Diachronic Self-making

David Mark Kovacs

Tel Aviv University

Abstract: This paper develops the Diachronic Self-making View (DSV), the view that we are the non-accidentally best candidate referents of our 'I'-beliefs. A formulation and defense of DSV is followed by an overview of its treatment of familiar puzzle cases about personal identity. The rest of the paper focuses on a challenge to DSV, the Puzzle of Inconstant 'I'-beliefs: the view appears to force on us inconsistent verdicts about personal identity in cases we would naturally describe as changes in one's *de se* beliefs. To solve this problem, the paper defends the possibility of overlapping people and addresses a number of objections to this idea.

1. Introduction

According to conventionalist views of personal identity, our persistence conditions are in some sense up to us.[1] And according to *private* conventionalist views, the conditions of each person's persistence are up to that very person.[2] In this paper I will develop a version of private conventionalism, according to which what I believe about myself affects what I refer to when I think 'I'.

I should flag at the outset what my brand of conventionalism implies and what it doesn't. According to *strong* conventionalist views, certain persistence facts obtain because of our conventions. I reject this view. Instead, I endorse only a *weak* kind of conventionalism, according to which each object's persistence conditions are mind-independent, but it's a

---

[1] Cf. Sidelle [1999] and Braddon-Mitchell and West [2001]

[2] Cf. Johnston [1989], Braddon-Mitchell and Miller [2004], and Zimmerman [2013: sec. 4]

matter of convention what our personal pronouns and names refer to.[3] According to the view I will develop, what we refer to with 'I' depends on what we think about ourselves. In recent work [Kovacs 2016] I defended a similar account of *synchronic* personal identity, the 'self-making view': each of us, at any time, is the best candidate referent of our 'I'-beliefs.[4] In this paper, I explore the prospects of a private conventionalist view of diachronic personal identity along similar lines. I will call the resulting view the Diachronic Self-making View (DSV).

The rest of the paper will go as follows. In section 2, I will formulate and defend DSV against the backdrop of a perdurantist ontology. In section 3, I will show how DSV deals with familiar cases from the personal identity literature. In section 4, I will address what I call the Puzzle of Inconstant 'I'-beliefs: DSV appears to force on us inconsistent persistence judgments in situations we would naturally describe as one in which a person's beliefs about her persistence conditions *change*. I will argue that DSV can handle this challenge, albeit at a cost; however, the cost is by no means unacceptable.

2. Background commitments, formulation, and defense

DSV is a *weak* conventionalist view: it doesn't say that any object's persistence conditions are up to you, only that within certain constraints, it's up to you which object (with mind-independent persistence conditions) qualifies as *you*. Suppose, for instance, that an evil

---

[3] See Kovacs [2016] for the weak vs. strong distinction. Self-described conventionalists are often not explicit about the distinction. See Eklund [2004] for weak and Miller [2013] for strong conventionalism.

[4] In Kovacs [2016] I used the terminology "I'-thought' rather than 'I'-belief, but as an anonymous referee pointed out to me, 'thought' has the connotation of an occurrent mental state. What's really at issue are *beliefs*, which may or may not be occurrent at the time when we assign them a referent.

scientist removes your cerebrum, leaving behind an empty-headed human vegetable. According to biological theories, you persist as the vegetable[5], whereas according to neo-Lockean ones as the cerebrum.[6] According to DSV, by contrast, there is something ('Thinker') that is a body-cerebrum composite before and a detached cerebrum after the surgery and something ('Breather') that is a body-cerebrum composite before and a human vegetable after the surgery. Both are intrinsically eligible candidate referents of your 'I'-beliefs, with mind-independent persistence conditions. But it depends on your 'I'-beliefs which one counts as *you*.

I said that Thinker and Breather were *intrinsically eligible* candidate referents. This is important because in itself, DSV isn't a full-fledged theory of personal identity. Rather, it's an account of which of various intrinsically eligible referents we refer to when we think 'I', combinable with all sorts of restrictions on what is eligible. A limiting case is the view that there is exactly one such thing; DSV would then have no work to do at all. This approach is not especially attractive: intuitively your 'I'-beliefs determinately refer to one thing, but it's far less obvious that they have only one *candidate referent*. Surely my upper body is intrinsically eligible to be a referent of 'I' (after all, I could be pared down to its size). Another extreme is an 'anything goes' approach with no constraint on the intrinsically eligible candidate referents of the first-person pronoun. I prefer a middle-of-the-road view: anything with not more than a certain amount of spatiotemporal or psychological discontinuity that has a cerebrum at some time, t, is a candidate referent of the 'I'-beliefs produced in that cerebrum at t. I won't argue for drawing the line here; readers with different leanings can accept appropriately strengthened or weakened versions of the conclusions I will draw.

---

[5] Olson [1997]

[6] Parfit [1984]

In developing DSV it will be convenient to presuppose a plenitude ontology, according to which for any distribution of matter over space-time there is an object made up exactly of those portions of matter at those times. (I should stress though that much of what I will say in the sections to follow are defensible against the backdrop of sparser ontologies. For example, I will argue that our beliefs about ourselves settle whether, when using 'I', we refer to biological or psychological continuers (sections 3.1, 4). This presupposes that there are both biological and psychological continuers, but this much is plausible even on non-plenitude ontologies: constitution theorists accept that there are both persons with psychological and human animals with biological persistence conditions[7], and even animalists are under pressure to concede that *cerebra* could persist through the death of their host animals.[8])

The most familiar plenitude ontology is Liberal Perdurantism (simply 'perdurantism' henceforth), which makes three claims: (1) material objects persist over time by having instantaneous temporal parts at different times, (2) for any instantaneous objects at any times, some temporally extended object is composed by exactly those objects at those times, (3) proper names and pronouns typically refer to such temporally extended objects. On the perdurantist picture, ordinary objects are four-dimensional 'worms': when the person typing this article and the person waking up in my bed tomorrow think 'I', they refer to a four-dimensional worm that has both the typing-slice and the waking-slice as temporal parts. Note the qualification 'typically' in (3). Although perdurantism construes persisting things as four-dimensional worms, I don't wish to rule out the possibility that instantaneous stages may be eligible referents of 'I'. The important point is that *non*-instantaneous things persist

---

[7] See, e.g., Baker [2000].

[8] See Olson [1997: 9–11].

by having instantaneous temporally parts, rather than by bearing temporal counterpart relations to future instantaneous things (see below).[9]

Other plenitude ontologies would also furnish DSV theorists with the objects needed to serve as candidate referents. One is stage theory, which accepts (1) and typically (2) but adopts a different semantics for referring expressions, according to which these refer to instantaneous stages of four-dimensional worms rather than the worms themselves.[10] Accordingly, persistence is analyzed in terms of temporal counterpart relations between stages. There are also plenitude three-dimensionalist views that dispense with temporal parts but mimic other features of perdurantism.[11] DSV could be easily combined with these views as well; what matters is that we overlap with a great number of 'temporal overlappers'[12] that are intrinsically eligible to be the referents of our 'I'-beliefs. For simplicity's sake, I will henceforth write as a perdurantist.

We can now sharpen up our initial first-pass statement of DSV. My official formulation is modeled on the synchronic self-making view and goes as follows:

> (*Diachronic self-making*) Of the four-dimensional worms that are candidate referents of your 'I'-beliefs, you are one of the non-accidentally best satisfiers of those 'I'-beliefs.[13]

---

[9] Perdurantists include Lewis [1976] and Hudson [2001].

[10] Stage theorists include Hawley [2001] and Sider [2001].

[11] See Miller [2006].

[12] This terminology is inspired by Madden [2016].

[13] Cf. Kovacs [2016: 1080]. The discussion of the next few paragraphs closely follows that of Kovacs [2016: 1077–1080].

Some clarification is in order here. First, DSV doesn't imply a descriptivist view of indexicals. Perry [1979] and others argued that certain pieces of information were essentially indexical: for example, you can know everything about the person making a mess in the supermarket without thereby knowing that *you* are that person. DSV doesn't attempt to overturn this insight: it takes for granted that anything eligible for having your 'I'-beliefs at a time has those of your parts that sustain mentality at that time. (This doesn't mean that you always had or will always have a cerebrum; see section 3.1.)

Second, the expression 'your 'I'-beliefs' might appear to suggest that your overlappers have no 'I'-beliefs. Not so. You and your overlappers share numerically the same 'I'-beliefs, and when I say 'share', I mean that they *have* those beliefs in the same sense you do. So, unlike some other philosophers I don't distinguish between having and merely hosting a mental state.[14] Rather, your overlappers have all your *de se* beliefs, but from their point of view these aren't *de se*: they are beliefs about you rather than themselves. So, I accept a version of Noonan's [1998] 'Personal Pronoun Revisionism' (PPR). According to Noonan, the 'I'-beliefs of human animals refer to the persons they coincide with rather than to themselves. Similarly, I hold that your temporal overlappers share your 'I'-beliefs but that these beliefs are self-referential only from your perspective. When a temporal overlapper that came into existence on your twentieth birthday thinks, *I was once a toddler*, it thinks something true about you, rather than something false about itself.[15]

---

14 Cf. Burke [2003] for the 'having' vs. 'hosting' distinction.

15 See Noonan [1998]. One question that naturally arises about PPR is what prevents my overlappers from having self-directed 'I'-beliefs. The (short) answer is that the assumptions that lead to overlapping subjects don't automatically lead to overlapping mental states. We can derive that my temporal segment that will cease

Third, of the various intrinsically eligible four-dimensional objects, the *best* candidate referent is the one that best satisfies (makes true) the total set of 'I'-beliefs you have right now. Beliefs likely to distinguish better and worse candidate referents include past-, present- and future-directed 'I'-beliefs (*I was once a fetus*; *I'm less than six feet tall*; *I will go on holiday next week*, etc.) and counterfactual beliefs (*I wouldn't survive total amnesia*; *I would survive the gradual replacement of my brain cells with silicon chips*; etc.). While most of us don't have beliefs directly about the counterfactual scenarios that dominate the personal identity literature, we have beliefs about matters bearing on them. For example, if you dread death but think that nothing bad happens to the characters of Star Trek when they teletransport, an object that would survive teletransportation better satisfies your 'I'-beliefs than an object that wouldn't.[16]

---

to exist half an hour before me has 'I'-beliefs at some time, t, from the assumptions that (1) a human being in my place has 'I'-beliefs at t, that (2) this human being could perish half an hour earlier than it actually will, and that (3) something's capacity to have 'I'-beliefs doesn't depend on future events. However, these assumptions don't imply that my temporal overlappers have 'I'-beliefs qualitatively similar to but numerically different from mine. But if they don't, then my token belief state *I am F* at some time t is numerically identical to my overlappers' token belief state *I am F* at t. Since these beliefs are one and the same, 'they' (it) cannot have numerically distinct referents. See also Kovacs [2016: 1087–1091] for a more detailed discussion of (and responses to) various other objections to PPR.

[16] One interesting question is whether the view could be formulated in terms of credences instead of beliefs. There are a couple of options here. The simpler one is to draw a threshold above which some level of credence in p counts as a belief that p. Alternatively, one might try to bypass matching credences to beliefs. One way of doing so would start by assigning a credence to every proposition p_i (as well as its negation) that S has entertained. Next, take every probabilistically coherent maximal conjunction that, for each p_i, contains either p_i or its negation; if there's none, find the conjunction that conjoins the largest probabilistically coherent subset. The referent of one's 'I'-beliefs, then, is the best non-accidental satisfier of the conjunction with the highest

Fourth, DSV wasn't meant to eliminate all referential indeterminacy. The Problem of the Many raises a general puzzle about boundaries, and on the assumption that we are material objects this puzzle extends to us.[17] The claim that you are *one of the* best satisfiers of your 'I'-beliefs is a reminder that DSV's job is not to eliminate all indeterminacy in the reference of 'I'. But to enhance readability, I will put indeterminacy to the side and often speak of 'the' best candidate referent.

Fifth, an intrinsically eligible candidate *non-accidentally* satisfies an 'I'-belief of yours iff it satisfies that 'I'-belief, and the 'I'-belief is neither based on misleading empirical evidence nor the result of an inappropriate response to your empirical evidence.[18] I cannot define 'empirical evidence', but it's meant to include all perceptual, memory-based and proprioceptive evidence that is neutral about which overlapper you are. It's meant to *exclude* evidence based on philosophical reasoning. Let me illustrate what I have in mind with some examples. (1) Your present temporal stage is spatiotemporally and psychologically continuous with a toddler stage that existed forty years ago. You think you have been around for at least forty years, but you do so on the basis of subsequently implanted fake memories that happen to match the way things were with that toddler. In this case, a more-than-forty-year-old worm satisfies your 'I'-belief, but only accidentally, since your fake memories only give you misleading empirical evidence that you are more than forty years old. (2) Your apparent memories were produced in the right way, but you are paranoid and think they

---

assigned credence. (Thanks to an anonymous referee for asking what a credence-based version of DSV would look like.)

[17] See Hudson [2001: ch. 1] for a survey.

[18] This is a slight departure from my earlier formulation [Kovacs 2016: 1079–1080]. See also Madden [2011] for a different approach, which ties the reference of 'I' to the best *knowledge*-maximizing assignment.

were implanted by a mad scientist. So, you form the belief that you came into existence twenty years ago. In this case, there is indeed a twenty-year-old four-dimensional worm that is the best satisfier of your belief. But it fails to be a *non-accidental* satisfier, since you responded to your evidence inappropriately. (3) In a similar situation, you believe that you came into existence only twenty years ago. You don't have any strange belief about how you acquired your apparent memories about the times from before; you think they were produced in the regular way. You just think the worm that used to be a toddler is merely spatiotemporally and psychologically continuous, but not identical, to you, and ceased to exist twenty years ago. Here, the twenty-year-old worm *is* non-accidentally the best candidate referent of your 'I'-belief. Of course, your response to your evidence seems inappropriate by ordinary standards (how can you simply ignore your apparent memories?), but you are not responding badly to your strictly *empirical* evidence. One moral of the overpopulation puzzles is that your empirical evidence doesn't distinguish you from your temporal overlappers (everything seems to them exactly the way it seems to you). So long as we agree on a metaphysically neutral description of your last twenty years and the period preceding them, believing that you are the non-identical continuer of a toddler is no worse than believing that you were that toddler. So DSV does imply that in the third case, you are a proper temporal part of what most would consider a human person. (Naturally, I will have more to say about this puzzling upshot in the second half of the paper.)

Hopefully, I have said enough to give a reasonably clear sense of what DSV amounts to. Why believe it? My reasons are analogous to the ones I offered in defense of the synchronic self-making view [Kovacs 2016: 1080–5], so I will be brief here. First, plenitude ontologists have special reasons to like DSV. These views suffer from a familiar 'overpopulation problem': every moment you share your place with countless overlapping objects. What

makes it the case, then, and how can you know, that – for example – you were once a child, given that you and your 'child-excluding' temporal parts all seem to have the same body of evidence?[19] DSV suggests a simple answer. The best candidate referent of your 'I'-beliefs *was* a child. Your childhood-excluding temporal overlappers share the belief *I was once a child*, but they too are right, since from their perspective these are not self-referential beliefs; they are beliefs about *you*. This is why the mechanics of self-reference not only guarantees that you were once a child but also allows you to know that you were. (Of course, if you thought – as I'm assuming you don't – that you were never a child or even a 20-year-old, that would be true and known by you as well, and for the same reason.)

Second, as Zimmerman [2013: 119–21] and I [2016: 180–2] both pointed out, there is an attractive analogy between 'I' on the one hand and 'here' and 'now' on the other. The latter two are 'impure indexicals' [Recanati 2001]: they come with contextually built-in restrictions on their possible referents but don't automatically pick out a determinate referent. Referential intentions plausibly play a significant role in determining the reference of impure indexicals. For example, when it rains on and off for an entire day, saying 'It's raining now' will be true even if the 2-second long duration of the utterance is rain-free, since the entire day is among the times eligible for me to think of as 'now'. But 'I' resembles 'now' in this regard: there are a host of intrinsically eligible candidate referents for my 'I'-beliefs. Accordingly, our referential intentions plausibly play a role in determining the reference of our 'I'-beliefs.

Third, considerations about constitutive rationality also support DSV. Lewis [1974, 1994] argued that it's a basic principle of folk psychology that a subject's attitudes rationalize

---

[19] See Olson [1997: 166–8], Hudson [2001: chs. 4 and 6], Zimmerman [2003: 501–3], Kovacs [2010] and Sutton [2014] for discussion.

her behavior. We tend to form beliefs that fit our evidence and cause behavior that serves our goals and should also assign beliefs to other agents according to these principles. DSV is a natural implementation of this idea: my 'I'-beliefs' best referent is the candidate whose behavior they best rationalize. For example, if you anticipate the pain that anything with your cerebrum would experience but not the pain that anything with your body would, the candidate that best rationalizes your beliefs is something that goes where your cerebrum goes.

The argument from constitutive rationality doesn't assume that persons are perfectly rational or that they cannot be mistaken about their beliefs.[20] It is commonplace that ordinary people's belief systems contain a number of hidden inconsistencies. In these cases the referent of some 'I'-beliefs is determined by which candidate best satisfies their largest coherent subset, where coherence is settled by internal consistency as well as how central the beliefs in question are ('I'-beliefs on which numerous other beliefs depend weigh more than peripheral beliefs the subject would be ready to revise without much reluctance). Nor does DSV deny that human persons are subject to cognitive illusions and biases and that we lack perfect access to our own 'I'-beliefs. What the view does assume is that we are *usually* rational; our irrationality is not so massive and systematic as to be incompatible with making sense of our behavior in terms of our folk 'theory of persons' [Lewis 1974].

To illustrate this point, take a person S_1 at t_1 who committed a horrible crime and is psychologically and biologically continuous with S_2 at time t_2.[21] The crime's gravity weighed so heavily on S_1's shoulders that by t_2 it has become unbearable for S_2, who has come to sincerely believe that *he* has never committed it. Does DSV recommend the verdict that S_2

---

[20] Thanks to an anonymous referee for requesting clarification here.

[21] Thanks to an anonymous referee for asking me about this case.

has no criminal temporal parts? It depends. If most of S_2's other beliefs best rationalize the behavior of someone who did commit the crime (he is hiding from the police because he believes they have reason to be after him; he has an urge to make up for the victims because he believes he somehow owes them; etc.), his 'I'-beliefs' best candidate referent is someone who committed the crime but believes that he didn't. If, by contrast, S_2's disownment of his apparent past is so complete that the best interpretation of him as a person doesn't ascribe to him a t_1 temporal part (for example, he goes to the police to report on the crime as a witness with 'quasi-memories' of the crime), then DSV may deliver the result that S_2 and S_1 are distinct persons. Such complete disownment of one's past may seem far-fetched, of course, because it *is*. But this just goes to show that even on DSV, self-deception will be more common than psychologically self-induced disruptions of personal identity. I take this to be a good result.

This concludes my positive case for DSV. Now is a good time to also say something about DSV's compatibility with the idea that generally, reference determination is a public matter. The issue arises because one might think that the arguments from impure indexicals and constitutive rationality could be turned on their heads and used to support conclusions contrary to DSV. Suppose I intend to use 'I' for an intrinsically eligible candidate, but everyone talking to me is using 'you' for another. Doesn't an analogy with 'there' imply that the people around me are referring to an object different from the one I believe myself to be? And don't considerations of constitutive rationality favor interpreting *them*, who vastly outnumber me, as being right? I don't think so, because there is an important asymmetry between 'I' on the one hand, and 'you', 'he' and 'she' on the other. When we use second- and third-person pronouns, it's part of our referential intention to use them for something that is disposed to use the first-person pronoun self-referentially; we aim to zone in on the centers

of the first-person attitudes entertained in their immediate vicinity. This principle of deference might, of course, clash with deep-seated views about persons. But these views aren't meaning-constitutive (it's no part of the semantics of 'I' or 'you' that they can only refer to beings with certain parts), so we shouldn't assume that they trump the principle of deference.[22]

By contrast, our use of first-person pronouns shows no such deference; it's no part of our referential intention to use them for what *others* use their second- and third-person pronouns for. Imagine that an entity you've been addressing with 'you' is the exoskeleton of an insect-sized alien controlling it from within. Plausibly, once you learn this you have a semantic obligation to start using 'you' for the alien inside. The alien on the other hand doesn't have to adjust to your use. She won't acquire any obligation to start using 'I' for her exoskeleton even if all humans she encounters use 'you' for it.

Admittedly, none of the three arguments for DSV is knockdown, and each depends on controversial assumptions. All I have been hoping to show is that DSV is not without motivation. Moreover, although it's a version of private conventionalism, it can accommodate the plausible principle that reference determination has an important communal component. It's time to turn to potential applications of DSV.

3. DSV at work

In this section, I will provide a brief overview of how DSV can handle some familiar cases from the personal identity literature. This will also clear the ground for a problem I will discuss in section 4.

---

[22] See Kovacs [2016: 1093–4] for a similar reasoning; see also Zimmerman [2013: 118–9] for the relation between individual referential intentions and deference to authority.

3.1. Fetuses

Have we ever been fetuses?[23] According to DSV, there is no general answer to this question. If you think you started off as a fetus, then other things equal the best candidate referent of your 'I'-belief is something that was once a fetus; if you think you didn't, it's something that wasn't. Either way, all the fetus-involving and fetus-excluding worms you overlap with refer to the same thing when they say 'I was [never] a fetus'.

Needless to say, fetuses don't have beliefs, let alone 'I'-beliefs. That's fine. DSV doesn't require that we have 'I'-beliefs at every moment of our existence; it says only that when we do, we pick out their best candidate referent. Of course, we can still ask third-person variants of these questions: under what conditions will *this entity* (presently a fetus or a small child) be around? Absent first-person beliefs, we might fall back on other linguistic conventions to answer such questions. For example, 'this toddler' arguably refers to something that (absent some unexpected tragedy) will survive toddlerhood, since virtually all of us use the expression for such a thing. The reference of 'this fetus', on the other hand, might be indeterminate, since the language users in our community are roughly evenly divided between those who think that fetuses persist as full-blown persons and those who don't. Perhaps we simply refer to different things when talking about fetuses: some of us to things with both fetus-like and adult-like temporal parts and some of us to things with only fetus-like temporal parts. Similar remarks apply to proper names given to newborns who will never develop a first-person perspective: since human beings without a first-person perspective lack 'I'-beliefs, the reference of such names may well be determined by public use. None of this is inconsistent with DSV, which implies only that your 'I'-beliefs (if you

---

[23] See Olson [1997: ch. 4] and S. Shoemaker [2008] for opposing answers.

have any) play a key role in reference determination, not that proper names and non-first-person indexicals fail to refer in the absence of 'I'-beliefs.

3.2. *Fission*

DSV implies that in most fission cases, we determinately go where one particular offshoot goes. Suppose you are about to divide into two offshoots, which will have different lives. Imagine, for example, that the fission occurs during your studies in medical school, which only one offshoot finishes – the other drops out and tries his luck in the art world (sadly, with little success). Which one is you? Suppose all other things are equal: your belief that you will study in med school is not the result of brainwashing, you didn't subconsciously want to be an artist, and so on (recall how we qualified the argument from constitutive rationality in section 2). In that case, I say you are the offshoot that finishes med school. Given your present beliefs, the post-fission doctor is a better non-accidental candidate referent of your 'I'-beliefs than the maverick artist.

Importantly, what matters here isn't that you *wanted* to become a doctor but that your plans to become one were serious enough to warrant the counterfactual belief *If I continued like this, I would become a doctor.* If you had merely been daydreaming about becoming a doctor, your belief that you would one day become one would be an inappropriate response to your empirical evidence. Moreover, although in this particular case you are psychologically continuous with the doctor offshoot, this is not the reason *why* you persist as that offshoot. (Suppose that nearing the age of 140, you are about to undergo fission. One of the offshoots will undergo a series of organ transplants to be able to live for another few decades; the other will soon fall into a persistent vegetative state, during which it will be studied by a medical team dedicated to figuring out what makes certain humans live such a long life. It

15

has been your plan for quite some time to allow this medical team to perform experiments on you for the benefit of the human race. In this case, DSV predicts that you persist as the vegetable offshoot, despite bearing no psychological links to her.)

The med school case brings out an important feature of DSV. To persist as either offshoot, you don't need to believe that you will be that offshoot; you don't even have to believe that fission will take place. All you need to have is a set of future-oriented or counterfactual beliefs about yourself that non-accidentally fit one of the offshoots better than the other. Perfect tie between the two offshoots will be rare, though not impossible. In those cases, it's indeterminate which offshoot you are. This is no great cost. Perfectly symmetric fission is everyone's problem, and according to DSV, even very fine-grained differences can break the tie between almost symmetric cases. The standard perdurantist solution treats *all* future-directed *de se* judgments as indeterminate in fission cases, even blatantly asymmetrical ones.[24] DSV eliminates indeterminacy in most cases that the standard perdurantist views would treat as symmetric.[25]

---

[24] Lewis [1976: 28–9]; cf. Sider [2018].

[25] This advantage might be shared by certain psychological approaches to personal identity, depending on how fine-grained the criterion is that they use to break ties between intrinsically eligible continuers. It seems to me, however, that many pairs of candidates that bear equally strong links of psychological continuity and connectedness (see Parfit [1984: 205] for these notions) to a pre-fission person will nonetheless differ in how many of the pre-fission person's future-oriented 'I'-beliefs they satisfy. On the face of it, the falsity of no particular future-oriented 'I'-belief I hold now weakens the psychological links I bear to a future person; however, even the most minor 'I'-belief can weaken a future person's claim to be a satisfier of my future-oriented 'I'-beliefs. (Thanks to an anonymous referee for asking how the traditional psychological approach fares against DSV with respect to fission.)

3.3. Teletransportation

You are about to enter the teletransporter. The machine records all your physical and mental features and instantly annihilates your body; the next moment someone exactly with your appearance and psychological profile materializes at the other side and seemingly remembers entering just a moment ago. Is this person you?

The answer depends on what you think will happen before you enter. If you think you will travel, then this is what will happen: the best candidate referent of your 'I'-beliefs will be something with a pre-teletransportation and a post-teletransportation temporal part. But if you think you will die, the best candidate referent of your 'I'-beliefs will be something the machine annihilated. Similarly to the previous two cases, the mechanism of self-reference turns your attitude to teletransportation into a kind of self-fulfilling prophecy.

4. The problem of inconstant 'I'-beliefs

Throughout the foregoing sections, I have been putting to the side an issue that by now has likely occurred to you. DSV deals nicely with the familiar puzzle cases so long as the various candidates' 'I'-beliefs don't clash. For example, things go smoothly in the teletransportation case if before entering the machine you expect to survive and the person leaving the teletransporter takes herself to be identical to the person who entered. But what if you think that you won't survive, while the person stepping out thinks she herself entered? Or what if both your doctor offshoot and your artist offshoot think they were *you*?

I will refer to such puzzle cases as instances of *the Problem of Inconstant 'I'-beliefs*. It's tricky to formulate the problem without begging some questions, but the general challenge is this. According to DSV, your 'I'-beliefs refer to their best candidate referent four-dimensional object. Which object this is in part depends on what you believe about your persistence

conditions and matters bearing on them. But surely these beliefs can change over time. And when they do, they seem to give us conflicting advice about how to assign a referent to your 'I'-beliefs.

Stage theorists have a unique way of dealing with this problem because they analyze persistence in terms of temporal counterpart relations, which can hold asymmetrically. For example, Sider [2018] argues that if Daytime Mitty lives the life of a superhero (Nighttime Mitty) every night, and Nighttime Mitty remembers the life of Daytime Mitty but Daytime Mitty neither anticipates nor remembers any of Nighttime Mitty's heroic deeds, then Daytime Mitty is the same person as Nighttime Mitty but Nighttime Mitty isn't the same person as Daytime Mitty. Were I a stage theorist, I could take a similar approach to the Problem of Inconstant 'I'-beliefs and say that if you expect to survive teletransportation but your post-teletransportation continuer doesn't take herself to have been around before the event, then you are the same person as this continuer but your continuer isn't the same person as you. While this option is intriguing, I won't explore it here; I want to see how far we can get with the more conservative resources of perdurantism.

I will contend that in apparent cases of belief change, the pre-event and post-event 'I'-beliefs are both correct but refer to distinct entities. This has a surprising consequence: it's possible for a perduring thinker to have self-referential 'I'-beliefs at some part of her career and non-self-referential ones at other times.[26] I will explain how the strategy works through the example of teletransportation; I will then briefly indicate how it generalizes to other cases.

---

[26] Braddon-Mitchell and West [2001] also allow for reference change in a series of *de se* beliefs that would normally be considered as belonging to one person, but they think that in these cases the reference of 'I' is indeterminate.

Suppose you enter the teletransporter thinking you won't survive it. Afterwards, your psychological continuer believes herself to have entered the machine. Which one of you is correct? Both of you are. At time t_1, you enter the machine thinking, *I won't survive this*. This is true, since the best candidate referent of your 'I'-beliefs at that time perishes shortly after you enter the machine (call it Perisher). At time t_2, the person emerging at the other side thinks, *Phew! I made it!* She is right, too: her 'I'-beliefs refer to something with both a pre-teletransportation and a post-teletransportation temporal part (call it Survivor).

At least two details in this story require comment. One is that whether your post-transportation psychological continuer is right that she herself entered the teletransporter in part depends on her beliefs. This might seem odd: it looks as if whether I was around at some past time t_1 depends on my beliefs at some later time, t_2. How is this possible without some spooky kind of backward causation? To make the case more dramatic, suppose that after your psychological continuer steps out of the teletransporter, she is unsure for a while about whether she herself earlier entered.[27] When thinking about such cases, we should constantly remind ourselves that DSV is a conventionalist view only in the sense that the reference of 'I' is to a large extent up to the subject, but not in the stronger sense of making it somehow a mind-dependent matter which objects have temporal parts at which times. In the aforementioned story, for example, there is no perduring object such that whether *that* object has a temporal part at t_1 depends on the 'I'-beliefs produced in that object's brain at some later time t_2. According to the perdurantist view I adopted at the outset, Perisher, Survivor and the post-teletransportation temporal part of Survivor (call it Reborn) all exist independently of anyone's 'I'-beliefs. Whether at t_2 your post-teletransportation continuer

---

[27] Thanks to an anonymous referee for raising this concern, and for offering this variation on the teletransportation case.

will form the belief *I was around at t_1* doesn't determine whether any of these objects have temporal parts at t_1: by t_2 it's already settled that Survivor and Perisher do while Reborn doesn't. The belief will only determine *which of these objects the 'I'-beliefs will refer to*; but of course, the determination of *this* fact at t_2 doesn't imply any kind of backward causation.[28]

Another issue is whether interpreting Survivor's t_2 belief as true forces us to infer that many of her past beliefs were false. If at t_2 Survivor is right that she earlier stepped into the teletransporter, didn't she mistakenly think at t_1 that she wouldn't survive teletransportation? The answer is 'No': although at t_1 Survivor shared your belief, *I won't survive this*, this wasn't a *de se* belief from Survivor's point of view. It was a belief about you, Perisher. And, of course, at t_1 Survivor was right to believe that *you*, Perisher, wouldn't survive teletransportation. There is one thing Survivor is admittedly wrong about: Survivor is wrong to think, at t_2, that at t_1 she mistakenly thought that she wouldn't survive teletransportation. To be more precise, at t_2 Survivor would be correct if she uttered the following direct quotational sentence: 'At t_1, I thought: "I wouldn't survive teletransportation"'. But she would be mistaken to utter the following *indirect* quotational sentence: 'At t_1, I thought I wouldn't survive teletransportation'. Survivor is wrong about the second sentence's truth value because she is unaware of a reference shift that took place between t_1 and t_2: unbeknownst to her, her past stages used 'I'

---

[28] A related question concerns the role of future facts in self-reference at the present. Suppose I have a large temporal part, David-minus, which will cease to exist half an hour before me, David – however, there's a possible world in which I die half an hour earlier. Doesn't it depend on future facts whether my present use of 'I' refers to David-minus or David? It does in the following sense: reference is a relation between a linguistic expression and a referent, and on the perdurantist view certain (future temporal) parts of the putative referent aren't present yet. Note, however, that there is nothing special about *de se* reference here: to the extent that this is an odd result, it also affects the reference of proper names, definite descriptions and other linguistic expressions. (Thanks, again, to an anonymous referee for requesting more clarity on this.)

with a different referent than her present stages. Perhaps it's odd, at first, to think that Survivor is making this mistake. But then again it was also surprising to learn that the reference of 'I' can shift like this in the first place, and once we remind ourselves of the unusual nature of teletransportation, the mistake starts looking more understandable.

One might think this verdict absurdly implies that Survivor was also wrong at t_2 about whether she was a person at t_1. However, this doesn't follow. While x's ability to have self-referential 'I'-beliefs is *sufficient* for x to be a person, I don't think it's also necessary. Patients in a reversible coma, who later wake up, are unable to have 'I'-beliefs while they are in coma; yet they are plausibly persons. On my view, other people are prevented from having self-referential 'I'-beliefs by overlapping with people that are better referents for those beliefs.

But doesn't this have the disastrous implication that there could be more than one *person* (not just thinker) sitting in your chair, thinking your beliefs? And aren't some of the most influential arguments in the recent personal identity literature based on the assumption that this result would be absurd enough to serve as a *reductio* of any view that entails it? I respond by pointing out that the literature routinely conflates two distinct problems. The first is the Problem of Almost-Persons: what prevents my overlappers from being persons, and how can I know I'm a person rather than a mere overlapper? The second is the Problem of Overlappers: what prevents my overlappers from being the referents of my 'I'-beliefs, and how can I know that my 'I'-beliefs typically ascribe true things to something with the right spatial and temporal parts (me), rather than an overlapper with the wrong parts?[29] The deeper problem is the second one. The primary motivation for denying that our overlappers are persons is the hope that this would help tell ourselves apart from them. So, the Problem of Almost-Persons borrows its importance from the assumption that if we can't solve it, we

---

[29] See Kovacs [2016: 1071–2] for this distinction.

can't solve the Problem of Overlappers either. But DSV gives us ample reason to drop this assumption: even if my overlappers are persons, I can distinguish myself from them by reflecting on my beliefs and noticing that they better fit the thing I think I am than my overlappers. I know who I am even if I can't be sure I'm the only person around.

Other 'change of heart' cases receive essentially the same treatment. Suppose that before your cerebrum is removed (at t_1), you think you will survive the surgery as Breather. At t_2, your cerebrum thinks: *contrary to my t_1 belief, I was around at t_1 and survived the surgery as a cerebrum.* In this case, too, both you at t_1 and your cerebrum at t_2 are right; you are just referring to different four-dimensional objects. However, similarly to the previous case, Thinker is mistaken to believe that she wrongly thought at t_1 that she would survive the surgery as Breather, since she didn't notice the subtle reference shift that took place between t_1 and t_2.

Or suppose that before fission you expect to finish med school, but after fission the lucky offshoot takes herself not to have existed pre-fission, while the unlucky one believes that her plans to become a doctor have been frustrated. As in the earlier cases, all these beliefs are true; they just refer to different four-dimensional objects. The only thing the unlucky offshoot is wrong about is the referent of her past belief, *I will finish med school*, which was not about herself but about a four-dimensional object that *did* finish med school.

5. Conclusion

In this paper I defended DSV, a version of weak private conventionalism about personal identity over time. Plenitude ontologists have several reasons to like this account. For one, DSV solves a recalcitrant problem for plenitude views: how can we tell ourselves apart from the countless temporal overlappers we share our place with? For another, it's independently

motivated by considerations pertaining to constitutive rationality and indexical reference. Moreover, the synchronic and the diachronic versions of the self-making view together make for an attractive unified treatment of personal identity at and over time.

Many outstanding issues remain to be solved. One of the more serious ones is Olson's 'Problem of Subpeople'.[30] If we overlap with other thinking beings, don't we have moral obligations to them that our standing ethical practices obviously violate? For example, by performing some thankless, boring task, aren't you literally *ruining the life* of overlappers whose entire existence is consumed by the chore and won't be around to reap the benefits? More generally, proponents of DSV are likely going to be under pressure to adopt revisionary views about the relationship between personal identity, moral responsibility and prudential concern.[31]

These issues are worth taking seriously, and I don't want to play down the challenge they pose to DSV. Unfortunately, I have to postpone their treatment to another occasion. For now, I'm happy to have put the view to use in addressing the formidable metaphysical and epistemological problems that beset plenitude ontologies of persons.[32]

---

[30] See Olson [2010]. Johnston [2017] refers to subpeople as 'personites'

[31] The view that personal identity is neither necessary nor sufficient for moral responsibility is not new with DSV; it's one aspect of Parfit's [1984] general view that 'identity is not what matters' and has more recently also been defended by D. Shoemaker [2012] and Khoury and Matheson [2018].

[32] For helpful comments on and discussions about earlier drafts of this paper I'm grateful to Matt Duncan, Matti Eklund, Rory Madden, Gideon Rosen, Cathy Sutton, anonymous referees, and audiences at the following conferences: 'Persons, Selves, and Organisms' (at Bilkent University), 'Self: Between Consciousness and Non-being' (at Hokkaido University), and the 2017 Paros Workshop on Personal Identity (at the Aegean Center for the Fine Arts).

References

Baker, Lynne R. 2000. *Persons and Bodies: A Constitution View*, Cambridge: Cambridge University Press.

Braddon-Mitchell, David and Caroline West 2001. Temporal Phase Pluralism, *Philosophy and Phenomenological Research* 62/1: 59–83.

Braddon-Mitchell, David and Kristie Miller 2004. How to Be a Conventional Person, *The Monist* 87/4: 457–74.

Burke, Michael B. 2003. Is My Head a Person?, in *On Human Persons*, ed. Klaus Petrus, Frankfurt: Ontos Verlag, 107–25.

Eklund, Matti 2004. Personal Identity, Concerns, and Indeterminacy, *The Monist* 87/4: 489–511.

Hawley, Katherine 2001. *How Things Persist*, Oxford: Clarendon Press.

Hudson, Hud 2001. *A Materialist Metaphysics of the Human Person*, Ithaca: Cornell University Press.

Johnston, Mark 1989. Relativism and the Self, in *Relativism: Interpretation and Confrontation*, ed. Michael Krausz, Notre Dame, IN: University of Notre Dame Press, 441–72.

Johnston, Mark 2017. The Personite Problem: Should Practical Reason Be Tabled?, *Noûs* 51/3: 617–44.

Khoury, Andrew C. and Benjamin Matheson 2018. Is Blameworthiness Forever?, *Journal of the American Philosophical Association* 4/2: 204–24.

Kovacs, David Mark 2010. Is There a Conservative Solution to the Many Thinkers Problem?, *Ratio* 23/3: 275–90.

Kovacs, David Mark 2016. Self-made People, *Mind* 125/500: 1071–99.

Lewis, David K. 1974. Radical Interpretation, *Synthese* 23/3–4: 331–44.

Lewis, David K. 1976. Survival and Identity, in *The Identities of Persons*, ed. Amelie O. Rorty,
Berkeley and LA: University of California Press, 17–40.

Lewis, David K. 1994. Reduction of Mind, in *A Companion to the Philosophy of Mind*, ed.
Samuel Guttenplan, Oxford: Blackwell, 412–31.

Madden, Rory 2011. Externalism and Brain Transplants, in *Oxford Studies in Metaphysics
Volume* 6, ed. Karen Bennett and Dean Zimmerman, Oxford: Oxford University Press,
287–316.

Madden, Rory 2016. Thinking Parts, in *Animalism: News Essays on Persons, Animals, and Identity*,
ed. Stephen Blatti and Paul F. Snowdon, New York: Oxford University Press, 180–207.

Miller, Kristie 2006. *Issues in Theoretical Diversity: Persistence, Composition, and Time*. Dordrecht:
Springer.

Miller, Kristie 2013. "Personal Identity" Minus the Persons, *Philosophical Studies* 166 (S1): 91–
109.

Noonan, Harold W. 1998. Animalism *Versus* Lockeanism: A Current Controversy, *The
Philosophical Quarterly* 48/102: 302–18.

Olson, Eric T. 1997. *The Human Animal: Personal Identity Without Psychology*. New York:
Oxford University Press.

Olson, Eric T. 2010. Ethics and the Generous Ontology, *Theoretical Medicine and Bioethics*
31/4: 259–70.

Parfit, Derek 1984. *Reasons and Persons*. Oxford: Clarendon Press.

Perry, John 1979. The Problem of the Essential Indexical, *Noûs* 13/1: 3–21.

Recanati, François 2001. Are 'Here' and 'Now' Indexicals?, *Texte* 127/8: 115–27.

Shoemaker, David 2012. Responsibility Without Identity, *The Harvard Review of Philosophy*
18/1: 109–32.

Shoemaker, Sydney 2008. Persons, Animals, and Identity, *Synthese* 162/3: 313–24.

Sidelle, Alan 1999. On the Prospects for a Theory of Personal Identity, *Philosophical Topics* 26/1–2: 351–72.

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

Sider, Theodore 2018. Asymmetric Personal Identity. *Journal of the American Philosophical Association* 4/2: 127–46.

Sutton, C.S. 2014. The Supervenience Solution to the Too-many-thinkers Problem. *The Philosophical Quarterly* 64/257: 619–39.

Zimmerman, Dean W. (2003). Material People, in *The Oxford Handbook of Metaphysics*, ed. Michael J. Loux and Dean W. Zimmerman, New York: Oxford University Press: 491–526.

Zimmerman, Dean W. 2013. Personal Identity and the Survival of Death, in *The Oxford Handbook of Philosophy of Death*, ed. Ben Bradley, Fred Feldman, Jens Johansson, New York: Oxford University Press: 97–154.