1.

The eponymous anti-hero of James Thurber’s story “The Secret Life of Walter Mitty” escapes his dreary suburban life by daydreaming.¹ Thurber’s Mitty’s dreams are various, 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, say. 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, always remembers the boring life of Daytime Mitty.² But Daytime Mitty cannot remember any of the nighttime events.

I would like to put forward two judgments about Mitty. 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 Nightime Mitty. If Daytime Mitty were to say “I will fight crime this evening”, that would be false. These two judgments, that Nightime Mitty was Daytime Mitty, and that Daytime Mitty will not be Nightime Mitty, together yield the conclusion that personal identity for Mitty is asymmetric. Asymmetric personal identity might seem metaphysically absurd; but before addressing that issue, let us attend first to the intuitive case in its favor.

Few will balk at the first judgment, since Nighttime Mitty remembers what Daytime Mitty does. Nightime Mitty will surely regard Daytime Mitty’s doings as his own, will judge himself responsible for Daytime Mitty’s misdeeds, will regret those misdeeds, and so forth; and these judgments from the inside match our own judgments from the outside. The second judgment will meet more resistance; but careful thought about the case, I believe, supports this

¹“Quasi-remembers” if you prefer (Shoemaker, 1970).

²Thanks to Karen Bennett, Andrew Chignell, Jenann Ismael, Daniel Manne, Kate Manne, Andrew McGonigal, Jill North, Erin Taylor, David Velleman, …

¹The New Yorker, March 18, 1939.
The presence of certain evaluative, rational, and moral relations are often used to “test” for personal identity. One person’s bearing these relations to—“mattering to”, in Parfit’s\(^4\) (1971) terminology—another counts heavily in favor of those persons being identical. Conversely, not mattering counts against identity.

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.\(^5\) This is a backward-looking test, and delivers a clear verdict in the case of Mitty. Nighttime Mitty can regret Daytime Mitty’s misdeeds, and so the test tells us that 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 her, and that if good or bad things will happen to her, then it is now good or bad for me that those things will happen—it is good news or bad news for me that those things will happen. This is a forward-looking test, and from the perspective of Daytime Mitty it speaks against his identification with Nighttime Mitty. Life as a crime-fighting vigilante is what Nighttime Mitty most wants. So if it is true that he will do just that—if he in fact will be Nighttime Mitty—then by the test, the fact that Nighttime Mitty fights crime is now good for Daytime Mitty. But this is surely not the case. Nighttime Mitty’s future exploits do Daytime Mitty no good at all. Thus the test tells us that Daytime Mitty will not be Nighttime Mitty. Similarly, it is intuitive that Daytime Mitty is not rationally required to care about Nighttime Mitty’s well-being;\(^6\) thus the test tells us again that Daytime Mitty will not be Nighttime Mitty.

\(^3\)The judgment would be less clearly correct if Nighttime Mitty sometimes followed through on Daytime Mitty’s intentions, so let us stipulate that this does not happen.

\(^4\)Parfit himself famously claimed that these tests are not always accurate, that identity is not ultimately what matters. We will discuss this view below.

\(^5\)The relevant sort of regret is “agent-regret” (Williams, 1981); but the principle would be trivialized if agent-regret were simply defined as regret for what one did oneself. We must, rather, pick out agent-regret ostensively, as it were, as a distinctive sort of regret that ordinarily is for what one did oneself, but which isn’t simply defined that way.

\(^6\)At least, not from a self-interested point of view, which is what is relevant to the test; see note 5.
Daytime Mitty’s ignorance of what happens at night may make it hard to assess whether Nighttime Mitty’s doings are good for Daytime Mitty. But imagine being Daytime Mitty, and ask yourself whether it would be good for you if the story described above were true. I think you would answer that it would not.

Furthermore, if you were then told that the story was true, 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 doing what you always wished you could do, and—the ultimate irony—doing it with your body.

Imagine someone using your body by remote control to fight crime, as a sort of puppet, while you slept. Here there is a clear sense in which the deeds are not “your own”, even though they are done by, or using, your body. Next remove the puppeteer and imagine the deeds done while sleepwalking. Here too the deeds are not your own. But now, finally, compare this sleepwalking story with Mitty’s. Nighttime Mitty’s deeds seem no more owned by Daytime Mitty as your sleepwalking deeds are owned by you.

Now that the case for asymmetric personal identity has been made, we may inquire into its source. What feature of the grounds of personal identity allows for asymmetry? The answer is that the relations that are criterial of personal identity are many in number and are not symmetric; and in some cases, the contribution from one of these relations to personal identity—to one person’s “identifying” with another, to counting the other’s perspective as lying in the first person’s future or past—is on just one “side” of the relation. When a later person remembers the doings of an earlier person, for example, the case for identification is strong from the later person’s perspective; but it 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. We might put this by saying that the relation of remembering is a “one-sided identification relation” and that it “identifies” from the later-self’s point of view. More briefly, we might call memory an “identification relation for later selves”. Ordinarily there is an intuitive case for identification both from the perspective of the earlier self and from the perspective of the later self, since ordinarily some one-sided identification relations identify from the earlier self’s perspective and some identify from the later self’s perspective. But in certain extraordinary cases like that of Mitty, there is an intuitive case from only one of these perspectives: although some one-sided identification relations identify from Nighttime Mitty’s perspective, no
one-sided identification relations identify (or identify strongly) 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, pp. 194–5) 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.” Thus anticipation is an identification relation in my terms. It is also, intuitively, a one-sided one: it is the anticipating earlier self who is led to identify, not the later self who is anticipated. Further, Daytime Mitty cannot anticipate Nighttime Mitty's experiences, for the simple reason that he knows nothing of Nighttime Mitty. 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.” (p. 198)

So far my argument that Daytime Mitty cannot anticipate Nighttime Mitty's experiences has been based on the fact that Daytime Mitty knows nothing of Nighttime Mitty. But, interestingly, this conclusion seems—to me, anyway—justified even if the story is varied so that Daytime Mitty learns of Nighttime Mitty's existence. When I put myself in Daytime Mitty's shoes, and imagine knowing in the abstract that—to put it question-beggingly—I have been dressing up as a superhero every night—I seem to be blocked from anticipating these events. Assuming this is so, why is it so? It's a bit puzzling. After all, once Daytime Mitty knows the score, he can form appropriate beliefs about what is about to occur as night approaches. Moreover, as noted above, Velleman stresses that when one anticipates a future experience, one expects the future experience to contain memories of the anticipation; but Daytime Mitty knows that Nighttime Mitty remembers Daytime Mitty, and thus will remember any anticipation. What seems to block the anticipation, somehow, would seem to be, first, the fact that Daytime Mitty knows he will wake up the following morning remembering nothing of the night, and second, the fact that he knows that this happens repeatedly. But why exactly do these facts block anticipation?
I’m not sure.

There is arguably a second earlier-self identification relation that ordinarily holds, but does not in Mitty’s case. As Marya Schechtman (1996) and others have maintained, narrative is an important part of what holds our lives together. Let us put the idea this way: the contribution of narrative to the identity of oneself with an earlier or later self $S$ is proportional to the extent that $S$’s doings are included in some coherent extended narrative that “is one’s own”. (Schechtman herself would not accept this, since she does not regard narrative as a criterion of personal identity in the sense of numerical identity.) Daytime Mitty fails to identify with Nighttime Mitty—even after being told of the nighttime exploits—in part because Nighttime Mitty’s doings are not included in the only narrative that is Daytime Mitty’s own: a story of unfulfilled dreams and suburban existence.

But why is the story of suburban existence the only narrative that is Daytime Mitty’s “own”? In particular, why doesn’t the “combined” narrative that includes the nighttime exploits in addition to the suburban drear count as Daytime Mitty’s own? The combined narrative certainly counts as Nighttime Mitty’s own; he would regard it as being his story.

The answer is that the combined story includes a large, natural, integrated part—the sum of its nighttime segments—from which Daytime Mitty is cut off: cut off in the sense of one-sided identification relations. In particular, Daytime Mitty does not remember any of this part. 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 being his own. It is of course possible to forget some parts of one’s own narrative. But Daytime Mitty cannot remember any of this large, natural, integrated part.

Thus the following relation between selves $S$ and $O$ is a relation of one-sided identification: there existing some narrative that is $S$’s own which includes the doings of $O$. The relation is identifying from the point of view of $S$, but not necessarily from the point of view of $O$, since there is no requirement that the narrative in question is $O$’s own. Nighttime Mitty bears the relation to Daytime Mitty, since there is a narrative—the combined narrative—that is Nighttime Mitty’s own and which includes the doings of Daytime Mitty. This contributes to Nighttime Mitty identifying with Daytime Mitty, but it does not contribute to Daytime Mitty identifying with Nighttime Mitty. Moreover,
Daytime Mitty doesn’t bear this relation to Nighttime Mitty, since no narrative that is Daytime Mitty’s own includes Nighttime Mitty’s actions.

2.

Personal identity is asymmetric only in extreme cases like Mitty’s, where there is a severe disconnect between different one-sided identification relations. But a milder disconnect is present in mundane cases, and this is significant as well. We ordinarily believe that personal identity is a symmetric matter, not just in the sense that if \( x \) was \( y \) then \( y \) will be \( x \), but also in the sense that the kinds of rational and psychological relations that bind our selves together do not create any important asymmetry between past and future selves. For example, consider the question of how an earlier self is rationally obliged to treat a later self, and the question of how the later self is morally constrained by the earlier self. We ordinarily assume, I think, that these questions have parallel answers. If there is a strong case to be made, for instance, that the later self’s well-being is of great rational concern to the earlier self, then, we assume, the later self must also, and for the same reasons, and to the same extent, be morally responsible for wrongdoing of the earlier self. But this thinking is undermined given what we have learned about one-sided 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 as well as from the point of view of the later self, the operative one-sided identification relations differ in each case, and so the nature or strength of the normative connection might differ. It is an open and interesting question just how equal in nature and strength the normative connections between our earlier and later selves ordinarily are.

To be concrete, consider the role of narrative in uniting one’s life. Late in life one strongly owns a narrative that includes the actions of one’s earlier self. But very early in life, in adolescence, say, this particular narrative—the one that will eventually come to constitute one’s life—is just an abstraction. It is one of many narratives that might become the adolescent’s, but it is not really yet hers to any significant degree. The only narratives that are, at this point, really hers may well not be owned by her later self. So as far as narrative is concerned, the ties binding the later self to the earlier self are much stronger than those binding the earlier self to the later self. Now, there may be other relations that bind the earlier self more tightly to the later self. 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? The current argument suggests that these questions might be answered quite differently.

3.

Let us turn, next, to making metaphysical sense of asymmetric personal identity. 
First of all, there is no cause to question the standard logic of the identity 
relation. The fact that $x = y$ whenever $y = x$ is as strongly confirmed as 
can be in other contexts—contexts that are more fundamental and in better 
theoretical shape 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 the case of “fission”, in which an earlier person $E$ is “split” into two later persons, $L_1$ and 
$L_2$, 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, p. 52). Logically, the puzzle is that although $L_1$ and $L_2$ are 
clearly distinct persons, $E$ seems to be the same person as $L_1$ and also as $L_2$ 
(for surely $E$ would be one of the successor persons if the other one had never 
existed, and surely the mere presence of the second successor person cannot 
destroy $E$’s ability to survive as the first); and if so, personal identity cannot be 
both transitive and symmetric.

Derek Parfit’s (1971) response to the puzzle 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 that we 
usually assume it has. We ordinarily assume that ceasing to exist is very bad, that 
one can only be responsible for what one does oneself, and so forth (recall the 
tests). But even though $E$ goes out of existence upon division, this is not bad for 
her, according to Parfit. Each of the successor persons $L_1$ and $L_2$ preserves what 
is important to the original person; identity is not “what matters in survival”.

But according to David Lewis, personal identity needn’t be divorced from 
what matters if one accepts an appropriate metaphysics of personal identity. In 
order for personal identity and what matters to coincide in the case of fission,
what a metaphysics of personal identity must deliver is that no person goes out of existence in that case. And this, Lewis says, can be achieved by saying that there was no single person such as E before the fission. Rather, each of the successor 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 lump 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 person-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, 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. On Lewis’s view, two persons sometimes share a single person stage. In such a case, if the first-person pronoun ‘I’ is uttered, to what does it 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 to say that pre-fission uses of ‘I’ are semantically indeterminate between the two coinciding persons, and to adopt a supervaluational semantics. 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 of the containing persons is F (supervaluational semantics), and neither true nor false if one is F and the other is not (neither supertrue nor superfalse).

So since Nighttime Mitty’s utterance of ‘I lived a boring suburban life earlier today’ is true, each Lewisian person containing the uttering Nighttime Mitty stage also lives a boring suburban life earlier in the day. But each person stage is part of at least one person.\(^7\) So there is at least one person containing the uttering stage who lives a boring suburban life earlier in the day. That is, there is

\(^7\)This is a condition of adequacy on any temporal parts account of persons. In Lewis’s own theory it follows from his definition of persons as maximal R-interrelated sums.
at least one person, \( P \), containing stages of both Nighttime Mitty and Daytime Mitty. 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 also exists a person, call her “Boring-Mitty”, made up of all and only the boring, daytime stages.\(^8\) 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 Nighttime Mitty can look forward to fighting crime. In the case of \( \text{fission} \), the “indeterminate” future that Lewis’s account delivers is one we can accept. Suppose that \( L_1 \) wakes up after \( \text{fission} \) in a red recovery room and \( L_2 \) wakes up in a blue recovery room. Lewis’s view then implies that a pre-\( \text{fission} \) utterance of “I will wake up in a red recovery room” is neither true nor false; and this, perhaps, is not intuitively unacceptable. But the intuitive facts in Mitty’s case are different: Daytime Mitty determinately will never fight crime.

We have seen that Lewis’s metaphysical response to Parfit does not extend to the case of asymmetrical identity. But Lewis’s is not the only such response. There is also the “stage view”, or “temporal counterpart theory” defended by Katherine Hawley (2001) and myself (1996; 2001). And unlike Lewis’s view, temporal counterpart theory does allow asymmetric personal identity.

According to temporal counterpart theory, persons are person stages, not aggregates of person stages.\(^9\) 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

---

\(^8\)This attempt to accommodate asymmetric identity in Lewisian terms also faces another challenge. If the Lewisian wants to preserve the truth of Nighttime Mitty’s utterance of “I lived a boring suburban life earlier today”, he must not count “Exciting-Mitty”, the aggregate of all the nighttime stages, 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.

\(^9\)See Sider (2006) for my currently preferred view on how the temporal extent of the stages that are persons is determined.
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 $A$ is true of some past temporal counterpart of $x$. Similarly for the future tense operator: $FA(x)$ is true if and only if $A$ is true of some future temporal counterpart of $x$. (Compare Lewis’s claim that $\diamond A(x)$ is true if and only if $A$ 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.

Temporal counterpart theory delivers different results from Lewis’s theory in the case of fission. As we saw, Lewis’s theory implies that a pre-/fission utterance of ‘I will wake up in a red room’ is neither true nor false, since one of the persons containing the uttering stage wakes up in a red room and the other does not. But according to temporal counterpart theory, on the other hand, that utterance is true since its counterpart-theoretic truth-condition is that the person uttering the sentence—namely, the stage uttering the sentence—has at least one future counterpart who wakes up in a red room.

Notice that the counterpart-theoretic account generates an odd tense logic in this case. Let $F_1$ be the tense operator for “it will be the case one day hence that”. The counterpart-theoretic semantics for this operator is this: $F_1A(x)$ is true if and only if $A$ is true of some temporal counterpart of $x$ that is located one day after the time of utterance. But then if we symbolize ‘$x$ will wake in a red room’ as $Rx$, $F_1Rx \land F_1(\sim Rx)$ comes out true of the pre-fission stage; for that stage has two counterparts one day hence, one who wakes in a red room and one who does not. Thus that stage 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 he 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”; $F_1(Rx \land \sim Rx)$ is false. Thus $F_1A \land F_1B$ fails to imply $F_1(A \land B)$. 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 that is demanded in the metaphysically odd case of fission, and will point out that classical logic in the extensional, tenseless metalanguage is not threatened.

The case of Mitty is straightforward for the counterpart theorist, since the counterpart relation can fail to be symmetric. Although Daytime Mitty stages are temporal counterparts of Nighttime Mitty stages, Nighttime Mitty stages are not temporal counterparts of Daytime Mitty stages. (The counterpart relation can be taken to be determined by the various nonsymmetric one-sided
identification relations.) Nighttime Mitty can therefore truly utter “I lived a
boring suburban life earlier today”, since he has a temporal 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
temporal counterparts that fight crime.

As with fission, this satisfying account comes at the price of a nonstandard
tense logic. For instance, since Nighttime Mitty fights crimes, but has past
counterparts—namely, daytime stages—with no future counterparts that fight
crime, he can truly utter “Although I fight crime, it was the case that it would
never be the case that I fight crime”: \( C x \land \neg P \neg \exists C x \). Thus \( A \) fails to imply \( H F A \),
where \( H \) is the Priorian operator “it always has been the case that”.

So if we are going to wheel in a metaphysics+semantics of persistence 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 asym-
metric personal identity, the right one to wheel in is temporal counterpart
theory, and not Lewis’s theory of overlapping aggregates of stages.

4.

Velleman argues that the case of fission provides another sort of asymmetric
personal identity.\(^{10}\) According to him, 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 sort of first-personal thought will do.
There is a sense in which one can think about others’ experiences—Napoleon’s
say—from the first-person point of view: 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, p. 188). But in the case of genuine first-personal thought, no
stipulation is needed; the centering on oneself is “automatic”. Velleman arrives
at this view through an analysis of imagination, memory, and anticipation; but
he then applies his view 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 to their thoughts

\(^{10}\) See Velleman (1996, pp. 200–2), especially note 53.
must be via stipulation, and thus amounts to mere imagining, rather than genuine anticipation. One cannot just think “what will I be doing tomorrow?” Thus we have an independent case for asymmetric personal identity.

The sort of “personal identity” involved in the asymmetry, according to Velleman, is not the persistence of the numerically identical self over time, which is simply numerical identity and therefore cannot possibly fail to be symmetric on Velleman’s view, but rather is a certain relation of “being a self for”. Nevertheless, with metaphysical views like Lewis’s and counterpart theory on the table, we may consider the argument as concerning persistence.

If Velleman is right about fission, one might take this as a further reason to favor temporal counterpart theory over Lewis’s metaphysics of persons. Temporal counterpart theory certainly handles the case smoothly: the pre-fission stage is a counterpart of each of the two post-fission stages, whereas, it might be said, the post-fission stages aren’t counterparts of the pre-fission stage (because they cannot be anticipated). Thus the pre-fission person, on this account, could say truly “I will not survive fission”, whereas each post-fission person can say truly “I existed before fission”. The Lewisian, on the other hand, can say no such thing. For if each post-fission person existed before fission, then before fission there were two stage-sharing persons each of whom will survive the fission.

But it is not clear that Velleman is right about fission. First, it is only in the case where the pre-fission person knows of the impending fission that Velleman appears right. If the person knows nothing of the fission, and anticipates doing certain things on the following day, there is no intuitive barrier to regarding her first-personal thoughts as attaching to each of the post-fission persons.

Second, even when the impending fission is known, this does not seem to be an insurmountable barrier to anticipation. For imagine fission to be a regularly occurring process. Each time, the subject enters a certain fission chamber, goes to sleep, and is then divided in two. One resultant person wakes up in a red recovery room, suppose, and the other wakes up in a blue recovery room. The first time, perhaps, the subject will find it hard to anticipate experiencing anything at all after fission. But consider the situation after ten iterations. The subject will remember entering the fission chamber ten times, and will remember waking up in a recovery room ten times. Suppose she remembers waking in the red recovery room five times, and in the blue recovery room five times. Suppose, moreover, that she compares notes with others who have gone through this process, and finds that, on average, they also remember waking in the red room half the time and the blue room half the time. Putting myself in
this subject’s shoes, when confronting the eleventh iteration, 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% likely to be red and 50% likely to be blue.

5.

If i) is right then Velleman is not right about fission, and so the case for coun-
terpart theory as against Lewis’s approach must rest solely on the case of Mitty.
But if ii) is also right, this generates a puzzle for counterpart theory. If each
post-fission person is a counterpart of the pre-fission person, then 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% confident, rather
than 50% confident, that she will wake in a red room, and also 100% confident
that she will wake in a red room.

The puzzle is not that the attitude recommended by counterpart theory is
probabilistically incoherent. For recall the counterpart-theorist’s nonstandard
tense logic: both “I will wake in a red room” and “I will wake in a blue room”
are true (though “I will both: wake in a red room and wake in a blue room” is
false). The problem is rather that this attitude conflicts with what seems to be
the rational attitude. For surely the attitude argued above to be the one we
would have (namely, 50/50 odds between red and green) is in fact the rational
one.

There is a similar puzzle that confronts the Everettian (“many-worlds”)
interpretation of quantum mechanics.11 According to that interpretation, when
a system is in a superposition of states $s_1, s_2, \ldots$, 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 probabilities and
uncertainty in this picture. Quantum mechanics gives, via the Born rule, the
probabilities of measuring the various states $s_i$ (these correspond to the branch
weights). We cannot simply give up on this aspect of quantum mechanics, since

11See Wallace (2012) for a recent defense and overview; see Lewis (2007) for a discussion of
the connections to the personal identity literature on fission. The puzzle I consider here is
that of how probability or uncertainty is even possible in an Everettian universe, and not the
puzzle of how to justify the particular probabilities given by the Born rule.
it is through such probabilistic predictions that the theory is confirmed by experiment. But given the Everett interpretation, 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 why do the branch weights count as probabilities, and how do they lead to uncertainty?

Measurement in an Everettian multiverse is, as has been noted many times, very 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; but 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. So the Everettian observer must somehow be uncertain about which future she will experience, even though she is—and may indeed know that she is—related by identification relations for future selves to each branch person.

The case of repeated fission discussed in the previous section seems to show that, at an intuitive level, our concept of uncertainty leaves room for uncertainty in an Everettian world. As we saw, it doesn’t seem unnatural in the slightest for a subject who has repeatedly undergone fission to rationally feel uncertainty about 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. Since the case of the Everettian observer is parallel, uncertainty for her seems equally intuitive. But it remains puzzling just how this uncertainty is possible, and puzzling how to accommodate it theoretically.

The two puzzles are not perfectly parallel; Wigginsian fission has no analog of the branch weights, for example. Still, one naturally expects a parallel solution to both. My own money is on an approach due to Jenann Ismael (2003), though I will develop it in my own way.\textsuperscript{12} The key move is to

\textsuperscript{12}This approach is at a different theoretical “level” from, and does not compete with, certain other solutions to the puzzle, such as those that argue for the rationality of certain preferences concerning future selves and then appeal to an operationalized conception of belief constituted by preferences over lotteries, as in decision theory (Greaves, 2004; Wallace, 2012). Thinking of the attitude in indexical terms reduces Greaves and Wallace’s reliance on operationalism about belief, and shows how the attitude that the Everettian must invoke is akin to other sorts of attitudes with which we are already familiar; Greaves and Wallace’s work enables a substantive account of the functional role of the indexical attitude.
hold that uncertainty in cases of fission—whether Everettian or Wiggensian—has an indexical element. What the subject is uncertain about—and what her credences attach to—are not propositions about what objectively will occur, since there is no uncertainty about such propositions. Rather, she is uncertain about what will happen in her future, in a certain indexical sense.

To bring out this indexical sense, consider first the situation after the fission operation discussed above, where one of the resulting persons is waking up in a recovery room but has not yet opened her eyes to see the color of her room. Everyone can agree that it is appropriate for her to be uncertain whether she is in a red or blue room. Moreover, the uncertainty here is “de se”, or indexical: 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 the key idea of the present approach is that there can be a correlative sort of de se uncertainty for the pre-fission subject. Intuitively, the pre-fission subject thinks of her future experiences as those of an unspecified one of her candidate future selves. Since the future self is unspecified, the pre-fission subject does not know anything about those experiences beyond what the post-fission subject knows before she opens her eyes (and otherwise perceptually attends to her surroundings). Thus the pre-fission subject has a sort of 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.

Like the uncertainty of the post-fission subject, the uncertainty of the pre-fission subject can also be regarded as indexical or de se, though it differs from the sorts most commonly studied. The attitudes most commonly studied are all “centered” on the subject: the subject thinks about herself using ‘I’, about her position using ‘here’, about her time using ‘now’. Uncertainty about the future could be indirectly accommodated within this orthodoxy, but only in a way that would not help in the case of fission: a subject can have an orthodox de se thought about herself and the current time, to the effect that she will experience certain things after that time. This is unhelpful because it delivers no uncertainty; in advance, the subject knows in the orthodox de se sense that she herself will wake up in a red room, and that she herself will wake up in a blue room. What we need is an irreducibly new de se attitude, which is

13See Casteñeda (1968); Lewis (1979); Perry (1993).
14It is worth emphasizing that even though temporal counterpart theory says that I (a stage) will be F in virtue of what happens to other objects (my future counterparts), it nevertheless says that I myself will be F. My counterparts are indeed involved in the analysis of the tensed property will-be-F, but I myself possess that tensed property. This is the temporal analog of a
directly centered on the future, so to speak. Call this attitude future ascription, and canonically express it thus: “I futurely will be F”.

Future ascription is akin to thinking about “there” rather than “here”. One is in effect pointing directly to the future—or rather, to a future self—and thinking about it, rather than pointing to oneself first, and then thinking about the future selves of the person thus pointed to. But the pointing is unspecific; in future ascription there need be no particular future self one is thinking about. Like ordinary de se thought, the thought is not fundamentally individuated by its referent, but rather by the cognitive role of the thought. But unlike ordinary de se thought, the cognitive role of the thought does not single out a unique referent.

Think of the having of an ordinary de se thought as the drawing of a kind of mental circle, and thinking about whatever is inside that circle. The subject draws circles using ‘I’, ‘here’, and ‘now’; and these circles are “automatically” centered on the subject herself, her location, and her time. The automaticity consists in this: the fact that she herself, her location, and her time are at the center of these circles does not hold in virtue of her possessing descriptive information about these objects (the circles are not descriptions); rather, the cognitive role of de se thought, together with the location of the thought itself, creates the circles; and the subject, her location, and her time just happen to be the only subject, location, and time in the circles. Future ascription is similar: the subject is thinking about what is going on in a circle located in the future, which has been drawn automatically by the cognitive role of future-ascription, together with her identity and location. Thus the fact that future ascription is about her future selves (though not about any one determinately), and not about the future selves of anyone else, or about her past or present selves, is not accomplished descriptively. But the circle is not determinately centered on any single one of her future selves.

It may be objected that we cannot make sense of this “indeterminate” centering. But here it is crucial to remember that even for orthodox sorts of de se thought which do have a unique referent, that referent is inessential to the thought’s role. To have a de se thought about a person, x, or a time, t, or a place, p, does not require the usual sorts of cognitive access (whether descriptive or causal) to x, t, or p. (One simply needs to be, or be located at, the referent.) Moreover, as Perry (1979) emphasizes, when we individuate thoughts by the

---

point that Lewis (1986, p. 196) famously made about modal counterpart theory; see also Sider (1996, pp. 437–8, 446–7).
kinds of behavior they tend to cause, we will group together de se thoughts with different referents: distinct speakers who think to themselves “I am making a mess” will behave similarly despite referring to different persons using ‘I’. The referent of a de se thought is extrinsic to that thought, then, both regarding the conceptual prerequisites for having the thought, and regarding certain distinctive aspects of the cognitive role of the thought. So there should be no barrier to recognizing a sort of de se thought that—unlike the usual sorts—does not possess a unique referent. The having of the thought is the drawing of the circle; whether anything, or any unique thing, is in the circle is secondary.

Of course, if the putative circle bore no relation to reference whatsoever, the manner of “thought” in question would be useless, and would have no role to play in a cognitive economy. It’s not good practice to have a sort of thought, and to be disposed to act on it, if thoughts of that sort are not reliably connected to parts of the world. But future-ascription does have a reliable connection to the world: the future selves of the future-ascriber are—albeit “indeterminately”—referents, and other selves (selves of other people, past or present selves of the future-ascriber) are not referents. This allows future ascription to have a nontrivial functional role, by virtue of which it deserves a place in our cognitive economy.

About that functional role, and in particular, its future-directed part: future ascriptions cause the kind of behavior that is caused by ordinary self-ascription of propositions about the future in cases that do not involve fission. Future-ascription of being in Ithaca in a year tends to cause one to prepare for life in Ithaca, for example, rather than life in Paris. Future-ascription-uncertainty (degrees of future ascription between 1 and 0) about whether one will be in Ithaca or 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 acceptance of certain appropriate bets on being in Ithaca or on being in Paris.

Future-ascribing attitudes might seem irrational to adopt. Why not just adopt attitudes of belief in future-tensed 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 halves, 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 ordinary belief in objective 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 lives in region A or B. The people in region A will express satisfaction with their lot, but not the people in region B. Subjects about to divide who make decisions on the basis of future ascription will, in general, act so as to benefit a larger number of their successor selves; and so people at the end 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 2/3 being in a cold room, let us assume, and accordingly will dress warmly. Subjects in region A, on the other hand, will, on average, be less satisfied at the ends of their lives with the prevalent decision-making method in that region. For in the example just considered, the agent before division would 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.\footnote{It is interesting to compare the case where fusion in addition to fission also regularly occurs. The preceding argument is much less clear, since subjects can no longer look back on “linear” histories. And indeed, it is unclear whether subjects of this sort would continue to regard their futures with uncertainty. (Fusion is more deeply alien than fusion because our pasts are more immediate to us (via memory) than are our futures (via anticipation).)}

The contents of future ascriptions can be modeled in the very same way as ordinary de se attitudes are: as sets of possible individuals or world-time-individual triples, depending on your metaphysical predilections. But the future-ascription attitude is different from the standard de se attitudes—for example Lewis’s attitude of “self-ascription”. Future-ascription is oriented toward the future, and, moreover, does not reduce to self-ascription: to future-ascribe a set S of centered worlds is not to self-ascribe the set S' of centered worlds in which the person at the center will in the future be the center of some member of S (for that would be of no use here).\footnote{One might adopt a relativist account of truth, as in MacFarlane (2008), for these attitudes.}

It is necessary to theorize in terms of future ascription only if fission—of an Everettian or other sort—regularly occurs. This does not mean that inhabitants of a world of fission are different “from the inside”, or need to make special mental preparations to think about their futures. What is different about a world of fission is how thinkers are related to the world; the regular fact of fission would require us to reconfigure our theory of that relation.

We have considered the problem that fission poses for temporal counterpart
theory, but fission poses a similar problem for Lewis, and a similar solution can be given.\footnote{Lewis (2004) himself argued that an Everettian observer would not be uncertain.} For Lewis, there are two coincident persons, $L_1$ and $L_2$, before fission, 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. Before fission, each person knows all the third-person, non-indexical facts. Moreover, since each is incapable of referring to herself uniquely using ‘I’, neither person, apparently, can wonder in the first-person singular: “will I wake up in a red room?”. Moreover, each already knows that the answer to the “first-person plural” question “will one of us will wake in a red room?” is yes.\footnote{Lewis’s account implies that each of ‘I will wake in a red room’ and ‘I will wake in a blue room’ is neither true nor false, and that their disjunction is true. Given natural assumptions this means that the subject should utter the disjunction while refraining from uttering either of its disjuncts. Wallace (2012, p. 269) suggests that this pattern justifies regarding the subject as being uncertain; but the same pattern holds in cases of vagueness, for instance, in which there is no uncertainty.} 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 can respond by embracing the uncertainty and regarding it as being indexical. Even though neither Lewisian pre-fission person can uniquely refer to herself using ‘I’, Lewis could still hold that each can wonder to herself “which of these coincident persons am I?” Just as the counterpart theorist invoked a de se attitude without unique de se reference to account for uncertainty in the case of fission, so too Lewis can hold that his overlapping persons can express de se uncertainty using ‘I’ despite being unable to uniquely refer to themselves. 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 attitude. Thus far, Lewis’s solution to the puzzle is parallel to that of the temporal counterpart theorist. But there is this difference: Lewis does not need to invoke a new sort of de se attitude. He needs only the familiar de se: the ‘I’ in ‘which of these overlapping persons am I?’ is centered on the utterer rather than the future.
References


