–443.
- Clarke, R. (2008). Intrinsic finks. *The Philosophical Quarterly*, 58, 512–518.
- Cross, C. B. (1986). 'Can' and the logic of ability. *Philosophical Studies*, 50, 53–64.
- DeRose, K. (1999). Can it be that it would have been even though it might not have been? *Philosophical Perspectives*, 13, 385–413.
- Fara, M. (2008). Masked abilities and compatibilism. *Mind*, 117, 843–865.
- Kane, R. (1996). *The significance of free will*. New York: Oxford University Press.
- Kenny, A. (1975). *Will, freedom and power*. Oxford: Blackwell.
- Kenny, A. (1976). Human abilities and dynamic modalities. In J. Manninen & R. Tuomela (Eds.), *Essays on explanation and understanding* (pp. 209–232). Dordrecht: Reidel.
- Kiourti, I. (2008). Killing Baby Suzy. *Philosophical Studies*, 139, 343–352.
- Kripke, S. A. (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.
- Lewis, D. K. (1973). *Counterfactuals*. Oxford: Blackwell.
- Lewis, D. K. (1976). The paradoxes of time travel. *American Philosophical Quarterly*, 13, 145–152.
- Lewis, D. K. (1986a). Counterfactuals and comparative possibility. In D. K. Lewis (Ed.), *Philosophical papers* (Vol. 2, pp. 3–31). New York: Oxford University Press (originally published 1973).
- Lewis, D. K. (1986b). Postscripts to "Counterfactual dependence and time's arrow". In D. K. Lewis (Ed.), *Philosophical papers* (Vol. 2, pp. 53–66). New York: Oxford University Press.
- Lewis, D. K. (1997). Finkish dispositions. *The Philosophical Quarterly*, 47, 143–158.
- Martin, C. B. (1994). Dispositions and conditionals. *The Philosophical Quarterly*, 44, 1–8.
- Sider, T. (2002). Time travel, coincidences and counterfactuals. *Philosophical Studies*, 110, 115–138.
- Smith, N. J. J. (1997). Bananas enough for time travel? *British Journal for the Philosophy of Science*, 48, 363–389.
- Vihvelin, K. (1996). What time travelers cannot do. *Philosophical Studies*, 81, 315–330.
- Vranas, P. B. M. (2009). Time travel: Two kinds of consistency paradoxes (in preparation).