

Penultimate draft of Sept 2021—please cite/quote from the published version.

I Think; Therefore, I am a Fiction¹

T. Parent (Nazarbayev University)

nontology@gmail.com

1. Introduction

We are all familiar with Descartes' *Cogito* argument for the existence of the thinking self. It runs something like this:

(Cog) I think.

(*) So, I exist as a thinking thing.

Even Descartes was aware that (Cog) seemed question-begging re: the existence of "I" as the thing that thinks. Regardless, the argument strikes many as compelling. (Descartes thought there was some sort of direct intuition of the premise.) However, if the argument is circular, then it seems just as legitimate to run the *Cogito* in reverse, as follows:

(*) I exist as a thinking thing.

(Cog) So, I think.

Is the reverse *Cogito* at all persuasive? Does the Cartesian "natural light" make the premise attractive? A *mental fictionalist* ought to be concerned. For if (*) is primitively compelling, then realism about thought follows—and mental fictionalism is false.² That is so, regardless of whether 'mental fictionalism' is defined as a type of "atheistic" eliminativism, or just as a more modest, "agnostic" view about the mental.³

In what follows, I shall explain why I do not find (*) compelling, and why you should not either. I will try to persuade you that the self is plausibly regarded as a fiction as well, although fictionalism about the self is here merely the "agnostic" and not the "atheistic" type. (Even so, I sometimes give the impression I am an atheist about the self. But while my doubts are strong, I do not believe that they are conclusive.)

I have more than one reason for opposing realism about the thinking self. Besides the threat it poses to mental fictionalism, it is also a threat to the kind of global fictionalism in ontology that I

¹ My thanks to Bill Lycan, an audience at the 2019 workshop on Mental Fictionalism at the Hungarian Academy of Sciences, and especially Julianne Chung for invaluable feedback on this paper.

² Something very much like this argument is given by anti-fictionalists Márton & Tózsér (2013; 2020). N.B., some might claim instead a direct intuition of 'Occurrent thoughts exist;' I address this variant of the argument later.

³ For more on such varieties of mental fictionalism, see section 2 of the introduction to this volume.

favor. Such global fictionalism, by the way, offers obvious support for mental fictionalism—so a problem here would indirectly bear on mental fictionalism too. At any rate, allow me first a few words about global fictionalism, since this will at least clarify the present motivations, and make clearer how ontological matters are understood. (But if preferred, the reader may skip the next section without losing the main thread of the paper.)

2. Background on Ontological Commitment

When it comes to ontology, I am commitment-phobic—I wish to avoid ontological commitments, strictly so called. That is because it is often right and natural for one’s existential beliefs to be questioned during an inquiry, without advanced warning. So I would not want to thwart such questioning by being categorically wedded or committed to claims about what exists.

Plausibly, many philosophers use ‘ontological commitment’ in a way that does not close off such questions. They see it as perfectly intelligible to be “ontologically committed” to electrons, say, while seriously contemplating whether to give up such a commitment. Now I would not want to fight over the term ‘commitment’. But to my ear, this way of speaking sounds needlessly confusing. I find it more natural to say that we have ontological *beliefs*, beliefs which are entirely defeasible, which we do not have any standing “commitment” to uphold.

In fact, I would most prefer to talk of ontological *assumptions relative to a context of inquiry*, where these assumptions vary from context to context. For it is a bit odd to say that I believe in electrons, while asking seriously whether I should relinquish this belief. (Shades of a Moorean paradox: “I believe in electrons, but there very well may not be any.”) However, if ‘there are electrons’ has the status of a provisional, contextual assumption, then there is a question about how to interpret the sentence. Again, I may say ‘there are electrons’ where this is meant to express realism about electrons. But I may say it without “committing” to the existence of electrons across all contexts. The question then is: In my utterance, is my existential quantifier commissive or not? Does my statement *entail* that there are electrons?

If it does not, then it seems like the quantifier is not commissive. However, the quantifier must be commissive if the sentence manages to express something ontological at all. If it is non-commissive, it would seem like the quantifier in ‘there is an even prime’ when uttered by a mathematical anti-realist. However, when I say ‘there are electrons’ I *am* expressing realism about electrons. And yet, I want to say that I am not *committed* to electrons—at least in the sense that, in some contexts, I would be entirely agnostic about electrons. For instance, I see mereological nihilism as a live hypothesis, where electrons would be mere heaps of more basic entities, superstrings or monads or what have you. I would not take such nihilism as seriously as I do, if I were simply “committed” to the reality of electrons.⁴

So when I say ‘there are electrons’, does my position entail that there really are electrons? One could hold that “my position” changes from context to context, but such vacillation in worldviews does not seem very respectable. And it seems *possible* to have a single (consistent)

⁴ Quine, in his classic (1948), avoids such issues by speaking of the ontological commitments of a *theory* rather than the ontological commitments of a person. But if preferred, the present issues can be re-framed in Quinean terms by observing that I often work with different theories (with different ontological commitments) in various contexts.

position, while freely engaging in different inquiries with different ontological assumptions. Sure, the different inquiries will have assumptions that are inconsistent with each other. But *I* am not being inconsistent, if I willingly adopt the different assumptions, relative to the different contexts of inquiry. I am not being inconsistent, just open minded!

But the basic datum remains: If I say ‘there are electrons’ in one context, and then refuse to say it in another context, it seems I am being inconsistent in realism about electrons. The way to resolve this, it seems, to see the ontological statements as tacitly prefixed with a fictionalizing operator. Thus, when I say ‘there are electrons’, I am really saying “According to the fiction adopted in the present context, there are electrons.” Now in one sense, such a statement is genuinely ontological. Ontologically significant statements *can* occur embedded in fiction, as when a story begins ‘once upon a time, *there was* a dragon...’ The story includes a sentence with a bona fide existential content about a dragon. But at the same time, it is just a story. In precise terms, the statement has an implicit fictionalizing prefix, whereby it expresses something like “According to the present fiction, at some time there was a dragon...” The prefixed statement does not entail realism about a dragon, yet the prefix hooks onto a *bona fide* existential statement—a statement which, apart from the prefix, has a heavyweight quantifier.⁵

This fictionalist proposal is how I hope to construe my ontological declarative expressions. Naturally, there still remains a difference between telling a fairy tale, and giving a theory about electrons in, e.g., the explanation of magnetism. In the fairy tale, it is obvious that we are just spinning a yarn. But with electrons, I am telling a narrative as well. Yet unlike with dragons, the existence of electrons is something I am *open* to. I am just not committed to it. Put differently, when I say ‘there are electrons,’ it is understood to have a tacit fictionalizing operator—and as such, the statement is not *anti*-commissive; it is only *non*-commissive.

This can be generalized into a *global fictionalism* regarding existential statements. The view would be that for any sentence of the form “x exists,” its truth condition is a condition of some model or fictional reality, which may or may not “correspond” to reality outside of fiction.⁶ In this manner, the quantifier is commissive in a *relative* sense—relative to the model that is salient in the context—but non-commissive if the statements are interpreted in an absolute sense. This is not a hermeneutic fictionalism, but rather a revolutionary one in Burgess & Rosen’s (1997) sense. It is not the claim that all existential statements in fact have a model-relative commissive quantifier—but rather that inquiry makes the most sense when we interpret them that way.⁷

⁵ My understanding of fictionalist semantics is highly influenced by Rosen (1990) and Nolan (2002).

⁶ On the metaphysics of nonactual, imaginary realities, see Parent (ms.).

⁷ There are interesting comparisons between the present view and Carnap’s (1950) framework-relative metaontology. Carnap could be read as a global fictionalist as, where any existential statement is context-relative due to a fictionalizing prefix (“According to the framework we are presently using, there exists...”) However, Carnap would seem to be more of a hermeneutic rather than a revolutionary fictionalist, given that existential statements are *described* as meaningless when used outside their proper framework. For my part, English is flexible enough that existential statements can have a meaning without any fictionalizing prefix (although I deny that it is usually the best way to understand such statements). I hope to elaborate further on these matters in future work.

3. Lycan's Realism

There are lots of reasonable objections you might raise at this point. Unfortunately, most cannot be addressed here (though I hope to do so in my next book).⁸ But there is an objection that was first raised to me years ago by Bill Lycan. Lycan pressed that every person is saddled with at least one unwavering realist commitment, namely, a commitment to the person herself. The argument he gave was something along the following lines:

- (L1) It is undeniable that I exist.
- (L2) If (L1), then I am committed to realism about at least one thing.
- (L3) So, I am committed to realism about at least one thing.

The conclusion obviously follows, and I would accept (L2) as true. Yet it is undeniable that I am a thing in the world? It should be clarified that we are not talking about the well-worn *diachronic* issue about personal identity, about the continued existence of numerically one person over time. Rather, we are talking about the *synchronic* question: Even if there is no such thing as a continuous self over time) is it *presently* true that a "self" or a person exists? And if so, what undergird its present existence? (Such a self might also persist through time, but we are leaving that aside.)

Having clarified that, why believe (L1)? Knowing Lycan, he might claim a Moorean commonsense intuition in its support.⁹ Yet it is the pleasure of philosophers to question commonsense, or so they say. Even so, others will naturally look to the *cogito* to support (L1). But to avoid question-begging, the *cogito* might be reformulated thus:

- (Cog') These presently experienced, occurrent thoughts exist.
- (L4) If (Cog') is true, then something produces these thoughts.
- (L5) Anything which thinks *these* thoughts is me.
- (*) So, I exist.

We can then get to (L1) if we add:

- (L6) The previous reasoning has undeniably true premises and is undeniably valid.

But what is the status of the premises? (Cog') is thought to be known in an especially immediate and direct way. But neither (L4) nor (L5) enjoy the indestructible certainty that Descartes sought. Yet they may remain "undeniable" by a reasonable standard. After all, in normal course of events, things are causally produced by some antecedent condition or state-of-affairs. The producer of thoughts is a thinker, and in the case of *these* occurrent thoughts, the thinker is me.

⁸ One footnotable objection: If I say that my ontological assumptions are context-relative, am I committed to the existence of folk psychological posits like *assumptions* and *contexts*? What I should say is that, relative to an "FP context," I assume the existence of assumptions and contexts. In which case, I will say that in the different contexts, I choose different ontological assumptions. But what if we are not in an FP context?" *Then, I won't have anything general to say about my ontological assumptions in the various contexts.* Regardless, as I confirm below, the present context *is* an FP context.

⁹ For examples of Lycan's Moorean approach, see Lycan (2001; 2007.)

Of course, in a volume on *mental fictionalism*, (Cog') hardly seems “undeniable.” That is so, assuming we recognize that mental fictionalism is not self-refuting.¹⁰ But mental fictionalism would have us eschew straightaway the *thinking* self, end of discussion. So I will not be assuming mental fictionalism in what follows. My aim, nevertheless, will be persuade you to resist a commitment to the thinking self, by resisting (L5). This, in turn is ultimately part of a defense of mental fictionalism, and part of a commitment-phobia regarding ontology generally.

4. The Paronymy of NPs (Lycan Rides Again)

In discussing the self, it is necessary to present one further preliminary. It is crucial to recognize that noun phrases (NPs) in natural language seem to be equivocal in widespread and systematic ways, including NPs that refer to oneself. This something Lycan himself has made clear in recent work (Lycan 2017). Lycan introduces the basic phenomenon as the *paronymy of proper names*, where “paronymy” is ambiguity that is not *sheer* ambiguity (as when different people share the name ‘Dan’). As an illustration, Lycan (p. 407) asks us to consider the name ‘Germany’ in the following uses:

- (G1) Germany is east of France.
- (G2) Germany voted Social Democrat.
- (G3) Germany invaded Austria.
- (G4) Germany won the World Cup.
- (G5) Germany loves potatoes.

Thus, consider that the relevant landmass east of France is incapable of voting, invading, winning matches, or loving potatoes.

Lycan observes, moreover, the paronymy of proper names creates confusion about identity in relation to persons (p. 408). Consider that Terri Schiavo died on March 31st, 2005. But loved ones might have rightly said that she left us before then, due to entering a persistent, vegetative state. Such a post-psychological condition is sufficient for her absence. In which case, the following has a true reading even if her living body is present in the room:

- (TS1) Terri Schiavo is no longer with us.

Even so, it can be simultaneously true that:

- (TS2) Terri Schiavo is at Pinellas Park hospice.

The explanation, again, is that the name ‘Terri Schiavo’ is paronymous; in particular, it refers to the living body in one case but not in the other.

¹⁰ See Wallace (2007), Joyce (2013), and Parent (2017, ch. 9).

For that matter, consider that (TS2) remains true on one reading, even just after the body dies. The name refers to the body regardless. But simultaneously, of course, there would be an even more palpable sense in which (TS1) is true.¹¹

Lycan further notices that such paronymy is not limited to proper names, but can arise with a variety of NPs, including the first-person indexical. He has us consider the following sort of contrast (p. 413):

- (Me1) After I die, I will no longer exist.¹²
 (Me2) After I die, I will be buried in the ground.

Here too, the indexical ‘I’ seems to shift in reference from the mind to the body.

In addition, Lycan observes that such paronymy explains some of the confusion with the “animalism” debate, i.e., the debate on whether persons are living animals. (Carter 1989, Snowdon 1990, Olsen 1997, Blatti 2012.) Animalism has very strong appeal, but it would seem to dictate that (TS1) and (Me1) are false, as long as the living body is present. Lycan’s response (p. 413) is that it simply depends on how interpret the relevant NP, and I think he is right about that.

So again, the paronymy of NPs can cloud the nature and existence conditions of the self, even as concerns synchronic existence. In fact, the paronymy goes beyond the distinction between the body vs. the mind. For there are cases where the self seems to be isolated to a proper *part* of the mind, e.g., when another part of the mind is introspected as an object. In such a case, the object appears distinct from the psychological subject, whence “the self” is not identical with the (entire) mind.

More concretely, consider that the following is often true in my own case:

- (Me3) When I am anxious, I calmly and mindfully observe the feeling and experience relief.

In such cases, the feeling of anxiety does not completely disappear. So in one sense “I” am continuing to feel anxiety. And yet in another sense, “I” do not feel anxiety, as long as I calmly and mindfully watch the feeling. In such cases, the anxiety seems to become an object of thought, and thus becomes separate from me, as the thinking subject. Once the feeling is thus compartmentalized, there is a clear sense in which “I” am no longer anxious.¹³

¹¹ Julianne Chung retorts that (TS1) might still be unequivocally false in these scenarios, for it is true that Terri Schiavo still exists as a memory. One worry, however, is that this is a kind of use/mention conflation. A memory is a *mental representation*, but Terri Schiavo is not—she is a person. I shall repeat such complaints more than once; regarding the self as a kind of idea is a very strong temptation.

¹² Questions of post-mortem existence are not meant to be summarily dismissed here. So we might instead see (Me1) as short for: After I die, I will no longer exist in the same way, viz., in an embodied form.

¹³ Lycan (in conversation) reports not hearing the paronymy in (Me3). Perhaps we are habituated to attributing contrary emotions to a single subject (anxiety and calm), where strictly and literally speaking, *I* am both calm and not calm. Nevertheless, I would submit that on at least one interpretation of (Me3), there seems to be a subject of anxiety, and something distinct experiencing calm.

If experiments in mindfulness do not appeal, similar examples arise with cases of critical self-reflection. Consider:

(Me4) When I start believing in a perfect God, I catch myself and curtail it.

These are cases where “I” intentionally try to alter what “I” believe. And the “who” that acts against the belief in naïve theism is not a naïve theist. Yet it is “I” who is sometimes a naïve theist, despite myself. The way to make sense of this, again, is to think of ‘I’ as paronymous vis-à-vis different parts of my mind.

The suggestion, then, would be that for each of (Me3) and (Me4), the second indexical differs slightly in reference from the first. After all, it is not as if every single neuron in your brain is needed to think a thought. In the case of naïve theism, then, the belief is likely realized by a subgroup of neurons. Other neurons can then manifest the introspective check on the belief; these other neurons are then the “me” who curbs “myself.” In which case, paronymy in ‘I’-talk arises even in reference to the mind.¹⁴

Paronymy in “me” talk may also explain why many researchers now talk of multiple selves. In a widely cited article, Ulric Neisser (1988) distinguishes between the ecological self, the interpersonal self, the extended self, the private self, and the conceptual self. Here are some examples based on Neisser’s observations:

- (UN1) “When you touch my shoulder, you are touching *me* even if a shirt and jacket interpose between your fingers and my skin.” (p. 39) [Ecological self]
- (UN2) If we write Neisser a letter, we are addressing *him*, even if (unbeknownst to us), he died several years ago. [Interpersonal self]¹⁵
- (UN3) Neisser is (now!) identical to a person who gave a talk in Aberdeen in 1998 (p. 47), even though Neisser died several years ago. [Extended self]

I expect such cases to be controversial, but *if* we accept the different “selves,” a puzzle arises. An individual human organism is supposed to have each kind of “self.” However, if Neisser has five different selves, which one is *him*? (There can only be one.) Neisser seemed to appreciate this conceptual difficulty only somewhat; ditto for many of those who cite him.¹⁶

The connections with Buddhism is now becoming impossible to ignore. For analytically-informed discussions about Buddhist fictionalism, see, e.g., D’amato (2013) and Guerrero (2018)

¹⁴ Split-brain patients or patients who suffer dissociative identity disorder may offer even more vivid examples of intra-psychic paronymy in their “me” talk. However, since such persons are statistically atypical, I prefer to stress other kinds of examples. Dennett (1986/1992) also cites the phenomenon of “talking to oneself” as an example where there are two different “me”s in the same psyche. I would prefer to say that there are two different referents for “me talk” in expressions like ‘I am talking to myself’ (and one could easily flesh out the example at (Me4) along these lines). For further comparison/contrast with Dennett (1986/1992), see section 7.

¹⁵ Your *idea* of Neisser is perhaps key in explaining your letter-writing behavior. Regardless the letter is being addressed *to him* and not just to one of your own ideas.

¹⁶ Neisser raises something like this issue in his concluding paragraph; however, his answer suggests that he treats the question as “how do I come to *represent* myself as one person?” since he talks here of the cognitive integration of information from disparate sources. But the question I am raising is “How can I *be* a single person if I have five selves?” The answer, I think, is that in one respect “I” am *not* a single person but a disunity (like “Germany”) and that in another respect, which self is “me” depends on the goals and purposes adopted in the context.

However, the difficulty can be resolved once we recognize the paronymy of NPs. Again, it just depends on how we interpret the relevant terms—and as Lycan implies, it can be reasonable to interpret them in different ways, depending on whatever goals and interests we have at the time.

5. Fictionalism about the Self.

So talk of a “person” or a “self” is rather like talk of “Germany.” Different parts or aspects of the German totality figure into the different present-tense truths about “Germany,” viz., the German landmass, the German electorate, the German military, the German football team, etc. Similarly, different parts or aspects of the “you” totality figure into different present-tense truths about you. Your mind, your body, your body plus your clothes, etc.

In light of such observations, there is a simple argument for fictionalism about “you.” For consider that we should be fictionalists about Germany, if “Germany” is understood to be a single entity which satisfies all the different predicates in (G1)–(G5). (Let us forbid wildly disjunctive or gruesome entities.) We speak as if there is one such entity, but that appears to be just a convenient *façon de parler*. Similarly, there is no single entity which satisfies the different predicates in (Me1)–(Me4), or (UN1)–(UN3), etc. For instance, there is no single spot in the brain which feels anxiety, is mindful of anxiety, occasionally believes naïve theism and simultaneously resists such belief, etc. At best, we have a hodgepodge of subsystems or modules, and none is the unequivocal reference of ‘I’.

This simple argument looks convincing as far as it goes. But it may not go that far. After all, you might object, even if there is no single “Germany,” there still exists a German landmass. And there still exists a German military. Etc. Similarly, you may complain that the simple argument just shows that our talk of “I” is semantically shifty and can refer to a variety of things. Nevertheless, my point remains that there is no one thing that counts as “I,” even if various “I-related” phenomena exist.

Also, waiting in the wings is a more interesting type of fictionalism about the self. Let us focus on the satisfier of (Me1).¹⁷ What is it that ceases to exist at the time of death? Certainly not the body. Yet assume that physicalism is true. Then, if I am not just a fiction, I am just the body, or at least a part of the body. But no part of the body ceases to exist at death. However, since I cease to exist, this must mean I am not any part of the body. So the thing that ceases to exist at death must have only existed in fiction.

We can reconstruct the reasoning here as follows:

- (1) If I exist outside of fiction, then I am identical to (some part of/) this biomass [= my body].
- (2) If I die at *t*, I cease to exist at *t*.
- (3) If I die at *t*, no part of this biomass ceases to exist at *t*.
- (4) Therefore, no part of this biomass is identical to me. [From (2), (3)]
- (5) Therefore, I do not exist outside of fiction. [From (4), (1)]

¹⁷ For the record, the argument about (Me1) could conceivably be applied to (TS1), although rather than the time of death, the argument would concern a time after which Shiavo was no longer aware/cognizant.

Some objections and replies are now in order.

The first premise is based on the physicalist idea that if I exist outside of fiction, then I must be something physical, presumably, something bodily. But “physical” existence might not be limited to *bodily* existence. After all, the physical includes not only matter but also forces, including electromagnetism, the force behind the attraction and repulsion of chemical ions within the firing of synapses. Upon death, thoughts stop because the synapses stop firing, meaning ultimately that certain electromagnetic forces disperse or fail to be coordinated in the necessary way. Thus, after death, “I” will not exist, in the sense that there will fail to be the right sort of pattern or arrangement of electromagnetic forces in the brain, of the sort that realizes a mind.

On second thought, however, the pattern of electromagnetism seems constitutive of a *mental state*, and not constitutive of *mind*. If the mental state in question is a thought, for instance, then the electromagnetic pattern in the brain is the token thought state, and not the thing that *produces* the thought. In attending to electromagnetic patterns, we have conflated the product with the producer.

The producer of the electromagnetic pattern, moreover, is the brain. So really, on this physicalist picture, what ceases to exist at death is only the thoughts and not the *thinker*, i.e., the producer of the thoughts. The brain still exists and is buried in the ground after death; accordingly, if ‘I’ refers to the brain in (2), then (2) is false. But (2) is not false. Such a physicalist view therefore appears to misfire.

However, is it fair to say that the brain is what *produces* thought? It may be more that the brain is where electromagnetic forces “assemble” into the right sort of pattern, when a thought token comes into existence. True enough, various contours of the brain might be instrumental in bringing about the right electromagnetic pattern. Yet the nature and magnitude of electromagnetism in the brain also partly explains how the patterns come about.

One might compare it to explaining a whirlpool in the bend of a river. The contours of the riverbed are instrumental causing the whirlpool, but it’s not the whole story. The physical features of water are also important to explaining the whirlpool. And when those features change in wintertime (when the water turns to ice), the whirlpool ceases to exist.

So what *produces* the whirlpool? Well, really, it’s better to say that the whirlpool is the product of various features of the riverbed and features of (liquid) water itself. It’s a bit misleading to say there is a separable *thing* that creates the whirlpool. Similarly, when it comes to the “thing that thinks,” it seems better to say that a thought is not really created by some distinct entity. A thought is instead a pattern of electromagnetism, explainable by the arrangement of molecules in the brain and their physical features, and by the dynamical features of intermolecular forces. These variegated, multi-level conditions are not *produced* by the brain, except in a loose way of speaking. The brain is more the “favorable environment” in which such conditions all co-assemble.

But if this is what we really want to say, then it supports rather than undermines that the thinker is unreal. For we are now saying that there is no entity which produces thinking, except in a loose way of speaking. In which case, (L5) is false

This might prompt us to backpedal, reverting to the view that the brain, strictly and literally, is the thing that thinks. But since the brain continues to exist after death, this could not be *me*. (What do I care if my *brain* continues to exist after death?)

This can prompt a different line where the self is identified with a *self-concept* or *self-conception*. We certainly have a concept such that (i) the “self” produces thoughts, and (ii) the “self” ceases to exist upon death. Both of these conditions are not satisfied by the brain, but the objector is now suggesting that the *self-concept* indeed satisfies both conditions. However, this conflates use and mention as regards mental representation: The self is not the same thing as the self-concept but is rather the thing represented by the self-concept. In addition, a self-concept is produced by the thinker, yet assuming there is no circular causation here, a self-concept would not produce itself. Thirdly, the brain can host more than one self-representation at a time, yet a brain is not supposed to host more than one self at a time (paronymy aside).

Yet another objection would be that the self is really the *mereological fusion* of this biomass (or a part thereof) with the neuroelectricity that dissipates upon death. Upon the dissipation, the fusion would cease to exist as well, even though no part of the biomass ceases to exist. The problem, however, would be that the neuroelectricity which comprises a thought would be *produced by an entity which has that very thought as a metaphysical part*. Or, to use the fashionable lingo, a thought would be partly “grounded” on itself, if the thought is partly constitutive of the thinking self. Such circular grounding is not usually looked upon favorably.¹⁸

6. Self-Location

So, at least under the mortally important precisification of ‘I’, (L1) is not true. Granted, fictionalism about the self is counterintuitive, but why is that? *Why* might Lycan have the Moorean intuition that the self is undeniably real? I suspect it owes to the vivid sense that the thinker has a location. But this needs to be qualified at bit. As Hume said, we never catch a glimpse of the self in inner sense. So whence the impression that we can pinpoint its location?

It can help with this puzzle to consider the street-view feature of Google maps. The street-view feature represents a location as if from a particular perspective. One can rotate the frame, thus creating an impression much like moving your head to take a look around. One thus gains the sense that there is a particular *spot* around which the visual representation is rotating, a hub around which the panoramic representation is being rotated, although the hub can never be part of the scene. (If it were, it would no longer be the hub.) But in the visual scene, there are subtle

¹⁸ A better version of the proposal is this. The thinker at t1 = the thoughts at t1 fused with the brain at t1. That fusion indeed produces the thoughts at t2, which in turn, are a metaphysical part of the thinker at t2 (along with the brain at t2). But there is no circular causation here. For the thinker at t1 produces only the thinking at t2, and not the thinking at t1. Nonetheless, in order to avoid a regress, one must posit thinking at t0 which was not produced by a thinker at a previous time, yet partly constitutes the thinker at t0. So the view has the funny consequence that there could be “thinkerless thinking” in the human organism—i.e. a thought without someone thinking the thought.

perspective lines that point back toward panoramic center, and this is what seems to provide a sense of self-location in Google maps. Something analogous seems to occur in our own non-virtual experience as well.

When it comes to non-virtual experience, is the panoramic center the location of the self? If so, then the self would have a location, hence, exist in physical reality. Now I do not doubt that one's own phenomenological hub corresponds to a location in physical reality. The location is a point in the brain, moreover, which is entirely real (as far as "points" go). But is this where the self dwells? I doubt it. At least, this point in the brain does not cease to exist at death, even though something important does.

Yet the phenomenological hub certainly *feels* like the abode of the psychological subject, the thing that is mortally important to "me." But we can erode this feeling by reconsidering the extended thought experiment from Daniel Dennett's (1981) "Where Am I?" The thought experiment is essentially an illustration of how the phenomenological center can patently *fail* to correspond to the self.

One might recall that in Dennett's tale, his brain is surgically removed from his body, to be preserved in a vat of nutrients, and yet:

each input and output pathway as it was severed, would be restored by a pair of microminiaturized radio transceivers, one attached precisely to the brain, the other to the nerve stumps in the empty cranium. No information would be lost, all connectivity preserved. (p. 311)

The drama starts to unfold when Dennett's body walks over to the vat to take a look at his brain.

I thought to myself: "Well, here I am sitting on a folding chair, staring through a piece of plate glass at my own brain . . . But wait," I said to myself, "shouldn't I have thought, 'Here I am, suspended in a bubbling fluid, being stared at by my own eyes'?" I tried to think this latter thought. I tried to project it into the tank, offering it hopefully to my brain, but I failed to carry off the exercise with any conviction. I tried again. "Here am I, Daniel Dennett, suspended in a bubbling fluid, being stared at by my own eyes." No, it just didn't work. Most puzzling and confusing. Being a philosopher of firm physicalist conviction, I believed unswervingly that the tokening of my thoughts was occurring somewhere in my brain: yet, when I thought "Here I am," where the thought occurred to me was *here*, outside the vat, where I, Dennett, was standing staring at my brain. I tried and tried to think myself into the vat, but to no avail. (p. 312)

The story is quite compelling, yet why should the protagonist be so inclined to see himself as located outside the vat? I suspect it is because the protagonist's phenomenology still points toward a center that is located inside the skull. And he is habituated into thinking that this phenomenological center is *his* location.¹⁹ But in the story, no point inside the skull corresponds

¹⁹ In conversation, Julianne Chung stresses that such self-phenomenology does not seem culturally universal. E.g., ancient Chinese philosophers identified the heart as the locus of cognition; see Rošker (2021).

to any region of the brain. And so, the protagonist is simultaneously attempting to understand himself as located elsewhere, viz., inside the vat. But the feat is difficult.

This shows the strength of the belief that the location of the self = the phenomenological center. And it also illustrates that this belief can be resisted: If these locations were literally numerically the same, then it seems unlikely that they could be separated in Dennett's narrative in such a compelling way. (The point is not conclusive, but it provides a basis for doubt).

If so, the phenomenology that allegedly discloses a self-location is non-demonstrative. Per the paronymy of NPs, it *can* correct to say the "self" has a location, e.g., when the relevant NPs refer to the body or a part thereof. But the phenomenological center may not be where the self resides. To use the earlier analogy, the center of the whirlpool is not the location of what produces the whirlpool. Indeed, there is really no single locus where the whirlpool is birthed—and in the same way, I have tried to make plausible that the psychological subject is not located anywhere in particular, except perhaps according to a fiction.

Thus, the sense of a self-location is no proof against fictionalism about the self. Some fictions are more cognitively ingrained than others, and the sense of self-location seems to be a case in point.

7. Dennett's *De Se* Fictionalism

Dennett himself is known for a kind of fictionalism about the self, yet as far as I know, he does not recruit the extended thought-experiment from "Where Am I?" in its support. More broadly, I would demur somewhat from how Dennett understands fictions in general as well as the self. First, consider that Dennett's favorite example of a fiction is a "center of gravity." Centers of gravity are used in Dennett (1992) to elucidate the nature of fiction in general, and the nature of the self in particular. (Dennett's readers may also recall centers of gravity as his 1991 exemplar of a "real pattern.") The only problem is that centers of gravity are not fictional (Haugeland 1993). A center of gravity is a perfectly real, concrete spatial point (or a point trajectory through time) at which the weighted relative position of the distributed mass of an object sums to zero. So it hardly strikes me as appropriate as an analog to the self.²⁰

Further, Dennett's broader view of fictional objects has contestable elements.²¹ Unlike Dennett (p. 106), I am unsure whether typical fictional objects violate the law of excluded middle. Many have argued for such a thing by considering whether Sherlock Holmes had a mole on his back. It

²⁰ Dennett bolsters the alleged fictional status of centers of gravity, writing: "Centers of gravity, as fictional objects...have only the properties that the theory that constitutes them endowed them with. If you scratch your head and say, 'I wonder if maybe centers of gravity are really neutrinos!' you have misunderstood the theoretical status of a center of gravity." But while it is true that the question misunderstands centers of gravity, it is not because they are really fictional. It is rather that a center of gravity is a spatiotemporal *point* (or point trajectory over time) defined by its physical-mathematical relations to the distributed mass of its object. It is not itself a physical object with dimensions (however small) in the manner of a neutrino.

²¹ Another contestable element, which I shall mention only in this footnote, is Dennett's view that typical fictional objects are abstracta. I would prefer to clarify that, in the mind-independent world, fictional objects aren't *anything*. But when we take into account mind-dependent phenomena, some fictional objects are imagined to be abstracta (e.g., Meinong's round square) while others are not. Pegasus is an imaginary object who is imagined not to be an abstractum, but rather a flesh-and-blood horse. For further elaboration, see Parent (ms.).

seems that the answer is indeterminate, given that Sir Arthur Conan Doyle never affirmed nor denied such a mole. But *pace* Dennett, we need not conclude that Holmes neither had nor lacked such a mole. Dennett finds it odd to think that there may be an undiscovered fact of the matter, given that Holmes is Doyle's *creation* and Doyle never decided on such a detail. Sure, but Doyle certainly did not conceive Holmes' moles to be counterexamples to the law of excluded middle. And so we seem entitled to assume that they are not such counterexamples. This would be akin to how Dennett, following Lewis (1978), treats the question of whether Holmes had three nostrils: Of course he didn't, even though Doyle did not specify the matter. In the absence of such a specification, we are entitled to assume that Holmes was normal in the matter of nostrils. Ditto, it might seem, with his moles and bivalence.²²

In any event, the indeterminacy of Dennettian fictional objects implies an indeterminacy in the Dennettian self. One dramatic illustration of this indeterminacy is that Dennett regards the self as possibly becoming *more determinate* through additional narration, much like John Updike could have made the character of Rabbit Angstrom more determinate in concocting further adventures about him. Dennett writes:

We cannot undo those parts of our pasts that are determinate, but our selves are constantly being made more determinate as we go along in response to the way the world impinges on us. Of course it is also possible for a person to engage in auto-hermeneutics, interpretation of one's self, and in particular to go back and think about one's past, and one's memories, and to rethink them and rewrite them. This process does change the "fictional" character, the character that you are, in much the way that Rabbit Angstrom, after Updike writes the second novel about him as a young man, comes to be a rather different fictional character, determinate in ways he was never determinate before. This would be an utterly mysterious and magical prospect (and hence something no one should take seriously) *if the self were anything but [fictional]*. (p. 110)

The idea of the self becoming more determinate seems to have two aspects. First, as time goes on, more stuff happens to you, thus fixing more and more facts of your autobiography or life story. Second, the self according to Dennett is capable of *rewriting* or *reconceiving* her own past in ways that also affect her autobiography.

But, to use a piece of Dennett jargon, this threatens to be a "deepity" (cf. Dennett 2013, ch. 2). It is banal that one can come to new understandings or new interpretations of what happened in the past. And since these new understandings are events in one's life, such events trivially add to the events constituting one's autobiography. However, Dennett seems to think that if the self were not a fiction, there would be something "mysterious and magical" in this. But if one is not a fictionalist about the self, how would these banalities translate into something mysterious and magical? Dennett is potentially suggesting something rather controversial—that we can *rewrite* or at least *fill in* our past and thus affect our past from the position of the present, much like Updike might fill in various bits of Rabbit Angstrom's backstory.

²² This sort of hard line is rarely defended, but Woods (1969; 1974) is an exception. I of course cannot defend the hard line here, but it seems worth asking whether we are entitled to assume bivalence with respect to Holmes.

However, not even Dennett seems to believe that we can literally change our past in this way, for he insists in the passage above that we “cannot undo those parts of our past which are determinate.” The fact is, moreover, that *all* of our past seems to be determinate. Our *understanding* of our past may change, but our understanding of the past is not somehow partly constitutive those past events themselves (although there may be a few bizarre counterexamples here). So what is Dennett saying here beyond the banalities? I myself am unable to say.

Dennett’s *de se* fictionalism has been challenged before, of course. E.g., Galen Strawson (2015) self-reports that *he* does not narrate his life to any degree (at least not consciously). So the idea that the self is the product of some act of narration strikes him as dubious, at least in his own case. Suffice it to say, however, that a *de se* fictionalist need not understand the self as the topic of some internal narrative. It is perfectly possible to create a fictional entity *sans* a storyline. Right now, for example, I can conjure up a duckbill platypus named “Dan” who has a small unicorn-horn, and I can leave it at that, without devising any plot line for him to follow. Besides, I take it that Dennett’s view of selves is that they are partly the product of a *collaborative* narrative, not just a single internal one. E.g., he speaks of the possibility that *others* fashioning multiple selves for a victim of dissociative identity disorder (p. 111). Whether this social constructivist line is plausible, however, is not something I shall investigate.

But there is one further matter about Dennett’s fictionalism which I would like to address. This concerns a distinction, which I adapt from Walton (1990), between fictions about X, where X is an actual object, and fictions about X, where X is not an actual object. Re: the former, consider that Napoleon, the actual French emperor, appears as a character in Tolstoy’s novel *War and Peace*. (The example is from Kripke 1973/2013.) If its implications about Napoleon are not 100% historically accurate, then *War and Peace* uses Napoleon as a “prop” for some creative story telling on the part of Tolstoy. In contrast, Sherlock Holmes seems to be Doyle’s *de novo* creation. So the descriptions of Holmes do not depend on an actual person as a “prop.”²³

Is the fiction about the self, according to Dennett, based on a prop or not? It seems to vary in the course of his discussion. The extensive comparison with Sherlock Holmes suggests that the fiction about the self, like Holmes, is *not* prop-based. However, things seem different when Dennett speaks approvingly of Michael Gazzaniga’s research:

According to Gazzaniga, the normal mind is *not* beautifully unified, but rather a problematically yoked-together bundle of partly autonomous systems. All parts of the mind are not equally accessible to each other at all times. These modules or systems sometimes have internal communication problems which they solve by various ingenious and devious routes...there is no conscious self that is unproblematically in command of the mind's resources. Rather, we are somewhat disunified. Our component modules have to act in opportunistic but amazingly resourceful ways to produce a modicum of behavioral unity, *which is then enhanced by an illusion of greater unity*. (pp. 111-112)

²³ Walton does not speak of props in relation to *fiction* but rather in relation to *pretending*. And it is best to distinguish a pretense theory of X, in the style of Walton, from fictionalism about X. (For relevant discussion, see Lewis 2005, Eklund 2011). So the above should not be seen as explicating Walton’s views; it is rather an analogue view which is obviously inspired by Walton.

This makes it sound like the fictional aspect of the self is the *unity* between the hodgepodge of modules. The modules themselves are treated as nonfictional; in fact, they collectively act as a nonfictional “prop” on which we project the fiction of the self, *qua* unified entity composed of the modules.

My prescription for the apparent inconsistency would be to invoke the paronymy in “self” talk. The paronymy means that sometimes such talk amounts to a fiction involving a prop and sometimes not. Indeed, when we compared ‘I’ with ‘Germany’, we saw that there are a variety of phenomena that can be the referent of such terms. But the fantasy is in thinking that there is a single, unified “Germany” or “self” between all such cases. Nevertheless, at times ‘I’ talk also bears the marks of a prop-free fiction. In sincerely asserting ‘When I die, I will cease to exist,’ I am apparently not referring to the aggregate of modules, or any aspect of the body more broadly. For I know very well that none of this biomass ceases to exist at the time of death. My talk apparently concerns some kind of entity whose existence supervenes on but is distinct from the mental life realized in the brain. But in the brain, there is only thoughts, and no separate producer of the thoughts, at least not beyond the biomatter. So in this case, the “self” is created *ex nihilo*, so to speak; it is a fiction that is not based on an actual prop.

8. Conclusion

The mortally important entity who allegedly experiences mental life between the ears is a fiction. Beyond the biomass, and the neuro-electricity surging through it, there is nothing else. The whirlpool example is meant to pump this intuition, but perhaps your laptop would be even more effective. Its electronic states are the result of electricity coursing through the hardware configuration. Thus, without electricity, the states disappear. But the system that realizes those states still exists; this is analogous to the physical body just after the time of death. Yet when the body dies, we also have a sense that a “subject” of thought has fallen out of existence. However, the laptop should be our model, where it is clear that only the electric states have dispersed, and not anything further.

Regardless, talk of a “thinking self” is largely unavoidable. We should therefore be fictionalists on the matter.

References

- Burgess, J. & Rosen, G. (1997) *A subject with no object*. Oxford: Clarendon Press.
- Carnap, R. (1950) 'Empiricism, semantics, and ontology', *Revue Internationale de Philosophie*, 4, 20-40.
- Carter, W. (1989) 'How to change your mind', *Canadian Journal of Philosophy*, 18, 1-14.
- D'Amato, M. (2013). 'Buddhist fictionalism', *Sophia*, 52, 409-424.
- Dennett, D. (1981) 'Where am I?' in his *Brainstorms: philosophical essays on mind and psychology*. Cambridge, MA: MIT Press, pp. 310-323.
- _____. (1991) 'Real patterns', *Journal of Philosophy*, 88, 21-51.
- _____. (1992) 'The self as a center of narrative gravity' in Kessel, F., Cole, P., and Johnson, D. (eds.) *Self and consciousness: multiple perspectives*. Hillsdale, NJ: Erlbaum, pp. 103-115
- _____. (2013) *Intuition pumps and other tools for thinking*. New York: W.W. Norton.
- Eklund, M. (2011) 'Fictionalism' in Zalta, E. (ed.) *The Stanford Encyclopedia of Philosophy*. Available at <http://plato.stanford.edu/archives/fall2011/entries/fictionalism>.
- Guerrero, L. (2018). 'Buddhist global fictionalism?', *Ratio*, 31, 424-436.
- Haugeland, J. (1993) 'Pattern and being' in Dahlbom, B. (ed.) *Dennett and his critics*. Malden, MA: Blackwell, pp. 53-69.
- Joyce, R. (2013) 'Psychological fictionalism', *The Monist*, 96, 517-538.
- Kripke, S. (1973/2013) *Reference and existence: the John Locke lectures*. Oxford University Press.
- Lewis, D. (1978) 'Truth in fiction', *American Philosophical Quarterly*, 15, 37-46.
- _____. (2005) 'Quasi-realism is fictionalism' in Kalderon, M. (ed.) *Fictionalism in Metaphysics*. Oxford University Press, pp. 314-321.
- Lycan, W. (2001) 'Moore against the new skeptics', *Philosophical Studies*, 103, 35-53.
- _____. (2007) 'Moore's antiskeptical strategies' in Nuccetelli, S. and Seay, G. (eds.) *Themes from G.E. Moore: new essays in epistemology and ethics*. Oxford University Press, pp. 84-99.
- _____. (2018) 'Metaphysics and the paronymy of proper names', *American Philosophical Quarterly*, 55, 405-419.
- Márton, M. & Tózsér, J. (2013) 'Mental fictionalism as an undermotivated theory', *The Monist*, 96, 622-638.
- _____. (2020) 'In defence of the phenomenological objection to mental fictionalism', *Organon F*, 27, 169-186.
- Nessier, U. (1988) 'Five kinds of self-knowledge', *Philosophical Psychology*, 1, 35-59.
- Nolan, D. (2002) *Topics in the philosophy of possible worlds*. New York: Routledge.
- Olson, E. (1997) *The human animal: personal identity without psychology*. Oxford University Press.
- Parent, T. (2017) *Self-reflection for the opaque mind: an essay in neo-Sellarsian philosophy*. New York: Routledge.
- _____. (ms.) 'Conservative Meinongianism: an actualist+ metaphysics', available at <http://tparent.net/ConservativeMeinong.pdf>.
- Quine, W. (1948) 'On what there is', *Review of Metaphysics*, 2, 21-38.
- Rosen, G. (1990) 'Modal fictionalism', *Mind*, 99, 327-354.
- Rošker, J. (2021). 'Epistemology in Chinese philosophy' in Zalta, E. (ed.) *The Stanford encyclopedia of philosophy*, available at <https://plato.stanford.edu/archives/spr2021/entries/chinese-epistemology/>

- Snowdon, P. (1990) 'Persons, animals, and ourselves' in Gill, C. (ed.) *The person and the human mind: issues in ancient and modern philosophy*. Oxford: Clarendon Press.
- Strawson, G. (2015) 'The unstoried life' in Leader, Z. (ed.) *On life-writing*. Oxford University Press, pp. 284-302.
- Wallace, M. (2007) 'Mental fictionalism', unpublished manuscript. A version lightly edited by the author now appears as chapter one of this volume.
- Walton, K. (1990) *Mimesis as make-believe: on the foundations of the representational arts*. Harvard University Press.
- Woods, J. (1969) 'Fictionality and the logic of relations', *Southern Journal of Philosophy*, 51-63.
- _____. (1974) *The logic of fiction: a philosophical sounding of deviant logic*. Berlin: Walter de Gruyter.