

# Asymmetric personal identity\*

THEODORE SIDER

*Journal of the American Philosophical Association* 2018: 127–46

---

## 1.

The eponymous antihero of James Thurber’s story ‘The Secret Life of Walter Mitty’ escapes his dreary suburban life by daydreaming. Thurber’s Mitty dreams of many things, of being a navy pilot, a famous surgeon, a notorious assassin, and so forth; but imagine a different Mitty, whose daydreams always involve a single persona, a cape-wearing crime-fighting vigilante. And further, imagine that each night, these daydreams become reality. Each night, Mitty gets out of bed and really does fight crime. Or better, to avoid begging questions let us say that ‘Daytime Mitty’ leads a boring suburban life and that ‘Nighttime Mitty’ fights crime. Nighttime Mitty, to continue the story, can remember being a frustrated suburbanite as Daytime Mitty, and can also remember fighting crime as Nighttime Mitty on earlier nights (‘quasi-remember’ if you prefer [Shoemaker 1970]). But Daytime Mitty cannot remember any of the nighttime events and has no other source of knowledge of them. He has no idea that Nighttime Mitty exists.

I offer two judgments. First, Nighttime Mitty *was* Daytime Mitty. Nighttime Mitty can say truly, ‘I lived a boring suburban life earlier today’. Second, Daytime Mitty will *not* be Nighttime Mitty. If Daytime Mitty were to say ‘I will fight crime this evening’, that would be false. These two judgments together yield the conclusion that personal identity for Mitty is asymmetric.<sup>1</sup> Asymmetric personal identity might seem logically absurd. But before addressing that issue, consider the case in its favor.

---

\*Thanks to Don Baxter, Karen Bennett, Tad Brennan, Phillip Bricker, Eddy Chen, Andrew Chignell, Andy Egan, Adam Elga, Hilary Greaves, Liz Harman, Jenann Ismael, Mark Johnston, Tom Kelly, David Kovacs, Peter Lewis, Nan Li, Daniel Manne, Kate Manne, Luke Manning, Andrew McGonigal, Jill North, Daniel Rubio, Jonathan Schaffer, Josh Schechter, Erin Taylor, Briana Toole, David Velleman, and referees.

<sup>1</sup>Consider also the Backward Lookers, who live their lives exclusively in the past and care nothing for the future. They remember their pasts clearly, delighting in or lamenting not only experiential memory but also the narrative coherence of their pasts. But they are indifferent to the arc of their future and their future well-being, and form no intentions whatsoever. The Backward Lookers clearly have pasts, but one might judge them to have no futures at all, that none of them will persist into the future.

Few will balk at the first judgment. Nighttime Mitty will surely regard Daytime Mitty's doings as his own, will feel responsible for Daytime Mitty's misdeeds, will regret those misdeeds, and so forth; and these opinions of his from the inside match our own from the outside. The second judgment will meet more resistance, but careful thought, I believe, supports it as well.

The presence of certain evaluative, rational, and moral relations are often used to 'test' for personal identity, particularly by advocates of the psychological approach. (Although a psychological criterion of personal identity will not appear as an explicit premise in the argument I am about to give, the argument is directed in the first instance at proponents of the psychological approach; I doubt it will move fans of a purely bodily approach.) One person's bearing these relations to another—one person's 'mattering' to another, in Derek Parfit's (1971) terminology—is thought to indicate that the persons are identical. Conversely, *not* mattering indicates nonidentity.

Distinguish forward-looking from backward-looking tests. Backward-looking tests apply from the perspective of the present looking back; forward-looking tests apply from the perspective of the present looking forward. The key to the case for asymmetric personal identity is that forward- and backward-looking tests can come apart.

One test says that regret for past actions is appropriate only if one committed those actions oneself. (The relevant sort of regret is 'agent-regret' [Williams 1981]. The principle would be trivialized if agent-regret were simply *defined* as regret for one's own past. We must, rather, pick out agent-regret 'ostensively', as a distinctive and familiar sort of regret, a kind we typically, though not definitionally, regard as restricted to our own pasts.) This backward-looking test delivers a clear verdict in favor of the first judgment: Nighttime Mitty can regret Daytime Mitty's misdeeds, so Nighttime Mitty was Daytime Mitty.

Another test says that if I will be a certain future person, then I have reason to care about what happens to him, and that if good or bad things will happen to him—good or bad things given *my* desires and values—then it is *now* good or bad for me that those things will happen—it is now good *news* or bad news for me that those things will happen. This forward-looking test, applied from Daytime Mitty's perspective, speaks against identifying him with Nighttime Mitty. Nighttime Mitty does what Daytime Mitty most wants to do. So if Daytime Mitty will be Nighttime Mitty, the fact that Nighttime Mitty will fight crime is now good for Daytime Mitty. But surely Nighttime Mitty's future exploits are no good at all for Daytime Mitty. Thus the test tells us that Daytime Mitty will not be Nighttime Mitty. Similarly, Daytime Mitty apparently has no

reason to care about Nighttime Mitty's well-being (not from a self-interested point of view anyway, which is what is relevant to the test); thus the test tells us again that Daytime Mitty will not be Nighttime Mitty.

Daytime Mitty's ignorance of what happens at night may make it hard to assess whether Nighttime Mitty's doings are good now for Daytime Mitty. But imagine being Daytime Mitty, and asking yourself the hypothetical question of whether it would be good for you *if* the story described above were true. I think you would answer that it would not. Imagine further that you were then told that the story *is* true, that for years Nighttime Mitty has been waking up and fighting crime, that this will continue, and that you never have had and never will have any memory of this during the daytime. I do not think you would regard this as good news, or that you would begin to identify with Nighttime Mitty. You would regard him as a lucky person living out your dream, and—the ultimate irony—doing it with your body. You would not 'own' the crime-fighting any more than you would if you did it while sleepwalking, or if some 'puppeteer' did it using your sleeping body.

(This is not to say that you would not be the sleepwalker or puppet. You surely would be, if only because of the sameness of body. This in turn is not to concede that sameness of body is generally sufficient for personal identity, so that Daytime Mitty will be Nighttime Mitty after all. Bodily sameness, I think, determines personal identity in the case of sleep only as a sort of extension by courtesy of an otherwise psychological account to stages of our lives that are not fully or actively psychological.)

It might be thought that Daytime Mitty is indifferent to Nighttime Mitty only because *good* things happen to Nighttime Mitty, and that Daytime Mitty would fear the future if he knew that Nighttime Mitty was to be tortured, say. Even if true this would not threaten the argument since the original version of the example, based on crime-fighting rather than torture, would still establish that personal identity can be asymmetric. But I think Daytime Mitty would not fear the torture, not after awhile anyway. He might be a little fearful after first learning of Nighttime Mitty's existence. But after falling asleep with apprehension, the next day he would wake up relieved, with no memory of any torture. As the days passed with no torture ever remembered, the fear would diminish and eventually disappear. It is sometimes said that certain kinds of anesthesia work by eliminating memory of pain rather than the pain itself. This is unnerving to be sure (there is a lot to be uncertain about) but not terrifying. We continue to pay for anesthesia.

It is not asymmetric personal identity if a drunkard passes out each night

and cannot remember what he did. But there are many differences between this case and Mitty's. The drunkard's amnesia is imperfect: there remain confusing, disconnected, embarrassing images of the drunken state. The transition to the drunken state is continuous and (to some extent) under the drunkard's control. The drunkenness has many repercussions in the drunkard's daytime life. And the asymmetries of anticipation, intention, and narrative to be discussed in the next section are less severe. Perhaps personal identity is asymmetric in some extreme version of the case of the drunkard, but our conviction that drunken blackouts do not produce asymmetric personal identity is based on more ordinary cases.

Consider an alternate version of Mitty's story, in which Mitty will die at dawn immediately after his first and only nighttime escapade. It is then clearly true, on the evening before the escapade, that Mitty will fight crime that night—this is just an ordinary case of a person who is about to unexpectedly wake at night. (The amnesia that figures centrally in my version of the story begins only the morning after the first escapade.) Thus in this truncated story, Daytime Mitty will be Nighttime Mitty. But then, should not the same be true in my version of the story, even after many escapades? (Thanks to Liz Harman for this argument.)

The objection assumes that whether Daytime Mitty will be Nighttime Mitty depends only on their intrinsic connection. But in my view, certain extrinsic factors are relevant, such as the fact that there have been many iterations of crime-fighting and failing to remember. (And perhaps that there will be many more—is it only later in the example when Daytime Mitty will fail to be Nighttime Mitty, or is personal identity asymmetric right from the start?) These extrinsic factors differ in the truncated story, which is why we judge it differently.

## 2.

A case has been made for asymmetric personal identity; but what is its source? What is it about the grounds of personal identity that makes asymmetry possible?

The relations that ground personal identity are many in number and are not symmetric. Further, the contribution from these relations to personal identity—to one person's 'identifying' with another, to counting the other's perspective as lying in her future or past—are often on just one side of the relation, so to speak.

For example, when a later person remembers the doings of an earlier person, the case for identification is strong from the later person's perspective, but is entirely lacking in force from the earlier person's perspective. If Daytime Mitty is told that Nighttime Mitty remembers his (Daytime Mitty's) deeds, this does not tend in the slightest to make him identify with Nighttime Mitty, to regard Nighttime Mitty's actions as his own. Let me put this by saying that memory is an 'identification relation' that 'identifies' only from the later-self's point of view. Or, more concisely, it is an 'identification relation for later selves'. Ordinarily there is a case for identification both from the perspective of the earlier self and from the perspective of the later self, since ordinarily some identification relations identify from the earlier self's perspective and some identify from the later self's perspective. But in extraordinary circumstances there can be a case from only one of these perspectives. Although some identification relations identify from Nighttime Mitty's perspective, none identifies from Daytime Mitty's perspective.

It is clear which identification relation for later selves Nighttime Mitty bears to Daytime Mitty: memory. But which identification relations for *earlier* selves does Daytime Mitty *fail* to bear to Nighttime Mitty?

One is anticipation. J. David Velleman (1996) stresses the importance of anticipation to survival into the future: 'What we most want to know about our survival, I believe, is how much of the future we are in a position to anticipate experiencing. We peer up the stream of consciousness, so to speak, and wonder how far up there is still a stream to see' (Velleman 1996, 194–95). Anticipation is not the mere fact that an experience will happen to one; it is an active mental state in which one expects the future experience, and, moreover, expects the future experience to be colored by one's expectation of it. As Velleman puts it, 'Within the frame of my anticipatory image, I glimpse a state of mind that will include a memory of its having been glimpsed through this frame—as if the image were a window through which to climb into the prefigured experience' (Velleman 1996, 198).

When it holds, anticipation counts in favor of identity from the earlier self's perspective. But anticipation seems not to hold in Mitty's case: Daytime Mitty cannot anticipate Nighttime Mitty's experiences. The fact that Daytime Mitty knows nothing of Nighttime Mitty again makes this hard to judge, but vary the story again so that Daytime Mitty learns of Nighttime Mitty's existence. When I put myself in Daytime Mitty's shoes and imagine learning all the facts about the situation, I just cannot reach out in anticipation to Nighttime Mitty.

Assuming this is so, why is it so? It is a bit puzzling. After all, once Daytime

Mitty knows the score, he can form appropriate *beliefs* about what is about to happen. And recall Velleman's point that one expects an anticipated future experience to contain memories of the anticipation: Daytime Mitty knows that Nighttime Mitty remembers Daytime Mitty, and thus knows that his attempts at anticipation will indeed be remembered. What blocks the anticipation, somehow, would seem to be Daytime Mitty's knowledge that he will wake up the following morning remembering nothing of the night, and second, that this happens repeatedly. But why exactly do these facts block anticipation? I am not sure.

Another earlier-self identification relation is intention. (Its importance to survival is also stressed by Velleman [1996].) In ordinary cases, we form many specific intentions for the future, which we generally expect to be, and often are, fulfilled. This relation seems identifying from the point of view of the earlier self, but Daytime Mitty does not bear it to Nighttime Mitty. If Daytime Mitty does not know of Nighttime Mitty's existence, for this reason alone he forms no specific intentions at all concerning the night. If Daytime Mitty learns of Nighttime Mitty's existence, he might at first try to form intentions. But on subsequent days, after consistently failing to have any memory of the night, it will feel increasingly odd to continue trying. It would be like sending instructions out into the dark, like trying to control what one will do in a dream by repeatedly imagining the desired action. Any subsequent attempts would surely not result in genuine intentions. (This seems true even if the attempts succeed, and are known to succeed, in influencing Nighttime Mitty; but if this is disputed we can stipulate that there is no such influence and that Daytime Mitty knows this.)

Yet another relevant identification relation is narrative. (See, for example, Schechtman [1996], although she denies that narrative is criterial of personal identity in the sense of numerical identity.) Narrative, we can say, contributes to one's identity with an earlier or later self *S* to the extent that *S*'s doings fit, or cohere with, one's narrative, the story of one's life. Daytime Mitty fails to identify with Nighttime Mitty—even after being told of the nighttime exploits—in part because Nighttime Mitty's doings do not fit Daytime Mitty's narrative: a story of unfulfilled dreams and suburban existence.

But is that Daytime Mitty's *only* narrative? What of a 'combined' narrative that includes the nighttime exploits in addition to the suburban dream? It certainly counts as *Nighttime Mitty's* narrative; that is how he would tell the story of his life.

The combined narrative is not Daytime Mitty's because it includes a large,

natural, integrated part—the sum of its nighttime segments—from which Daytime Mitty is cut off: he cannot remember any of it. Thus even though memory is an identification relation for *later* selves, and hence plays no direct role in determining whether the earlier self Daytime Mitty identifies with the later self Nighttime Mitty, it nevertheless plays a role indirectly. The failure of Daytime Mitty to remember the *earlier* nighttime segments of the combined narrative is what disqualifies that narrative as belonging to him. It is of course possible to forget some parts of one's narrative. But Daytime Mitty cannot remember *any* of this large, natural, integrated part.

Narrative, then, is a nonsymmetric identification relation. The relation born by self *S* to self *O* just when *O's deeds fit some narrative that belongs to S* is nonsymmetric since a narrative belonging to *S* need not belong to *O*. And if the relation holds in just one direction, it is surely identifying only from the point of view of *S*, the owner of the narrative, not from *O's*. Nighttime Mitty bears the relation to Daytime Mitty since some narrative belonging to Nighttime Mitty—the combined narrative—includes the doings of Daytime Mitty. This contributes to Nighttime Mitty's identifying with Daytime Mitty, but not yet to Daytime Mitty's identifying with Nighttime Mitty since the relation is identifying only from the perspective of the owner of the narrative. Moreover, Daytime Mitty does not bear this relation to Nighttime Mitty, since no narrative belonging to Daytime Mitty includes Nighttime Mitty's actions.

The identification relations for earlier selves that I have considered, intention, anticipation, and narrative, fail from Daytime Mitty's perspective to identify him with Nighttime Mitty. But my strategy for grounding asymmetric personal identity relies on there not being any other such relations that would make this identification. Are there any such relations? Sameness of character is sometimes claimed to be criterial of personal identity; and if it is, it surely identifies from the perspective of earlier selves (as well as, presumably, from the perspective of later selves). (The same issue would arise if sameness of body were criterial of personal identity.) But sameness of character strikes me as not criterial at all of personal identity. Certainly it counts for nothing in isolation from other factors: there is no case whatsoever for identifying me with a person on a distant planet who happens to share my character traits. Another criterion sometimes advanced is the possession of beliefs caused by states of an earlier self. (Thanks to Mark Johnston here.) But this relation, it seems to me, at best counts in favor of identification from the perspective of the later self.

I have not portrayed the discussion above as an argument for asymmetric personal identity, but rather as an account of how asymmetric personal identity

is possible (that possibility having been established in the previous section). But someone who was antecedently convinced that psychological factors like memory, anticipation, intention, and narrative jointly ground personal identity might well take this discussion as a further argument in favor of asymmetric personal identity—perhaps even a more compelling one. For if those factors are indeed the joint grounds of personal identity, it is natural to take each as counting in favor of personal identity from one direction only, which then leads to asymmetric personal identity when the factors from one direction are missing.

### 3.

Personal identity is asymmetric only in extreme cases like *Mitty's*, where there is a severe discrepancy between different identification relations. But a milder discrepancy is present in mundane cases, and this is significant as well.

We ordinarily believe not only that personal identity is symmetric, but also that the evaluative relations born by our past to future selves are parallel in a certain sense to the relations born by our future to past selves. We assume, for example, that if the earlier self is rationally obliged to care about the later self's well-being then the later self must also, and for the same reasons, and to the same extent, be morally responsible for any wrongdoing of the earlier self. These assumptions are undermined by what we have learned about identification relations, even in cases that are not so extreme as to amount to asymmetric personal identity. For even if the case for identification is sufficiently strong both from the point of view of the earlier self and from the point of view of the later self, the operative identification relations are different in the two cases, and so the nature and strength of the evaluative connection might differ. It is an open and interesting question just how equal in nature and strength these connections normally are.

To be concrete, consider narrative. The narratives we embrace late in life are typically overarching, including even our childhoods. But to an adolescent, the narrative she will eventually embrace is just an abstraction. It is one of many narratives that *might* become hers, but she does not yet own it. And any narrative she does own need not be owned by her later self. So as far as narrative is concerned, then, the backward-looking connection is stronger than the forward-looking one. Now, it may be that the various identification relations for earlier selves compensate, resulting in connections of comparable



strength. But in any case there is the possibility of an asymmetry in strength or nature of connection. And this asymmetry might matter. How irrational is it to smoke, knowing that one's future self will be put at risk? How morally responsible are we for the misdeeds of our past selves? These questions might be answered quite differently.

#### 4.

Asymmetric personal identity would be absurd if it conflicted with the standard logic of the identity relation. The fact that  $x = y$  if  $y = x$  is as strongly confirmed as can be in other contexts, contexts that are better understood than the theory of personal identity. What we need is a way to conceptualize the phenomenon that does not require denying that the identity relation itself is symmetric.

There are other cases in which personal identity has been argued not to obey the standard logic of identity. Most notoriously there is 'fission', in which an earlier person is 'split' into two later persons, perhaps by dividing the brain of the original person and transplanting the halves into two cloned, brainless bodies, as in the thought experiment of David Wiggins (1967, 52). Logically, the puzzle is that although the later persons  $L_1$  and  $L_2$  are distinct, the earlier person  $E$  seems to be the same person as  $L_1$  and also as  $L_2$  (for surely  $E$  would have been one of the later persons if the other had never existed, and surely the mere presence of one later person cannot destroy  $E$ 's ability to survive as the other); and if so, personal identity cannot be both transitive and symmetric.

Derek Parfit's (1971; 1984, chap. 12) response to the puzzle of fission was twofold. First,  $E$  goes out of existence upon division, which solves the puzzle's logical aspect. Second, personal identity does not have the rational and moral significance we ordinarily take it to have. We ordinarily assume that ceasing to exist is very bad, that one can be responsible only for what one does oneself, and so forth (recall the tests). But even though dividing causes  $E$  to stop existing, this is not bad for her, according to Parfit. For *each* of the later persons,  $L_1$  and  $L_2$ , the existence of that person preserves what is important to  $E$ , even though  $E$  is not identical with either. Identity, Parfit says, is not 'what matters in survival'.

But according to David Lewis (1983*b*), personal identity need not be divorced from what matters if one accepts an appropriate metaphysics of personal identity. Assuming fission is not bad in the way that death is, fission must not

result in any person going out of existence, if personal identity and what matters are to coincide. And this, Lewis says, can be achieved by saying that there was no single person  $E$  before the fission. Rather, each of the later persons,  $L_1$  and  $L_2$ , was ‘there all along’. The pre-fission relation between  $L_1$  and  $L_2$  is like the relation between a statue and the quantity of matter from which it is made: although numerically distinct,  $L_1$  and  $L_2$  are then intrinsically alike, have the same mass, spatial location, material parts, and so forth. They differ merely in their future-looking properties: they will later go on to do different things. For Lewis, this pre-fission ‘coincidence’ between  $L_1$  and  $L_2$  is possible because they, like all persisting things, are aggregates of temporal stages; the only ‘wholly present’ entity during the time of coincidence is the segment of person-stages that the aggregates  $L_1$  and  $L_2$  share.

Whatever the merits of this approach to fission (I argue against it in Sider [1996]), it runs into trouble in the case of asymmetric identity. The facts of the case, as argued above, are these: Nighttime Mitty was Daytime Mitty, but Daytime Mitty will not be Nighttime Mitty. Or better: an utterance by Nighttime Mitty of ‘I lived a boring suburban life earlier today’ would be true, but an utterance by Daytime Mitty of ‘I will fight crime later tonight’ would be false. Lewis cannot accommodate these facts.

To see this, we must examine the semantics that Lewis pairs with his metaphysics of persons. Return to the case of fission. Suppose that at some time before division,  $L_1$  and  $L_2$  utter the first-person pronoun ‘I’ (via their shared stage); to what does ‘I’ refer? ‘I’ is normally taken to refer to *the* person uttering it, but here there is no unique person doing the uttering. Lewis’s answer is that pre-fission uses of ‘I’ are indeterminate in reference between  $L_1$  and  $L_2$ , and that a supervaluational semantics governs such indeterminacy. Thus a pre-fission utterance of ‘I am  $F$ ’ is true if each of the coinciding persons is  $F$  (this is called supertruth), false if neither is  $F$  (superfalsity), and neither true nor false if one person is  $F$  and the other is not.

So since Nighttime Mitty’s utterance of ‘I lived a boring suburban life earlier today’ is true, each person containing the uttering Nighttime Mitty stage also lives a boring suburban life earlier in the day, and hence contains Daytime Mitty stages that are located earlier in the day. But each person stage is part of at least one person. (This is a condition of adequacy on any temporal parts account of persons—assuming the stage is embedded in an appropriate sequence of stages, anyway. In Lewis’s own theory it follows from his definition of persons as maximal R-interrelated sums.) So there is at least one person,  $P$ , containing both stages of Nighttime Mitty from that night and also stages of Daytime

Mitty from that day. But that means that if Daytime Mitty uttered ‘I will fight crime later tonight’, it would not be (super>false. For there is at least one person—namely, *P*—containing the uttering stage who *does* fight crime that night.

At best, Lewis could claim that such an utterance would be neither true nor false. He could do so by claiming that in addition to *P*, there exists another person, call him ‘Boring-Mitty’, made up of all and only the boring, daytime stages. Daytime Mitty’s utterances of ‘I would then be indeterminate between *P* and Boring-Mitty; since one of these persons fights crime and the other does not, ‘I will fight crime later tonight’ is neither true nor false. But this verdict is not adequate to the example. It is surely false, and not merely untrue, that Daytime Mitty can look forward to fighting crime. In the case of *fission*, the ‘indeterminate’ future that Lewis’s account delivers is not unintuitive. Suppose that  $L_1$  wakes up after fission in a red recovery room and  $L_2$  wakes up in a blue recovery room. Lewis’s view then implies that a pre-fission utterance of ‘I will wake up in a red recovery room’ is neither true nor false, which isn’t so hard to swallow. (Although see below.) But Mitty’s case surely involves no such indeterminacy. It is determinate that Daytime Mitty will never fight crime.

(This attempt to accommodate asymmetric identity in Lewisian terms also faces another challenge. Since Nighttime Mitty’s utterance of ‘I lived a boring suburban life earlier today’ is true, ‘Exciting-Mitty’, the aggregate of all and only the nighttime stages, cannot count as a person. But there is a certain symmetry between Boring-Mitty and Exciting-Mitty. For instance, if Exciting-Mitty is disqualified from personhood because it is a proper part of the person *P*—say, by appeal to something like the maximality clause in Lewis’s definition of personhood—that would also disqualify Boring-Mitty.)

Lewis might instead try to account for the phenomenon by positing contextual shifts in the extension of the predicate ‘person’, and corresponding shifts in the referents of names and pronouns. (Thanks to Jonathan Schaffer here.) For instance, he might hold that in some contexts, Boring Mitty (the aggregate of the daytime stages) counts as a person, and in other contexts ‘Full Mitty’, the aggregate of all daytime and nighttime stages, counts as a person, but in no context do they both count as persons. And he might claim that in the context in which Daytime Mitty says ‘I will fight crime later tonight’, only Boring-Mitty counts as a person, ‘I’ refers to Boring Mitty, and the utterance is false; whereas in the context in which Nighttime Mitty says ‘I lived a boring suburban life earlier today’, only Full Mitty is a person, ‘I’ refers to Full Mitty, and the utterance is true. But this view predicts that in the former context, the

sentence ‘It is sometimes the case that a person remembers (in the right kind of way) some past person, but nevertheless was not that past person’ is true; whereas, I say, this sentence is false in *all* contexts, on the grounds that memory is criterially sufficient for personal identity.

Lewis, then, cannot accommodate asymmetric personal identity. But Lewis’s metaphysics of persons is not the only one that can align identity with what matters. There is also the ‘stage view’, or ‘temporal counterpart theory’, defended by Katherine Hawley (2001) and me (1996; 2001). And as I show below, temporal counterpart theory can accommodate asymmetric personal identity. So if we are going to wheel in a metaphysics (and associated semantics) of persons to resolve the mismatch between the strict logic of identity and the multifaceted logic of our identifying attitudes, including the mismatch in the case of asymmetric personal identity, the right one to wheel in is temporal counterpart theory, and not Lewis’s theory of overlapping aggregates of stages.

(Counterpart theory also seems to outperform Lewis’s theory in the case of an extremely long-lasting person who changes continuously. What Lewis says about this case is roughly that each aggregate of person-stages of a certain fixed temporal length counts as a person [Lewis 1983*b*, section 4]. Thus at most moments in such a life, there are infinitely many overlapping persons, extending forward and backward in time to varying degrees. But some of these people are located almost entirely to the past of the moment in question, and cease to exist only a millisecond later, which means that there are precisifications on which an utterance at that time of ‘I will exist for at least two more milliseconds’ is false. Lewis might say that the sentence is nevertheless true to a very high degree, since it is true on almost all precisifications [Lewis 1970, 64–65]. But surely it is completely true; and the counterpart theorist can agree.)

According to temporal counterpart theory, persons are person stages, not aggregates of person stages. (See Sider [2006] for my currently preferred view on how the temporal extent of the stages that are persons is determined.) Ordinary uses of personal names and personal pronouns refer to person stages at the time of utterance. Tensed sentences containing such terms can nevertheless be true, for they are governed by a temporal version of David Lewis’s (1968) counterpart-theoretic semantics for modal operators. A current utterance by me of the past-tensed sentence ‘I once was four feet tall’ may be regimented with a Priorian tense operator **P** for ‘it was the case in the past that’: ‘**P**(I am four feet tall)’. According to temporal counterpart theory, **PA**(*x*) is true if and only if *A*(*x*) is true of some past temporal counterpart of *x*. Similarly for the future tense operator: **FA**(*x*) is true if and only if *A*(*x*) is true of some future

temporal counterpart of  $x$ . (Compare Lewis's claim that  $\diamond A(x)$  is true if and only if  $A(x)$  is true of some (modal) counterpart of  $x$  in some possible world.) Thus even though I am only a person-stage, 'P(I am four feet tall)' is true since some of my past temporal counterparts are four feet tall.

Like Lewis's view, temporal counterpart theory implies that no person goes out of existence in the case of fission; thus it too aligns identity with what matters. The counterpart-theoretic truth condition of 'I will exist after fission' is that some counterpart of the utterer be located after fission, and the pre-fission subject—a person-stage, according to counterpart theory—does have a counterpart after fission. Indeed, she has counterparts 'on both branches'.

Temporal counterpart theory does generate logically odd results in this case. Let  $F_1$  be the tense operator 'it will be the case one day hence that'. Its counterpart-theoretic semantics is this:  $F_1A(x)$  is true if and only if  $A(x)$  is true of some temporal counterpart of  $x$  that is located one day after the time of utterance. But then if we symbolize 'x wakes in a red room' as  $Rx$ , the sentence  $F_1Rx \wedge F_1\sim Rx$  comes out true of the pre-fission person, since she has two counterparts one day hence, one who wakes in a red room and one who does not. Thus she can say truly, 'In one day it will be the case that I wake in a red room, and in one day it will be the case that I do not wake in a red room'. But she cannot say truly 'In one day it will be the case that: I wake in a red room and do not wake in a red room';  $F_1(Rx \wedge \sim Rx)$  is false of her, since she has no counterparts one day hence who both do and do not wake in a red room. Thus  $F_1A \wedge F_1B$  fails to imply  $F_1(A \wedge B)$ . Some will reject the theory on the grounds that it generates an unacceptable tense logic, but the counterpart theorist will insist that this is the logic demanded by the metaphysically odd case of fission, and will point out that classical logic in the extensional, tenseless metalanguage is not threatened.

Asymmetric personal identity is straightforward for the temporal counterpart theorist, since the temporal counterpart relation needn't be symmetric. The temporal counterpart theorist is free to say that although Daytime Mitty stages are counterparts of future Nighttime Mitty stages, Nighttime Mitty stages are not counterparts of past Daytime Mitty stages. The idea would be that whether an earlier thing is a counterpart of a later thing depends on the holding of identification relations for later selves (such as memory), whereas whether a later thing is a counterpart of an earlier thing depends on the holding of identification relations for earlier selves (such as anticipation, intention, and narrative). Nighttime Mitty can therefore truly utter 'I lived a boring suburban life earlier today', since he has a counterpart earlier in the day who lived a

boring suburban life; but an utterance by Daytime Mitty of ‘I will fight crime later tonight’ would be false since he has no future counterparts that fight crime.

Note the importance of tense—here understood in Priorean terms—for describing the phenomenon of asymmetric personal identity. What we want to say is that from the perspective of Nighttime Mitty looking back at Daytime Mitty, identity holds, and that from the perspective of Daytime Mitty looking ahead, it does not. These perspectives can be cashed out as times of utterance of tensed sentences.

As with fission, this satisfying account comes at the price of an odd tense logic. Since Nighttime Mitty fights crimes but has past counterparts—Daytime Mitty stages—with no future counterparts that fight crime, he can truly utter ‘Although I am fighting crime, it was the case that it would never be the case that I fight crime’:  $Cx \wedge P\sim FCx$ .

## 5.

Velleman argues that there is a sort of asymmetric personal identity in the case of fission (Velleman 1996, 200–2, especially note 53). He himself regards the ‘personal identity’ involved as being, not the persistence of a numerically identical self over time, but rather the holding of a certain relation of ‘being a self for’; he is a Parfittian about the issues of the previous section. Nevertheless, with metaphysical conceptions of persons like Lewis’s and temporal counterpart theory on the table, we may consider the argument as concerning persistence.

According to Velleman, the distinctive relation one bears to one’s past and future selves is a certain mode of reflexive thought: one can think about their experiences—via memory, for past selves, and via anticipation, for future selves—in the first person. But not just any mode of reflexive thought will do. For in imagination one can think first-personally about someone else’s experiences: one can imagine a certain perspectival experience that Napoleon in fact had, and ‘center that image on him’ by stipulation, by stipulating to oneself that it is Napoleon one is imagining being (Velleman 1996, 188). The difference is that in genuine first-personal thought, no stipulation is needed; the centering on oneself is ‘automatic’.

Velleman arrives at this view through a general analysis of imagination, memory, and anticipation, but he then applies it to the case of fission. Each of the two persons resulting from fission can access via memory the experiences

of the original person in a genuinely first-person way. But the original person cannot access the thoughts of either of those two persons via genuine first-person anticipation, Velleman claims. Since there are two of them, any first-personal access must be via stipulation, and thus amounts to mere imagining, rather than genuine anticipation. The pre-fission subject cannot think to herself, ‘what will I be doing tomorrow?’

If Velleman is right that fission blocks anticipation, we have the beginnings of an independent case for asymmetric personal identity. But it is not clear that he is right. Anticipation of post-fission experiences seems especially possible if fission regularly occurs. Imagine that subjects frequently enter a certain fission chamber, go to sleep, and then are divided in two, one waking in a red recovery room and the other in a blue recovery room. Perhaps a fission rookie would find it hard to anticipate experiencing anything at all afterward, but consider a veteran. She remembers entering the fission chamber many times and waking in a recovery room each time. Moreover, if she is a typical veteran, she remembers waking in a red room roughly half the time and in a blue room roughly half the time. She also knows that other veterans on average remember waking in red rooms half the time and blue rooms half the time. Putting myself in the veteran’s shoes, when entering the fission chamber the next time, I feel fairly sure that my attitudes would be these: (i) I would anticipate waking in a recovery room, and (ii) I would be uncertain as to the color of the room—in particular, I would regard the color as being 50 percent likely to be red and 50 percent likely to be blue.

## 6.

If (i) is right, then fission does not generally block anticipation, which casts doubt on the independent case for asymmetric personal identity. But if (ii) is also right, then there is a problem for counterpart theory (see also Tappenden [2011]). If each post-fission person is a counterpart of the pre-fission person, then, according to counterpart theory, the pre-fission person can say truly both that she will experience a red recovery room and that she will experience a blue recovery room. And if this is known to the pre-fission person, then, it would seem, she should be 100 percent confident, rather than 50 percent confident, that she will wake in a red room, and also 100 percent confident that she will wake in a blue room. This problem for counterpart theory will occupy us for the remainder of the paper.

The problem is not that the attitudes recommended by counterpart theory are probabilistically incoherent. Being 100 percent confident that:

(R) I will in one day wake in a red room

and also that:

(B) I will in one day wake in a blue room

while maintaining 100 percent confidence that:

It's not the case that in one day I will wake in both a red room and a blue room

is coherent given counterpart theory because the counterpart theorist's tense logic counts these three statements as being logically compatible. The problem is rather that the attitudes argued above to be the one we would in fact have—namely, 50 percent confidence in both (R) and (B)—are surely rationally permissible attitudes, whereas counterpart theory seems to predict that we rationally must be fully confident in both (R) and (B) (or at any rate that anyone who is certain that counterpart theory is true must be fully confident in both).

I insist only that 50 percent confidence in both (R) and (B) is rationally permissible, and not that it is mandatory, because I think that there are two reasonable perspectives, an objective one in which nothing seems uncertain and a subjective one in which one is 50 percent confident in both (R) and (B). I return to this issue below.

There is a similar puzzle that confronts (one version of) Everettian, 'many-worlds' quantum mechanics.<sup>2</sup> According to that view, when a system is in a superposition of sufficiently isolated states  $s_1, s_2, \dots$ , reality in fact contains multiple parts, or 'branches', each of which contains the system in just one of the states. So in a sense, all possible outcomes of any given measurement are actualized, each on one of the branches. The puzzle is where to locate quantum probabilities in this picture. Quantum mechanics gives, via the Born rule, the probabilities of measuring the various states  $s_i$ . We cannot simply give up on this aspect of quantum mechanics, since it is through such probabilistic

---

<sup>2</sup>See Greaves (2007) for an overview, Wallace (2012) for a recent approach, and Lewis (2007) for a discussion of the connections to the personal identity literature on fission. The puzzle considered here is that of how probability or uncertainty is even possible in an Everettian multiverse (Greaves's 'incoherence problem'), and not the puzzle of how to justify the numerical values for probability given by the Born rule (Greaves's 'quantitative problem').



predictions that the theory is confirmed by experiment. But given the Everettian view, nothing seems uncertain, since we know in advance what will occur: *each* outcome  $s_i$  will occur on some branch. Yes, each branch has a ‘weight’, which is a number assigned to it in virtue of facts about the wave function, but in what sense do the branch weights count as ‘probabilities’ if nothing is uncertain?

Measurement in an Everettian multiverse is, as has been noted many times, similar to the case of fission as discussed in the personal identity literature. Each possible outcome of the measurement process is experienced by some observer on some branch, and each of these observers is related to the original pre-measurement person (or person stage) by the sorts of relations that normally unite persons over time, such as memory.

The case of repeated fission discussed in the previous section seems to show that our ordinary concept of uncertainty is not intrinsically incompatible with uncertainty in an Everettian world. As we saw, a subject who has repeatedly undergone fission can be uncertain what she will experience, even though she knows exactly what will happen in an impersonal sense. Moreover, this uncertainty does not depend on the truth of, or belief in, facts about personal identity that break the physical symmetries and thus fail to supervene on the physical facts. But it remains puzzling just how this uncertainty is possible, and puzzling how to accommodate it theoretically.

Although the two puzzles are not perfectly parallel—Wiggensian fission has no analog of the branch weights, for example—each points to the need for an account of how pre-fission uncertainty is possible when one knows all the post-fission facts. My own money is on approaches that take the uncertainty to be subjective, and in particular on an approach due to Jenann Ismael (2003), though I develop it in my own way.<sup>3</sup>

Consider first the situation *after* the fission operation discussed above, when one of the resulting persons wakes up in a recovery room but has not yet opened her eyes (Vaidman 1998, 254). Everyone can agree that it is appropriate for her to be uncertain whether she is in a red or blue room. The uncertainty is ‘*de*

---

<sup>3</sup>My approach complements rather than competes with the decision-theoretic approach to the puzzle, which argues for the rationality of certain preferences concerning future selves and then appeals to an operationalized conception of belief constituted by preferences over lotteries, as in decision theory (Deutsch 1999; Wallace 2012). Thinking of the attitude in indexical terms reduces the decision-theoretic approach’s reliance on operationalism, showing that the states the Everettian requires are akin to ordinary belief states; and the decision-theoretic account can enable a more substantive account of the functional role of the indexical attitude I am about to introduce.

*se*': to express it, she must use the first person pronoun, 'Am *I* in a red or blue room?'

So far we have only uncertainty for the post-fission subject; but according to Ismael, there is a correlative sort of indexical uncertainty for the *pre*-fission subject. My idea is that to account for this correlative sort of uncertainty, we must recognize a distinctive sort of indexical thought. (There are hints of the need for this in the literature: Wallace's [2012, 285–86] remark that the temporal counterpart theorist will require 'a new kind of uncertainty, one which has no analogue in non-branching situations', and perhaps Lewis's [2004, 14] talk of 'expectations'.) Thinking about the future in these cases, wondering 'What will happen to me?', involves a type of thought that does not reduce to theoretically more familiar categories such as (aperspectival) belief or even *de se* belief. The uncertainty of the *pre*-fission subject involves this type of thought.

I call this type of thought *de se futura*. The *pre*-fission subject has *de se futura* uncertainty whether she will experience red or blue, corresponding to the post-fission subject's *de se* uncertainty as to whether she is in a red or blue room. *De se futura* thought is thought about one's own future—hence the *de se* part of the term. But futurity is irreducibly bound up in the attitude—hence *futura*—in that *de se futura* thought is not the same thing as *de se* thought about one's future. The *pre*-fission subject is *de se futura* uncertain whether she will experience red or blue, even though she knows *de se* what will happen to her. She knows exactly which person she is (which person stage, that is); and she knows what will happen to this person: she knows that this person will experience red (because she has a future counterpart who experiences red) and will experience blue (because she has a future counterpart who experiences blue).

The standard approach to *de se* thought was a departure from an older orthodoxy, which held that belief consists in a subject bearing a certain relation, call it belief, to a certain sort of content, call it a proposition. How exactly to conceive of propositions is a matter for theory; all that matters here is that they are aperspectival in being true or false absolutely, rather than relative to persons, places, or times. According to Hector-Neri Casteñeda (1968), David Lewis (1979), John Perry (1993), and other critics, this older orthodoxy cannot accommodate certain thoughts about oneself, one's spatial location, and the present moment that one expresses using the indexical words *I*, *here*, and, *now*. (The reason, in a nutshell, is that one might know all the relevant propositions but still be uncertain who one is, or where one is, or what time it is.) Such *de se* thoughts, on Lewis's version of the view anyway, consist in the subject's bearing a distinctive relation of 'self-ascription'—a relation that differs from

belief—to distinctive contents, ‘centered propositions’. Call a *location* a four-tuple  $\langle w, s, p, t \rangle$ , with  $w$  a possible world,  $s$  a person,  $p$  a place, and  $t$  a time; and call a *centered proposition* a set of locations. Say that person  $s'$  is *located at* location  $\langle w, s, p, t \rangle$  at world  $w'$  and time  $t'$  if and only if:  $w = w'$ ,  $s = s'$ ,  $t = t'$ , and  $p$  is the spatial location of  $s$  at  $t$  in  $w$ . (Assume that a person has no more than one spatial location at a time and world; this is mostly terminological.) To self-ascribe a centered proposition is, intuitively, to think to oneself: *my current location—i.e., (the actual world, me, here, now)—is a member of the centered proposition*. A person who thinks ‘It is now raining’ self-ascribes the set of locations where it is raining at the time of the location; a person who thinks ‘I am Napoleon’ self-ascribes the set of locations in which the person in the location is Napoleon; and so on.

*De se futura* thought, as I conceive of it, also consists in bearing a certain relation to a centered proposition. Thus the contents of *de se futura* thoughts are the same sorts of entities as the contents of *de se* thoughts. But the relation one bears to those contents in a *de se futura* thought is not self-ascription. It is rather a relation that we can call *future-ascription*, and canonically express thus: ‘I futurally will be  $F$ ’. Intuitively, one future-ascribes a centered proposition if—or to the degree that—one expects one’s *future* location (not one’s present location) to be a member of the centered proposition.

Not only is future-ascription distinct from self-ascription, it also cannot be defined in terms of it. In particular, it is crucial that future-ascribing a centered proposition  $S$  not be defined as self-ascribing the centered proposition that one’s location *will* be (in the counterpart-theoretic sense of ‘will be’) in  $S$ . (Thus future ascription is like thinking about ‘there’ rather than ‘here’. One is in effect pointing directly to one’s future and thinking about it, rather than pointing to oneself first, and then thinking about the future of the person thus pointed to.) More exactly, the definition to be rejected is this: one future-ascribes a set  $S$  of centered worlds to degree  $d$  if and only if one self-ascribes to degree  $d$  the set of locations  $\langle w, s, p, t \rangle$  where for some counterpart  $s'$  of  $s$  at some future time  $t'$ , located at place  $p'$ ,  $\langle w, s', p', t' \rangle \in S$ . Given that definition, the proposed solution to the problem of uncertainty in cases of fission would collapse. According to the solution, the pre-fission subject’s 50 percent confidence that she will wake in a red recovery room amounts to future-ascription to degree 0.5 of the centered proposition that she is in a red room. But the definition equates this with self-ascription to degree 0.5 of the centered proposition that she has a future counterpart in a red room, which is not the case as she is certain that she has such a counterpart.

Like any attitude, future-ascription is closely associated with a distinctive functional role, a distinctive way of causing and being caused in a person's cognitive economy. I am not going to attempt to define that role, but I can say a bit about it, in particular about its future-directed part: future ascriptions cause the kind of behavior that is caused by self-ascription of centered propositions about one's future in cases that do not involve fission. Future-ascription of being in Princeton in a year tends to cause one to prepare for life in Princeton, rather than life in Paris. Future-ascription to degrees between 1 and 0 of being in Princeton and being in Paris will tend to cause some amount of preparation for each, or perhaps postponement of certain decisions, depending on the situation, and will tend to cause one to place certain bets on being in Princeton or on being in Paris.

Future-ascribing attitudes might seem irrational to adopt. Why not just adopt attitudes of belief in future-tensed aperspectival propositions, since these bear a more direct relation to the objective facts? The forward-looking causal role of future-ascription yields an answer. Suppose division to be a regular occurrence; and indeed, suppose people to frequently divide into not just two, but sometimes three or more. Suppose further that in region *A* of the world, everyone makes decisions on the basis of future-ascription, whereas in otherwise similar region *B*, everyone makes decisions on the basis of belief in aperspectival future-tensed propositions. And finally, imagine polling each person at the end of her life, and asking her whether she is glad, from a self-interested point of view, that she lived in the region she did. The people in region *A* will express satisfaction with their lot, but not the people in region *B*. Subjects who regularly divide and who make decisions on the basis of future ascription will generally act to benefit a larger number of their successor selves; and so people at the ends of their lives will, on average, regard the adoption of this sort of decision-making as having benefitted them. For instance, if a subject is about to divide into three, and knows that two of the three will emerge in an uncomfortably cold room and one will emerge in an uncomfortably hot room, she will future-ascribe to degree  $\frac{2}{3}$  being in a cold room, let us assume<sup>4</sup>, and accordingly will dress warmly. Inhabitants of region *B* will tend to be less satisfied at the ends of their lives with the prevalent decision-making method in that region. In the example just considered, the agent before division would

---

<sup>4</sup>Uniform degrees of future-ascription over the branches seems intuitively right, but there is a question of what justifies it. Compare the 'quantitative problem' in the Everettian case—note 2.

regard it as certain that she will emerge in a hot room and that she will emerge in a cold room; and whatever that would cause one to do, it presumably would not particularly favor dressing warmly.

It might be objected that the *B*-inhabitants would try to maximize the welfare of their future selves as a group. Perhaps we could imagine such people, but I am imagining that the *B*-inhabitants make decisions in a much more ordinary way: from a self-interested point of view and on the basis of first-person credences. Thus they think, ‘I don’t want to be either too hot or too cold, but I am sure that I will emerge in a hot room, and also that I will emerge in a cold room. So what on earth should I wear?’ The point of the example is that in certain circumstances, decision-making of the ordinary sort would be a more reasonable practice if based on future-ascription credences rather than ordinary ones.

I have advocated counterpart theory and *de se futura* thought. But do we really need both? What different roles do the two play?

Counterpart theory and *de se futura* thought are very different beasts. Counterpart theory is a semantic theory of a tensed language (or, if you like, a metaphysics of persistence); *de se futura* thought is a kind of attitude.

To be sure, attitudes and semantics are connected: one can have attitudes toward the semantic values delivered by a semantic theory. But we must recognize both *de se futura* attitudes as well as more standard attitudes toward the semantic values delivered by counterpart theory. The core of this section’s puzzle is that there are two distinct sorts of attitudes we can have toward our futures: we can be certain about what will happen to us in an objective (or aperspectival) sense, while remaining uncertain what will happen in a subjective (or perspectival) sense. An adequate resolution of the puzzle must do justice to each. The subjective uncertainty is *de se futura*, I say, whereas the objective certainty involves counterpart theory. Counterpart theory is a theory of (statements about) our objective futures, and certainty about such futures consists of familiar relations of belief to propositions associated by counterpart theory with future-tensed sentences.

And even setting aside the need for an account of attitudes concerning *our* objective futures, counterpart theory is needed to deliver the objects of our attitudes concerning *others’* futures. (Although see the discussion of *de illo* attitudes below. I am also tempted to think that *de se futurity* has no role to play in the compositional semantics for tensed sentences, that counterpart theory is the whole story there.)

More concretely, consider the following attempt to do without the *de se*

*futura*. On my version of counterpart theory, each successor person in a case of fission is a counterpart of the original person. But a counterpart theorist might propose a different version of that theory, according to which it is indeterminate which successor person is a counterpart of the original person (although, plausibly, the original person is determinately a counterpart of each successor person). In the fission case discussed earlier, the following sentences would then be indeterminate, as uttered by the original person before undergoing fission:

(R) I will in one day wake in a red room

(B) I will in one day wake in a blue room

An advocate of this approach might then claim that believing each to degree 0.5 is reasonable—‘believing’ in the ordinary, aperspectival sense. Thus the phenomenon would allegedly be accommodated without invoking *de se futura* thought.

Two objections to this approach are clear, given the preceding discussion. First, the approach accounts only for the sense in which the subject is uncertain what will happen to her; it has no account of the sense in which she is certain what will happen—the sense in which she knows, objectively, what will happen to her. Second, it implies that a bystander ought to be uncertain what will happen to the subject, which seems clearly wrong.

(The approach also faces a further objection: numerical degrees of belief—nonzero ones, anyway—seem inappropriate for claims that are believed to be indeterminate. It is inappropriate to wonder whether a borderline pink/red patch of color is red or pink, and thus inappropriate to have degree of belief 0.5 in each proposition.)

We have considered the problem that fission poses for temporal counterpart theory, but fission also poses a problem for Lewis’s account of the metaphysics of persistence. (Lewis [2004] himself argued that an Everettian observer would not be uncertain.) For Lewis, before fission there are two coincident persons, one of whom will wake in a red room and the other of whom will wake in a blue room. On the face of it, there is nothing to be uncertain about given Lewis’s metaphysics, since each of the coincident persons knows all the third-person, aperspectival facts. Yet as noted, uncertainty about the color of the room in which one will wake just does seem to be rational.

Like the counterpart theorist, Lewis could respond by embracing the uncertainty and regarding it as being indexical. He could hold that each of the

coincident pre-fission persons can wonder to herself ‘Will I wake in a red room?’, even though neither person can uniquely refer to herself using ‘I’. Saunders and Wallace (2008) make this suggestion in the case of Everettian fission. The suggestion is more plausible than it may at first seem. After all, everyone agrees that *de se* thought does not require the ability to uniquely identify the object of the thought in any way that is not assisted by the circumstances: one achieves reference using ‘I’, ‘now’, and ‘here’ simply by courtesy of the circumstances. But then, being in unfavorable circumstances, such as the lack of a unique referent for ‘I’, should not undercut one’s ability to have the same *de se* thought. The thought comes first, reference second. (Lewis himself took a different line [1983*b*, postscript A] in the context of replying to Parfit [1976]—a reply he would forfeit by taking the suggestion.)

There is this difference between the suggested Lewisian solution and my own: Lewis does not need the distinctive form of *de se futura* thought, only the familiar *de se*. Thus Lewis’s solution is more conservative.

But in the case of Everettian quantum mechanics, Lewis’s account arguably cannot remain conservative; it must move in my direction. Suppose I face a doom so catastrophic that I have no chance whatsoever of surviving. The present moment is, with chance 1, my last. Suppose also that I know this, and indeed, know the whole truth about the entire Everettian multiverse. Might I not still wonder about the future, about what the world will be like after I am gone, just as at earlier moments I wondered what *my* future would hold?<sup>5</sup> Such end-of-life uncertainty about the future could not be *de se* uncertainty, since at that moment I know exactly which Lewis-person I am. So what would it be?

One wants to call it uncertainty of which branch is *mine*, uncertainty of which complete linear path through the tree-like Everettian multiverse *I* inhabit (compare Saunders and Wallace [2008, 301].) This is on the right track, but misleadingly suggests that the uncertainty is just the familiar sort of *de se* uncertainty. It is not, since if I knew everything about my past and present, and that the present moment is my last, I would know exactly where I am in the multiverse and still have the uncertainty. The uncertainty must rather be taken to be irreducibly demonstrative, ‘*de illo*’. What I am wondering is ‘Which

---

<sup>5</sup>I said above that bystanders could not be uncertain about the outcome of fission. (In the terminology I introduce here, there is nothing like *de illo* uncertainty about what will happen to *that person*.) If end-of-life uncertainty is indeed possible, the difference between it and bystander uncertainty would seem to be that we can have memories of sequences of experiences along just one branch, whereas bystanders can have memories of experiences of multiple products of a given process of fission.

branch is *this branch?*’, thereby expressing an attitude in which the ‘this’ cannot be eliminated, an attitude that cannot be reduced to standard attitudes towards propositions. Alternatively, in counterpart-theoretic terms, taking the branch to be its current stage, my uncertainty would be ‘*de futuro illo*’: I am thinking irreducibly about *this future branch*.

## References

- Casteñeda, Hector-Neri (1968). “On the Logic of Attributions of Self-Knowledge to Others.” *Journal of Philosophy* 54: 439–56.
- Deutsch, David (1999). “Quantum Theory of Probability and Decisions.” *Proceedings of the Royal Society of London* 3129–37.
- Greaves, Hilary (2007). “Probability in the Everett Interpretation.” *Philosophy Compass* 2(1): 109–28.
- Haslanger, Sally and Roxanne Marie Kurtz (2006). *Persistence: Contemporary Readings*. Ed. Sally Haslanger and Roxanne Marie Kurtz. Cambridge, MA: MIT Press.
- Hawley, Katherine (2001). *How Things Persist*. Oxford: Clarendon Press.
- Ismael, Jenann (2003). “How to Combine Chance and Determinism: Thinking about the Future in an Everett Universe.” *Philosophy of Science* 70: 776–90.
- Lewis, David (1968). “Counterpart Theory and Quantified Modal Logic.” *Journal of Philosophy* 65: 113–26. Reprinted in Lewis 1983a: 26–46.
- (1970). “General Semantics.” *Synthese* 22: 18–67. Reprinted in Lewis 1983a: 189–232.
- (1979). “Attitudes *De Dicto* and *De Se*.” *Philosophical Review* 88: 513–43. Reprinted in Lewis 1983a: 133–56.
- (1983a). *Philosophical Papers, Volume 1*. Oxford: Oxford University Press.
- (1983b). “Survival and Identity.” In Lewis (1983a), 55–77. With postscripts. Originally published in Amelie O. Rorty, ed., *The Identities of Persons* (Berkeley: University of California Press, 1976), 17–40.



- (2004). “How Many Lives Has Schrödinger’s Cat?” *Australasian Journal of Philosophy* 82: 3–22.
- Lewis, Peter J. (2007). “Uncertainty and Probability for Branching Selves.” *Studies in History and Philosophy of Modern Physics* 38: 1–14.
- Parfit, Derek (1971). “Personal Identity.” *Philosophical Review* 80: 3–27.  
Reprinted in Perry 1975: 199–223.
- (1976). “Lewis, Perry, and What Matters.” In Amelie O. Rorty (ed.), *The Identities of Persons*, 91–107. Berkeley: University of California Press.
- (1984). *Reasons and Persons*. Oxford: Clarendon Press.
- Perry, John (ed.) (1975). *Personal Identity*. Berkeley: University of California Press.
- Perry, John (1993). *The Problem of the Essential Indexical and Other Essays*. New York: Oxford University Press.
- Saunders, Simon and David Wallace (2008). “Branching and Uncertainty.” *British Journal for the Philosophy of Science* 59(3): 293–305.
- Schechtman, Marya (1996). *The Constitution of Selves*. Ithaca, NY: Cornell University Press.
- Shoemaker, Sydney (1970). “Persons and Their Pasts.” *American Philosophical Quarterly* 7: 269–85.
- Sider, Theodore (1996). “All the World’s a Stage.” *Australasian Journal of Philosophy* 74: 433–53. Reprinted in Haslanger and Kurtz 2006: 91–117.
- (2001). *Four-Dimensionalism*. Oxford: Clarendon Press.
- (2006). “Beyond the Humphrey Objection.” Available at [http://tedsider.org/papers/counterpart\\_theory.pdf](http://tedsider.org/papers/counterpart_theory.pdf).
- Tappenden, Paul (2011). “Expectancy and Rational Action Prior to Personal Fission.” *Philosophical Studies* 153(2): 299–306.
- Thurber, James (1939). “The Secret Life of Walter Mitty.” *The New Yorker* March 18.

- Vaidman, Lev (1998). "On Schizophrenic Experiences of the Neutron or Why We Should Believe in the Many-Worlds Interpretation of Quantum Theory." *International Studies in the Philosophy of Science* 12(3): 245–261.
- Velleman, J. David (1996). "Self to Self." *Philosophical Review* 105: 39–76.  
Reprinted in Velleman 2006: 170–202.
- (2006). *Self to Self: Selected Essays*. New York: Cambridge University Press.
- Wallace, David (2012). *The Emergent Multiverse*. Oxford: Oxford University Press.
- Wiggins, David (1967). *Identity and Spatio-temporal Continuity*. Oxford: Basil Blackwell.
- Williams, Bernard (1981). *Moral Luck*. Cambridge: Cambridge University Press.