

Remember me? First person thought, memory and explanations of IEM

Léa Salje

To cite this article: Léa Salje (02 Aug 2024): Remember me? First person thought, memory and explanations of IEM, Philosophical Psychology, DOI: [10.1080/09515089.2024.2386150](https://doi.org/10.1080/09515089.2024.2386150)

To link to this article: <https://doi.org/10.1080/09515089.2024.2386150>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 02 Aug 2024.



Submit your article to this journal [↗](#)



Article views: 80



View related articles [↗](#)



View Crossmark data [↗](#)

Remember me? First person thought, memory and explanations of IEM

Léa Salje

School of Philosophy, Religion and History of Science, University of Leeds, Leeds, UK

ABSTRACT

An important focus in discussions of immunity to error through misidentification (IEM) is on what explains putative cases of it, and correspondingly, on what significance its presence should be taken to have. This focus is important for the literature: the interest of the phenomenon naturally depends on why it is supposed to be significant when it arises. This paper looks to the claimed IEM of memory-based judgments as a case study to advocate for a pluralist account of explanations of IEM. I argue that Evans was right that Shoemakerian q-memories fail to undermine the supposed IEM of memory-based first personal judgments, but that he was wrong about *why*. In fact, cases of q-memory reveal a surprising referential underdetermination in first person thought, which subvenes an underdetermination of the explanation of IEM for memory-based judgments. Given this plurality of explanations active in just a single case study, we should be cautious about claims to IEM having just one source of significance.

ARTICLE HISTORY

Received 8 January 2024



Accepted 23 July 2024

KEYWORDS

Memory; IEM; first person thought

First personal judgments, made on certain grounds, are often taken to be immune to error through misidentification (IEM) relative to the use of the first person concept – that is, it's taken to be the case that the subject is in a position to know, on those grounds, that *someone* instantiates the ascribed property, but is wrong only in taking that person to be themselves (thought of first personally).¹ What is the explanation, and corresponding significance, of claims like these? This paper explores this question by asking, by way of example: what is the significance of the supposed IEM of first personal judgments made on the basis of episodic memory?²

A commanding answer to this question is that it indicates that the given grounds – memory, in this case – is a directly first personal way of knowing about oneself; a way of knowing oneself *as oneself*, under a first personal guise or mode of presentation. It's not a way of finding out about myself as

CONTACT Léa Salje  l.c.salje@leeds.ac.uk  School of Philosophy, Religion and History of Science, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, UK

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

an object encountered in the world, akin to spotting my name in a newspaper account of a past event and only then realizing the story is about me; rather, it's a way of directly finding out about how things once were with me, *thought of as such*. Why is this? Because according to an attractively simple explanation of IEM, it arises just in case there is an absence of an identification component in the grounds on which the judgment was formed or in the presuppositions underlying the judgment's formation. This means that I do not, in the course of arriving at the judgment, at any point have to think of myself, thought of in a "me-ish" way, as identical to someone thought of in some other way. If there is no identification in the formative or presuppositional background to my judgment, then there is no question that its formation involved an apprehension of myself under some non-first personal guise, which I then identified with myself (thought of as such). So, it must be that when I come to judge things about myself on the basis of memory, those grounds allow me to apprehend myself *directly* as myself. The significance of the putative IEM of first personal memory-based judgments, on this view, is that it shows memory to be a directly first personal mode of self-apprehension.³

Simple as this answer sounds, in this paper, I argue that it cannot be the full story – by which I mean it cannot be the story in *every* case of IEM. I will argue that different instances of IEM will receive different best explanations, which means that the presence of IEM will signify different things in different cases. We should therefore be cautious about accepting claims to a uniform account of the significance of IEM.

The way I want to get to this result is through consideration of the IEM of episodic memory judgments, and more specifically, episodic memory judgments apparently had from a first person perspective (I will not consider so-called "field" or "observer" memories). The putative IEM of first personal memory-based judgments has classically been taken to be under threat from the conceivability of so-called *q-memory* cases. Introduced by Sydney Shoemaker, *q-memory* is a capacity just like ordinary memory, except that it leaves open the identity of the causal origin of the memory-impressions. Gareth Evans famously argued that this challenge fails. The deliverances of memory, he argued, are always first personal in content, so *q-memory* can be at best a way of generating systematic illusions about how things were in *my* past, not a way of knowing how things used to be with someone else. This is enough to secure the claim that the self-knowledge generated by episodic memories is identification-free; it does not result from prior knowledge of an identity between oneself thought of non-first personally and oneself thought of first personally.⁴ And *this* is enough to preserve the IEM of first personal memory judgments.

In what follows, I am going to argue that although Evans is right that Shoemaker's challenge fails, he was wrong about *why* – or, rather, he was

wrong that his explanation is the only explanation. I'll start by setting out the Shoemaker-Evans debate about q-memory (Section 1). Drawing on work from John Campbell I then argue that the thinker-reflexive rule (TRR) that individuates the first person concept is underdetermined, because ambiguous between (at least) three different relations we each normally bear to our own thoughts (Section 2). Disambiguating that reference rule results in differential reference assignments in cases of q-memory (Section 3). As it happens, all three of the predicted assignments allow for the preservation of IEM of memory judgments, but what explains the IEM is different each time (Section 4). In a way, then, the offering of this paper is a contribution to the Shoemaker-Evans debate on q-memory in which I side with Evans insofar as I offer a way of thinking about the cases that have them uniformly turn out to fail as a challenge to the IEM of first personal memory judgments. But, the more important intended contribution uses this discussion of memory as a case study to illustrate the variety of explanations there can be for instances of IEM. I end with some reflections on what we should take away from all of this about the nature of first person thought (Section 5).

First, we'll need to get a working definition of IEM on the table, as it applies to judgments based on memory impressions:

Memory-IEM: A judgment *I was F* formed on the basis of memory-impressions of being *F* is IEM relative to the use of the first person concept iff it cannot be mistaken in the following way: the subject knows, on that basis, that *someone was F* (*de re* or existential) but has made a mistake only in misidentifying the person that they thereby know to be *F* with the person referred to by their use of *I*.⁵

1. Shoemaker and Evans on Q-Memory

It is standardly taken that ascriptions of IEM are relative to a judgment's grounds; *which* grounds give rise to judgments with IEM, however, is still an open question. Memory seems like an excellent candidate – there is something compellingly direct about the epistemic access to the past we get from remembering an event, as contrasted with reading about it or being told what happened by a friend. If one remembers it, the very challenge of a potential misidentification can seem absurd. (“Of course it was me, I *remember* it happening!”).

Following a suggestion by Grice, however, Shoemaker diagnosed this intuition as resting on a piece of linguistic trickery. It is, he points out, just a stipulative feature of all cases we are liable to classify as memory that the earlier experiencing subject is identical to the later rememberer, and so that the represented past is one's own. But, all this shows is that “we refuse to *call* someone's having knowledge of a past experience a case of his remembering it unless the past experience belonged to the rememberer himself.”

(Shoemaker, 1970, p. 270).⁶ Remove this stipulation, argued Shoemaker, and there's no reason to think that these judgments couldn't be mistaken in exactly the way ruled out by claims to IEM. Shoemaker coined a new faculty for this purpose, *quasi-* or *q-memory*. Q-memory is just like ordinary episodic memory, except that it is neutral about the identity of the earlier experiencer and later rememberer.⁷

Take the following case of q-memory:

Memory Transplant. A highly advanced team of neuroscientists succeed in precisely identifying and removing the multiple distributed parts of Earl's brain (call these collectively the "memory trace") encoding information from his past experience of looking up at the colourful lozenge windows in the Sagrada Familia. They then graft this memory trace onto Larry's brain, meticulously fitting new brain-bits to old, connecting up trailing synapses, and so on. Upon awakening from surgery, Larry is struck by personal-level conscious memory impressions of sunlight streaming through high bright windows in the Sagrada Familia. As it happens, Larry has never been to Barcelona. Larry forms the judgement on these grounds *I was in the Sagrada Familia*.

It certainly seems as if the judgment in this case is mistaken because (and only because) although Larry knows that *someone* was in the Sagrada Familia on this basis, he's gone wrong in misidentifying himself as that person. First person judgments made on the basis of memory impressions might have *de facto* IEM – that is, mistakes of the kind ruled out by the status of IEM can't *in fact* occur, given the world we happen to be in. But, cases of q-memory like this one appear to show that they do not have logical immunity – that is, that these mistakes are impossible in *any* world, as a matter of conceptual or logical necessity. The IEM of memory-based judgments is just a contingent fact about how things happen to be set up in the actual world.

Evans rejected outright the supposed threat to the IEM of memory judgments coming from the intelligibility of such cases. Telling stories like this, he argued, does nothing to settle the question whether the contents of the apparent memories are person-neutral or first personal. Indeed, Evans insists that we cannot think of them as person-neutral – memory states always present themselves as memories *of oneself*. Recall now that other than identity-neutrality between earlier experiencer and later rememberer, q-memory is supposed to be just the same as ordinary episodic memory. What this means is that if Evans is right, then what is introduced with the conceptual possibility of q-memory is not a systematic source of knowledge about someone else's past, but a source of systematic illusions about how things once were with oneself; he explains, merely defining a notion of q-memory "leaves the question of the *content* of memory states quite untouched; it can still be right to say [...] that an apparent memory of ϕ -ing is necessarily an apparent memory of oneself ϕ -ing." (p. 248). The

q-memory challenge, so the charge goes, is itself a piece of linguistic trickery on Shoemaker's behalf.

2. Three TRRs of Reference for I

Evans' claim that memory impressions are always first personal – “that an apparent memory of ϕ -ing is necessarily an apparent memory of oneself ϕ -ing” – in content is *prima facie* compelling. Memory impressions certainly don't seem to be neutral with respect to whose past is being represented, and indeed, even those who have disagreed with him on the question whether memory judgments are IEM have tended to accept this point.⁸ The claim also finds widespread support in the current psychological literature; its self-referential or auto-noetic nature is widely cited as a distinguishing feature of episodic – as contrasted with semantic or mere event – memory.⁹

There are, however, two sorts of claims that might be being offered here which need distinguishing. On the one hand, the claim might be that these mentally represented contents always involve the use of a first person concept – a *me-ish* mode of apprehension or presentation. On the other, the claim might be that these mentally represented contents *are always about the rememberer*, thought of first personally. In Fregean terms, one is a question of sense and the other of reference. What is compelling, it seems to me, is the part of this picture that is introspectively evident to the thinking subject: that is, that the *sense* involved is a first personal one, reflected in the described case by the fact that Larry takes it up into a first personal judgment. It is compelling that q-memories present the remembered subject in a *me-ish* kind of way.

Now, in most circumstances, this would be enough to settle the question of reference too. At least, according to a plausible and widespread view according to which uses of the first person concept always refer to their thinkers (and here we leave Evans behind us¹⁰), once we know that it is a first person concept in play, we will know that it refers to its thinker. Call this the TRR model of first person thought. The TRR model says that the first person concept is individuated by a rule of thinker-reflexive reference, or the rule that uses of the first person concept reflexively refer to their thinkers. I propose to grant the TRR model of first person thought.

My suggestion now is that even if we accept Evans' compelling claim that q-memories always involve a first person sense, there is a surprising question of first person *reference* that remains open.¹¹ This is not because in cases of q-memory first person contents no longer follow the rule of thinker-reflexive reference – since, we have granted the TRR model, if it didn't, it would no longer be a first person content. Rather, the suggestion is that in that in the highly unusual cases of q-memory it's no longer clear how to apply the rule.

It's unclear for the following reason. While the TRR is widely accepted, there is much less agreement on exactly how to spell it out. Just a few formulations in current currency include:

'I' refers to whoever uses it. (O'Brien, 2007, p. 7)

'I' refers to whoever has the control over its production. (O'Brien, 2007, p. 68)¹²

$\forall x \forall \text{event of thinking } \theta: x$ is the reference of a use of an instance of the [self] type in the event θ of thinking iff x is the producer (agent) of that event θ of thinking. (Peacocke, 2014, p. 83)

The reference of the first person is fixed by the simple rule that any token of 'I' refers to whoever produced it. (Campbell, 1999a, p. 621)

[A]ny token first person thought will be about the subject whose thought it is. (Morgan, 2015, p. 1801)

[F]irst-person thought is [...] reflexive thought: in grasping a first person thought x thinks of x . (Madden, 2016, p. 190)

Even in this compressed sample there's conspicuously little overlap in the rule's phrasing. We find mention of being a thought's *producer* (or *having control over its production*), being its *agent*, being the *user* of the token first person concept, *grasping* the thought, and *being the subject whose thought it is*. None of these seem obviously wrong – and, of course, in ordinary contexts very little hangs on their differences.

The fact that there are multiple adequate specifications of the relation thinkers bear to their thoughts has its theoretical uses. One writer who has made powerful explanatory use of it is John Campbell, in his influential discussion of the schizophrenic delusion of thought insertion (Campbell, 1999b). This is a delusional belief that thoughts occurring in one's stream of consciousness have been implanted there by an external source. A question that arises here is how to intelligibly characterize the delusion's content. If we say that it involves believing that one is having thoughts that are not one's own, it might seem, as Campbell writes, "like an Escher drawing, an illusion which is compelling but has no coherent content" (1999a, p. 620). Campbell avoids this imputation of contradiction by distinguishing two senses of thought ownership. On the first, "[f]or a token thought to be truly mine, it must have been generated by me" (1999a, p. 36), whereas the second is a sense of thought ownership "on which what makes a psychological state mine is the possibility of self-ascription of it by me" (1999a, p. 35). The content of the delusion, then, involves claiming and denying ownership in two distinct senses; the subject "has first person knowledge of token thoughts which were formed by someone else. And there is no contradiction in that" (1999a, p. 620).

Now, Campbell's motivation for teasing apart these different relationships we bear to our thoughts had something special to do with finding a way to render the contents of these delusions intelligible.¹³ That is not our concern here. But even unlatched from the initial theoretical context of their identification, what Campbell has done is to draw our attention to an important feature of the relationship we bear to our own thoughts: namely, that that relationship is not as simple or as uniform as we might at first have thought. It is this complexification of that relationship that I now want to draw on here.

So, what "thinker-roles", as we might call them, does Campbell bring into view in his discussion? Let's call one of his notions of thought ownership in the paragraph above the *entertainer* thinker-role – a subject instantiates it with respect to a given conscious thought just in case she is able to immediately self-ascribe it on the basis of introspection alone. The other is a notion of ownership in which I own my thought *because it was generated by me*. I can think of at least two ways in which this might be the case. A thought might be generated by me insofar as the conscious episode is physically generated or sustained by neural activation in my brain. Call this the *neural* thinker-role. Alternatively, we might say that a given thought was generated by me because I am the one who *thought the thought* at the personal-level – I am the subject whose recent mental history made it the case that the thought had the intentional content that it did. This might be as an inferential or otherwise rational product of preceding thoughts, or as a member of a loosely strung-together chain of thoughts, as a long-standing thinking project that I am disposed to return to, or because it is generated by my specific psychic constellation of concerns, values, beliefs and experiences, or any one of many other ways in which one might generate a thought at the personal-level. Either way, a thought is mine in the relevant sense if its intentional content is formed or generated by me at the personal-level of mental content, rather than at the underlying neural level. Call this the *author* thinker-role. The author thinker-role is likely what Campbell himself had in mind for the first of his notions of thought-ownership; as he says, these thoughts are mine in virtue of "the fact that they are products of my long-standing beliefs and desires" (1999a: 261).

Given these three thinker-roles, there are at least three ways in which the TRR might now be precisified¹⁴:

Entertainer. I-thought reflexively refers to its thinker qua entertainer, or the subject who is in a position to immediately self-ascribe it on the basis of introspection.

Neural Thinker. I-thought reflexively refers to its thinker qua neural thinker, or the subject whose brain or brain-parts physically sustain the thought episode.

Author. *I*-thought reflexively refers to its thinker qua author, or the subject who determined the intentional content of the thought at the personal-level.

In ordinary conditions of thinking, there is no practical or theoretical gain to be had in distinguishing these formulations of TRR – typically, the author, neural substrate and entertainer of a given *I*-thought are combined in one individual, so in normal circumstances is no predictive or explanatory difference on the scene to tell them apart. The crucial point for our purposes, however, is that there is no obvious conceptually necessary alignment between them. At least on paper, each of these formulations of the TRR alights upon a conceptually distinct relation we each normally bear to our own thoughts. Who's to say that they can't come apart?

As it turns out, we already have a case in hand to demonstrate this conceptual possibility of dissociation – in the next section, I turn back to our q-memory case, and survey the reference-assignments for the use of the first person concept in that case under each of these TRR disambiguations.

3. Three reference assignments in *Memory Transplant*

The question that faces us now is, who is the referent of the first person contents in *Memory Transplant* under each of the three disambiguations of the TRR?

The *Entertainer* disambiguation of the TRR is the easiest to settle. Ex hypothesi, Larry, the later q-rememberer, is the one enjoying memory impressions of being in the Sagrada Familia in his introspectively accessible stream of consciousness. As such, Larry alone is positioned to immediately self-ascribe those contents on the basis of introspection. According to *Entertainer* this makes him the referent of the *I*-thought. For many, this is the most pretheoretically intuitive assignment of reference between the three – and presumably, the assignment Evans, and others following him, have had in mind.

Take *Author* next. In *Memory Transplant*, the intentional content of the memory is determined by a memory trace encoded at the earlier time by Earl's experience of being in the Sagrada Familia. Larry isn't even on the scene at this point. So again, it seems fairly straightforward to say that so long as we have the *Author* interpretation of the TRR before us, Earl is the *I*-thought's unique referent.

A bit too straightforward perhaps. This way of describing the case employs the simplifying fiction that the memory impressions entertained at the later time are fully determined by the memory trace encoded at the time of the earlier experience, much like saving a document to a floppy disk on one computer and opening it later on another. But of course, episodic memory is not really like this. The empirical sciences involved in memory

studies have long since converged on the view that episodic memory is a largely reconstructive process.¹⁵ Remembering, it is agreed, is an active process on the part of the subject in which fragmentary information encoded in memory traces laid down at the earlier time – often only preserving what has come to be called the “*gist*” of the past event – is combined with independent information gained in the intervening time, and filtered through all sorts of influences at the later point of recall. Research on these influences on memory has exploded in the time since this reconstructive model first emerged in the 1930s¹⁶; to give a flavor of their scale, a passage from John Sutton records some of those we know about so far:

There are hormonal and neuromodulatory, genetic and pharmacological, developmental and age-related influences; there are influences of arousal, stress, gender, mood, emotion, sleep and personality; there are unconscious, schematic or semantic influences, and there are influences of context, situation, task and environment. (2011: 355)

Indeed, the determinative power of these factors is so strong that Sutton suggests that it is wrongheaded to think of these factors as influencing an independently existing memory at all; ‘remembering itself *is* the shifting and fragile resultant or vector of inner and outer influences operating and interacting at distinct levels and timescales. There need be no user or consumer of these influences: it could be influences all the way down.’ (Sutton, 2011, p. 357) We needn’t go as far as Sutton on this point, and most memory scientists still allow a role for the relevant contributory memory traces to the determination of the final memory’s content. But, Sutton’s suggestion serves to bring out an important feature of the reconstructive model. These influences are not only operative in defective episodes of remembering, but in all cases. The process of remembering is *inherently* reconstructive, and that reconstruction is *always* largely determined by factors of this kind.

How does this move to the reconstructive model of memory affect the interpretation of our case of q-memory? It certainly suggests we should treat Larry, the later rememberer, as a contributor to what is remembered along a number of potential dimensions. Strictly, it is false to say that the full intentional content of the memory, in all its detail, is entirely determined by the memory trace encoded at the earlier point of the experienced event. This is an interesting and underexplored aspect of the q-memory debate in its own right. But, in this case, it needn’t change our minds about who is the thought’s thinker according to *Author* — who determines the intentional content at the personal-level. That’s because what we’re interested in here is the determination of the intentional content endorsed in the propositional judgment *I was in the Sagrada Familia*, the candidate target of an IEM

claim. Whatever details or distortions are introduced in the gap between the experience-event and the remembering-event, or during Larry's later reconstructive act of recall, the determination of the propositional content of the memory *I was in the Sagrada Familia* is entirely counterfactually dependent on Earl's earlier experience and the memory trace that causally resulted from it – none of *this* content came from Larry.¹⁷ Even accepting these interesting complications arising from reverting to a more empirically adequate model of memory, then, Earl is still assigned as thinker-reflexive referent under *Author*.

Now, we get to the really interesting reference assignment. It's much harder to say who the thought's thinker is in terms of being the subject whose brain is the thought's neural substrate. At first pass it might seem obvious that it should be Larry. After all, he is the subject in whose head the brain-parts are being activated, so surely he is the subject whose brain is physically sustaining the conscious episode in question.

This, however, would be too quick. The description of the transplant above does not involve injecting some un-physically-realized content into Larry's brain, a tiny slip of a ghost inserted into a new machine. What happens, rather, is a splicing together of (parts of) two thinking apparatuses from distinct thinkers; a collection of bits of Earl's brain get grafted onto the relevant areas of Larry's, so that what we end up with is a functionally integrated patchwork of neural matter coming from them both. The reason this is crucial to the description of the case is because it is now a simple matter to trace a causal path between Earl's earlier experience and Larry's later memory impressions, without which it couldn't count as a case of q-memory.¹⁸ Let's add the plausible assumption that it is these transplanted bits of neural matter comprised of specialized trace-cells that are dominantly physically sustaining the q-memories.¹⁹ To know who their producer is, we must now settle a metaphysical question about who these brain-parts belong to.²⁰

I can see at least three ways of going. Under the first, Larry is a thinker whose brain now incorporates secondhand parts donated by Earl. Larry's head, the idea might be, so Larry's brain. Earl is a distinct thinker, whose thinking-organ is now smaller than it used to be. This option is likely to appeal to those who think that physical contiguity or functional integration is what matters for the individuation of organs. Given that the parts of the hodgepodge neural object causally sustaining the conscious memory impressions are Larry's – as *all* of its parts now are – on this reading he is the neural thinker of these contents.

A second option agrees with the first that there are two distinct brains here, but disagrees about which brain the parts in question are parts of. Under this second reading they are still strictly Earl's: his brain-parts have simply been rehoused in Larry's skull. This is similar to the way you

might use another's body to temporarily host and regenerate parts of your liver-tissue without thereby making it theirs, or the way a gestating mother contains the body-parts of her fetus within her bodily boundaries. (To encourage this reading of the case we might add the further detail that Earl has only consented to a short-term loan of these brain-parts, which will be returned to him as soon as the philosophical experiment is over.) This would make Earl's brain a kind of spatially distributed object – a bizarre result, but no more bizarre than other features of this case. On this second reading, Earl is the contents' thinker *qua* neural thinker. This option will likely appeal to those with ontogenetic or etiological intuitions about organ individuation, those moved by the thought that Earl is the one who grew this neural matter in his body, who fixed the neural grooves and built up the synaptic pathways. This makes them forever his, even if he consensually makes them available for prosthetic use by someone else.

On the other hand, by now we might suspect that even if we have ways of massaging our intuitions one way or the other, they can't be made to bear any weight. Perhaps, there *isn't* a single correct overarching answer about who these brain-parts uniquely belong to. This brings us to a third reading of the scene, on which neither subject has unique claim to being the thought's neural thinker. This reading will appeal to those who think that the details of the case corrupt our ordinary ways of individuating brains so much that there's no longer a single right answer to the question of whose brain these brain-parts are part of.

It seems to me that this third reading fits well with our intuitions about other kinds of organ transplant. It is compelling to think that there *is* no single right answer about whether a transplanted heart is still the donor's heart in a new body – a way of keeping part of a loved one alive – or whether it now counts unambiguously as the recipient's. Perhaps, the suggestion goes, there is likewise no single right answer whether the brain-parts in the q-memory case belong to Earl or Larry. Or maybe we should say that both have equal claim to those brain-parts, just as it might be argued that a transplanted heart is now *both* the recipient's and the donor's. The q-memory case would now involve overlapping thinking-organs whose shared regions belong to them both. (We might describe Earl and Larry on this reading as a spatially scattered version of craniopagal twins.) Whether they both have claim to these parts, or whether it is indeterminate whose parts they are, the third option is that there is no unique owner of the brain-parts, so no unique referent of the first person contents under *Neural Thinker*.

I leave it open which of these ways of attributing brain-parts is correct. It is enough for our purposes that the answer isn't obvious. This means that under *Neural Thinker* it turns out to be non-obvious whether the first

person contents in *Memory Transplant* uniquely refer to Earl, Larry, or fail to uniquely refer to anyone at all.

Let's take stock. First person contents refer reflexively to their thinkers. To give this up would be to abandon a central and plausible tenet of an orthodox view of first person reference. What the arguments of §2 brought out, however, is that being a content's thinker is a multiply ambiguous notion, in a way that is normally hidden from view by the alignment of the disambiguations in typical conditions of thought. In this section, I have argued that in the highly artificial case of q-memory, how the reference rule is disambiguated makes a difference to who is assigned as the thinker-reflexive referent. In our case, Larry is the thinker of these memory impressions *qua* entertainer, Earl is their thinker *qua* author, and it is unclear which (if either) of them counts as unique thinker *qua* neural thinker. How, now, does this connect with our opening discussion about the IEM of memory judgments?

4. IEM explanations

Here is a restatement of the conditions under which judgments based on memory impressions are IEM:

Memory-IEM: A judgment *I was F* formed on the basis of memory-impressions of being *F* is IEM relative to the use of the first person concept iff it cannot be mistaken in the following way: the subject knows, on that basis, that *someone* was *F* (*de re* or existential) but has made a mistake only in misidentifying the person that they thereby know to be *F* with the person referred to by their use of *I*.

I have been arguing that the first person contents in q-memory cases are referentially underdetermined by the TRR. So long as we take it that these contents are what are endorsed in Larry's first personal judgment made on their basis, this results in the claim that the use of the first personal concept in that judgment is likewise referentially underdetermined. Putting this together with our starting discussion about the IEM of memory judgments, it looks like the question whether q-memories involve errors of misidentification – and so whether they really constitute a threat to the claimed IEM of memory judgments – depends on how this underdetermination question is resolved.

Suppose first that the use of *I* in the q-memory judgment *I was in the Sagrada Familia* refers to Larry, the transplant recipient. This might be because the TRR is disambiguated in favor of the *Entertainer* reading, or the *Neural Thinker* reading under the first way of settling the metaphysical brain-individuation question. The judgment under this reference assignment is mistaken – ex hypothesi, Larry has never been to Barcelona. But, it's not mistaken because of an error of misidentification.

The grounds for this judgment never gave Larry warrant to judge anything but a first person judgment about himself; Evans, in other words, was right on this reading that q-memory is not a systematic source of knowledge for Larry about someone else's past, but a systematic source of illusion about how things were in his past. His mistake is a commonplace error of mispredication (albeit an easy one to make in the circumstances). He got it wrong, for good reason, that he was once in the Sagrada Familia.

On this first reference assignment, then, *Memory Transplant* does not undermine the IEM of memory judgments, because it does not provide an example of a memory-impression based judgment that is in error through misidentification. The explanation of the preserved immunity-claim is just the one described in the introduction. Even in such causally deviant cases as this one, there is no identification involved in the judgment's grounds. Memory impressions are an identification-free form of self-knowledge, a way of knowing about oneself *as oneself*.

The second option from §3 was that the use of *I* in the q-memory judgment refers to Earl, the memory-trace donor. This would result from either the *Author*-disambiguation of the TRR, or the *Neural Thinker*-disambiguation under the second way of counting brains. In this case, of course, the judgment isn't in error at all – Earl really was once in the Sagrada Familia, looking up at the colorful windows. And if there is no error here, *a fortiori* there is no error of misidentification.

This second assignment, then, again returns the verdict that *Memory Transplant* is not a problem for the claimed IEM of memory judgments. This time, however, what explains the immunity isn't exhausted by the Evansian claim that memory is identification-free. Under this reading, rather, the challenge is eliminated by making it the case that the first person contents refer not to the later judger, but to the q-memory's source. So under this reading, although identification-freedom is part of the explanation of the presence of IEM, it's not the whole of it – it is also partly explained by the possibility of thinker-reflexive-reference-at-a-distance.²¹ What cases like *Memory Transplant* now show, strikingly, is not only that memory is an identification-free form of self-knowledge but also that it is possible to have first personal memory impressions occurring in one's stream of consciousness, and first person judgments formed on their basis, that thinker-reflexively refer to someone else.²²

The third referential assignment (or rather, *non*-assignment) is the most intriguing; viz., that the first person contents of q-memory cases don't uniquely refer to anyone – perhaps the reference is split between the two thinkers, or maybe the thought fails of reference altogether. This would correspond to the third way of individuating brains considered on the *Neural Thinker*-disambiguation of the TRR. What should we now say

about the error-profile of Larry's q-memory judgment *I was in the Sagrada Familia*?

The judgment is now in error, but not through a misidentification. Here's why. An error of misidentification was defined at the start of this section as requiring two things: (i) the grounds in question must supply warrant for *some* knowledge claim – either existentially general knowledge that the property is instantiated, or *de re* knowledge that it is instantiated in a particular person – and (ii) the final judgment must be in error only because its singular component picks out either the wrong witness to that existential, or picks out a different person from the person in whom the property is known on those grounds to be instantiated.²³

Now, assuming the no-unique-reference verdict to our referential underdetermination problem, how does the q-memory judgment fare with respect to these conditions? There's no reason to think it fails the first condition – the q-memory impressions really do seem to supply warrant for a knowledge-claim that someone was in the Sagrada Familia, whether understood existentially or *de re*. It *does*, however, look to fail the second requirement. That's because the second requirement involves a presupposition of successful singular reference: it's built in to the requirement that the singular component of the judgment refers to someone other than the "someone" relevant to the warrant mentioned in the first requirement. But, of course, we're currently considering the verdict that the q-memory judgment fails this presupposition. So there can be no error of misidentification under this reading. In brief: error through misidentification presupposes singular reference, so where there is no singular reference there can be no error through misidentification.

This is a surprising source for a verdict of IEM. Logical IEM is a negative modal property; it holds just in case the possibility of error through misidentification is ruled out.²⁴ If it turns out that q-memory cases fail the referential presuppositions of that kind of error, then the possibility of that kind of error *is* ruled out, and the IEM of memory judgments is preserved. But in this case, it isn't because of the identification-freedom of the judgment's grounds, but because those judgments are semantically defective in a way that prohibits the possibility of error through misidentification in the first place. (From here, whether we say that Larry's faulty judgment is in *error* or defective in some other way depends on quite general questions about whether propositions of the form *a is F*, in which *a* has empty reference, should be treated as false or lacking a truth value.)

The aim of this section was to review the implications of each of the three possible first personal reference assignments in *Memory Transplant* for the question whether that case undermines the putative IEM of memory judgments. We leave the exercise with a surprising result. It turns out that the IEM of memory judgments is preserved on all three reference assignments,

but that's not the surprising thing. The surprising result is that *the explanation for the presence of IEM is different each time*. If this is right, then we should be cautious about accepting claims about a uniform significance to the fact that a given kind of judgment has IEM.²⁵ This result stands independently of the plausibility of the different descriptions of the case study used to reach it. *If* our example of q-memory was understood in each of these different ways, the presence of IEM would receive these different explanations. That is all I need to make the broader point that there can be different operative explanations of IEM, and so different corollary significance to its presence (or absence) in different cases.

5. conclusion

There is still an outstanding matter to be settled. The argument presented in this paper for pluralism about explanations of IEM made use of a hidden referential ambiguity in the TRR governing the first person concept. This is a bold claim about the first person concept that can't be left as it stands. How, when all the chips are in, *should* we disambiguate the TRR governing the first person concept?

We've already seen that the TRR by itself won't settle it. Perhaps there are theoretical resources external to the rule that could provide guidance about how to proceed: independently supported knowledge/truth/rationality-maximization principles, teleosemantic considerations, or perhaps evolutionary-biological claims about the adaptive function of memory.²⁶ A quite different way ahead, however, is one that resists the idea that there is a single disambiguation of the TRR that could accurately track stable and general truths about first person thought. Our first person concept just isn't more determinate than the TRR as it stands, in all its ambiguous richness.

I think we have reason to favor this non-reductive option, and I want to end by briefly saying something about why. The first person concept is not a theoretical term of art, subject to theory-led refinements and stipulations. The first person concept, rather, is a way that we in fact think about ourselves, in the ordinary conditions of thinking that we in fact find ourselves in. Moreover, it is one that plays a central and ineliminable role in our cognitive economies just as it is. Now, importantly, the contexts in which we ordinarily think of ourselves this way – contexts in which we develop and deploy our first person concepts – never normally call for finer-grained disambiguations of the TRR. So if we're pursuing a deeper understanding of the nature of first person thought understood as a natural semantic kind, as a way we in fact think about ourselves, then the fact of the matter is that there *is* no fact of the matter about which of these disambiguations captures our "real" or "underlying" mode of first person thought. All three thinker-roles identified in [Section 2](#), and surely others

too, are equal contributory strands to the TRR governing our use of *I*, and they cannot be isolated, disentangled or ranked without loss.

This view leaves us with an open question about what to say about the highly unusual cases in which the different thinker-roles ordinarily enmeshed in its reference-rule come apart like *Memory Transplant*. One is tempted to say that strictly speaking there is indeterminacy of reference here – the highly artificial conditions of thinking described simply don't deliver a determinate reference-assignment for the TRR, which was not developed for use under such conditions. But, of course, this is perfectly compatible with taking pragmatic cues on board as to the most intelligible or intuitive assignment in context. If this is right, then by playing around with the pragmatic settings of q-memory cases (as, of course, I did in [Section 3](#)) we ought to be able to manipulate our referential intuitions one way or the other.

This sketch of a non-reductive view of the first person concept stands apart from the central argument of this paper. With it, the upshot of the paper is that *even on a simple TRR view of the reference determination of first person thought*, we can retain a psychologically plausible degree of fuzziness in the way we model these thoughts, and that this fits together nicely with a newly argued pluralistic case for what explains first person judgments with IEM. Without it, the argued claim is that there is no single fact of the matter about what explains the failure of q-memory cases as a threat to the IEM of memory judgments, and so no single fact of the matter about what the significance of that claim is. Evans was right in his verdict that the threat fails, but wrong about why it matters.

Notes

1. There is a question whether IEM is an epistemic property that can only apply to knowledgeable judgments, or also to warranted/justified judgments; this is not an issue that affects the arguments of this paper, so I will leave it to one side and simply talk in terms of knowledgeable judgments. Thanks for an anonymous reviewer for raising this point.
2. *Episodic* memory, in which a particular past experienced event is represented through experiential or quasi-experiential reconstruction, contrasts with *semantic* memory in which a propositional fact is retained over time (cf. Tulving, 1972, 1985); see Mahr and Csibra (2018) for a recent overview of episodic memory in the psychological literature. By “memory” throughout I mean “episodic memory”.
3. The idea that IEM marks out forms of self-awareness through which we have a distinctively subjective perspective on ourselves can be traced back to Wittgenstein's discussion of the use of *I as-subject* (Wittgenstein, 1958, p. 67) and to both Shoemaker and Evans' early formulations of the property (Evans, 1982:220; Shoemaker, 1969, p. 561).
4. See Evans (1982, pp. 180–1) for his discussion of identification-freedom and IEM, and 1982:235–48 for his discussion of memory and quasi-memory.

5. This formulation stays close to Shoemaker's original presentation (1968:556–7), but is relativized – as is standard – to judgmental grounds, and presented in terms of judgments rather than statements. The “someone” is intended to be inclusive of both sides of Pryor's *wh-/de re* distinction (Pryor, 1999) – that is, either as an existential, or as a *de re* identification of an individual. This is an issue that won't matter much for our purposes. The judgment's grounds should be taken to include background presuppositions (cf. Coliva, 2006; Wright, 2012).
6. It is a mark in favor of this conceptual claim that in his original definition of episodic memory, Tulving (1972) defines episodic memory as memory for personally experienced past events, a feature that continues to characterize episodic memory across otherwise differing accounts of episodic memory in the psychological literature (see, e.g., Mahr & Csibra, 2018 and references contained therein).
7. Shoemaker (1970, p. 271).
8. See e.g., Pryor, 1999, p. 296 and Coliva (2006, pp. 406–407).
9. See, e.g., (Dokic, 2001; Klein, 2013, 2014, 2015; Klein & Nichols, 2012; Perner, 2001; Perner & Ruffman, 1995; Perner et al., 2007).
10. Evans' own view of first person thought involves a mode of presentation that is in part individuated by a specified set of information links to its referent – a sort of “demonstrative view” of first person thought, as it is often called (1980: Chap 7). This is a marginal view in the contemporary debate, so I put it to one side in what follows.
11. Is this incompatible with a Fregean framework? Not in principle, so long as we accept Evans' moral in “Understanding Demonstratives” (Evans, 1981) that there can be context-sensitive Fregean senses.
12. These quotes are for the first person pronoun, but O'Brien is explicit that she intends her account to extend to first person reference in both thought and language; see (2007) pp. 66–7. The second quote is her interpretation of the thinker-reflexive rule as given in Peacocke.
13. This way of describing thought insertion cases is a matter of controversy (see e.g., Bortolotti & Broome, 2008, Billon, 2013, Coliva, 2002), but nothing in what follows depends on this being the correct account of the delusion.
14. See Palmira (2022) for another example of someone doing something similar with the token reflexive rule for a different theoretical purpose.
15. The canonical source for the reconstructive model is Bartlett (1932).
16. See Mahr and Csibra (2018), p. 5 for a survey of recent experimental findings on top-down influences on memory reconstruction.
17. The classic readings on the possibility of introducing misinformation into a memory on this reconstructive model are from Elizabeth Loftus (e.g., Loftus, 1975, 2005).
18. Q-memory was defined as the same as ordinary memory in all respects except its neutrality about the identity of the earlier experiencer and later rememberer (p. 3); a suitable causal link between the earlier experience and later memory is a standard condition on ordinary memory (cf. Martin & Deutscher, 1966).
19. I bracket the complication of how exactly this transplant would go if we think of the memory trace as taking the form of distributed synaptic structures; to deny the intelligibility of the transplant would be to dismiss these cases too soon.
20. See also Schechtman (2014) Chapter 6 for an excellent extended discussion of the ways in which different descriptions of classic puzzle cases can pull our ownership intuitions in transplant cases in different directions.
21. See (Madden, 2011) for a different argument for this possibility.

22. This really is striking; it allows, contra Campbell (1987), for failures of trading on the identity of the referent of “I” in inferences.
23. The disjunction is included to accommodate Pryor’s (1999) *de re/wh*- divide.
24. Or, sufficiently modally safe (McGlynn, 2016).
25. This finding has natural allies in Peacocke and Campbell who likewise deny the univocity of IEM explanations (Campbell, 1999a; Peacocke, 2008); its natural opponents posit a univocal explanation for all instances of IEM (Morgan, 2012; Wright, 2012).
26. See Mahr and Csibra (2018), pp. 6–8 for a survey of some of the leading views of the adaptive function of episodic memory in the current psychological literature.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.
- Billon, A. (2013). Does consciousness entail subjectivity? The puzzle of thought insertion. *Philosophical Psychology*, 26(2), 291–314. <https://doi.org/10.1080/09515089.2011.625117>
- Bortolotti, L., & Broome, M. (2008). A role for ownership and authorship in the analysis of thought insertion. *Phenomenology and the Cognitive Sciences*, 8(2), 205–224. <https://doi.org/10.1007/s11097-008-9109-z>
- Campbell, J. (1987). Is sense transparent? *Proceedings of the Aristotelian Society*, 88, 273–292. <https://doi.org/10.1093/aristotelian/88.1.273>
- Campbell, J. (1999a). Immunity to error through misidentification and the meaning of a referring term. *Philosophical Topics*, 26(1), 89–104. <https://doi.org/10.5840/philtopics1999261/237>
- Campbell, J. (1999b). Schizophrenia, the space of reasons and thinking as a motor process. *The Monist*, 82(4), 609–625. <https://doi.org/10.5840/monist199982426>
- Coliva, A. (2002). On what there really is to our notion of ownership of a thought. *Philosophy, Psychiatry, & Psychology*, 9(1), 41–46. <https://doi.org/10.1353/ppp.2003.0005>
- Coliva, A. (2006). Error through misidentification: Some varieties. *The Journal of Philosophy*, 103(8), 403–425. <https://doi.org/10.5840/jphil2006103824>
- Dokic, J. (2001). Is memory purely preservative? In C. Hoerl & T. McCormack (Eds.), *Time and memory: Issues in philosophy and psychology* (pp. 213–232). Oxford University Press.
- Evans, G. (1981). Understanding demonstratives. In H. Parret & J. Bouveresse (Eds.), *Meaning and understanding* (pp. 280–304). W. de Gruyter.
- Evans, G. (1982). *The varieties of reference* (J. McDowell Ed.), OUP.
- Klein, S. B. (2013). Making the case that episodic recollection is attributable to operations occurring at retrieval rather than to content stored in a dedicated subsystem of long-term memory. *Frontiers in Behavioral Neuroscience*, 7, 1–14. <https://doi.org/10.3389/fnbeh.2013.00003>
- Klein, S. B. (2014). Autonoesis and belief in a personal past: An evolutionary theory of memory indices. *Review of Philosophy and Psychology*, 5(3), 427–447. <https://doi.org/10.1007/s13164-014-0181-8>

- Klein, S. B. (2015). What memory is. *WIREs Cognitive Science*, 6(1), 1–38. <https://doi.org/10.1002/wcs.1333>
- Klein, S. B., & Nichols, S. (2012). Memory and the sense of personal identity. *Mind*, 121(483), 677–702. <https://doi.org/10.1093/mind/fzs080>
- Lofthus, E. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, 7(4), 560–572. [https://doi.org/10.1016/0010-0285\(75\)90023-7](https://doi.org/10.1016/0010-0285(75)90023-7)
- Lofthus, E. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory: Figure 1. *Learning & Memory*, 12(4), 361–366. <https://doi.org/10.1101/lm.94705>
- Madden, R. (2011). Brain transplants and externalism. In K. Bennett & D. Zimmerman (Eds.), *Oxford studies in metaphysics* (Vol. 6, pp. 287–316). OUP.
- Madden, R. (2016). Thinking parts. In S. Blatti & P. Snowdon (Eds.), *Essays on animalism* (pp. 180–207). OUP.
- Mahr, J., & Csibra, G. (2018). Why do we remember? The communicative function of episodic memory. *Brain and Behavioural Sciences*, 19, 1–93. <https://doi.org/10.1017/S0140525X17000012>
- Martin, C. B., & Deutscher, M. (1966). Remembering. *The Philosophical Review*, 75(2), 161. <https://doi.org/10.2307/2183082>
- McGlynn, A. (2016). Immunity to error through misidentification and the epistemology of *De Se* thought. In M. García-Carpintero & S. Torre (Eds.), *About oneself* (pp. 25–55). Oxford University Press.
- Morgan, D. (2012). Immunity to error through misidentification: What does it tell us about the *de se*? In S. Prosser & F. Recanati (Eds.), *Immunity to error through misidentification* (pp. 104–124). CUP.
- Morgan, D. (2015). The demonstrative model of first person thought. *Philosophical Studies*, 172(7), 1795–1811. <https://doi.org/10.1007/s11098-014-0391-z>
- O'Brien, L. (2007). *Self-knowing agents*. OUP.
- Palmira, M. (2022). Questions of reference and the reflexivity of first-person thought. *The Journal of Philosophy*, 119(11), 628–640. <https://doi.org/10.5840/jphil20221191140>
- Peacocke, C. (2008). *Truly understood*. OUP.
- Peacocke, C. (2014). *The mirror of the world: Subjects, consciousness, and self-consciousness*. Oxford University Press.
- Perner, J. (2001). Episodic memory: Essential distinctions and developmental implications. In C. Moore & K. Lemmon (Eds.), *The self in time: Developmental perspectives* (pp. 181–202). Routledge.
- Perner, J., Kloos, D., & Stöttinger, E. (2007). Introspection & remembering. *Synthese*, 159(2), 253–270. <https://doi.org/10.1007/s11229-007-9207-4>
- Perner, J., & Ruffman, T. (1995). Episodic memory and auto-noetic consciousness: Developmental evidence and a theory of childhood amnesia. *Journal of Experimental Child Psychology*, 59(3), 516–548. <https://doi.org/10.1006/jecp.1995.1024>
- Pryor, J. (1999). Immunity to error through misidentification. *Philosophical Topics*, 26(1), 271–304. <https://doi.org/10.5840/philtopics1999261/246>
- Schechtman, M. (2014). *Staying alive: Personal identity, practical concerns, and the unity of a life*. OUP.
- Shoemaker, S. (1969). Self-reference and self-awareness. *The Journal of Philosophy*, 65(19), 19. <https://doi.org/10.2307/2024121>
- Shoemaker, S. (1970). Persons and their pasts. *American Philosophical Quarterly*, 7(4), 269–285.
- Sutton, J. (2011). Influences on memory. *Memory Studies*, 4(4), 355. <https://doi.org/10.1177/1750698011414082>

- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. Academic Press.
- Tulving, E. (1985). *Elements of episodic memory*. OUP.
- Wittgenstein, L. (1958). *The blue and Brown Books*. Blackwell Publishing.
- Wright, C. (2012). Reflections on Francois Recanati's 'immunity to error through misidentification: What it is and where it comes from. In F. Recanati & S. Prosser (Eds.), *Immunity to error through misidentification* (pp.247–280). CUP.