



Is conscious thought immune to error through misidentification?

Manuel García-Carpintero

LOGOS/BIAP-Departament de Filosofia, Universitat de Barcelona, Barcelona, Spain

ABSTRACT

Wittgenstein distinguished between two uses of “I”, one “as object” and the other “as subject”, a distinction that Shoemaker elucidated in terms of a notion of *immunity to error through misidentification* (“IEM”); first-personal claims are IEM in the use “as subject”, but not in the other use. Shoemaker argued that memory judgments based on “personal”, *episodic* memory are not strictly speaking IEM; Gareth Evans disputed this. Similar issues have been debated regarding self-ascriptions of conscious thoughts based on first-personal awareness, in the light of claims of “thought insertion” in schizophrenic patients. The paper aims to defend a Shoemaker-like line by critically engaging with some compelling recent contributions. Methodologically, the paper argues that to properly address these issues the all-inclusive term “thought” should be avoided, and specific types of thoughts countenanced.

ARTICLE HISTORY

Received 1 February 2024
Accepted 29 April 2024

KEYWORDS

Self-conscious thought; first-personal reference; *de se attitudes*; self-knowledge; immunity to error through misidentification

1. Preamble: immunity to error through misidentification and self-reference

Wittgenstein (1958, 66–67) distinguished between two uses of “I”, one “as object” and the other “as subject”, a distinction that Shoemaker (1968) elucidated in terms of a notion of *immunity to error through misidentification* (“IEM” henceforth, also for *immune* . . .).¹ Shoemaker (1968, p. 557) specifies the sort of mistake that IEM excludes as follows: “to say that a statement ‘*a* is φ ’ is subject to error through misidentification relative to the term ‘*a*’ means that the following is possible: the speaker knows some particular thing to be φ , but makes the mistake of asserting ‘*a* is φ ’ because, and only because, he mistakenly thinks that the thing he knows to be φ is what *a* refers to”. Shoemaker (1970, p. 270) offers an illustration involving memory: “if I claim on the strength of memory that I saw John yesterday, and have a full and accurate memory of the incident, it cannot be the case that I remember someone seeing John but have misidentified that person as

myself; my memory claim ‘I saw John’ is subject to error through misidentification with respect to the term ‘John’ (for it could have been John’s twin or double that I saw), but not with respect to ‘I’. The error can also affect “I” if, e.g., I base the claim on a photograph of what I take to be me looking at John. I may know on that basis that someone is looking at John but misidentify him as me. This would be a contrasting use of “I” “as object”.

At first sight, the self-ascription of conscious thoughts (*I am thinking that I should murder Lissi*) made based on the personal access that underwrites introspection (here called *subjectivity*, Section 3) appears intuitively to be IEM. However, Campbell, Sugden (1999) argued that cases of thought insertion in schizophrenia patients show that they are not IEM, on a particular interpretation of what the relevant self-ascription amounts to along the lines of earlier suggestions by Stephens and Graham (1994), Campbell (2002) and Coliva (2002a, 2002b) debated the issue.

In this paper I want to develop an account of these cases that I have briefly suggested elsewhere (García-Carpintero, 2016, 2018), which elaborates on and defends Campbell’s view. I rely on an account of IEM that differs from Coliva’s and Campbell’s, which I’ll outline in this and the following section. My proposal strengthens distinctions that Campbell drew, by relying on recent debates on inner speech and conscious thought. I won’t critically engage here with Coliva’s or Campbell’s views; I hope that my proposal makes sufficient sense in its own terms and is thus a helpful addition to their philosophical discussion. The paper emphasizes the methodological point that to properly address these issues the all-inclusive term “thought” should be avoided, and specific types of thought countenanced.

Shoemaker defines IEM for linguistic acts, while the phenomenon I am interested in concerns mental attitudes – occurrent conscious thoughts like propositional imaginings and judgments. However, as I’ll explain in more detail in Section 3, here I’ll understand judgments as inner assertions – assertions in inner speech, on the assumption that inner speech, at least in some cases, is actual speech as opposed to merely imagined speech (Gregory 2016; Roessler 2016; Kompa 2023). Shoemaker’s characterization is thus good enough as a starting point. IEM is a phenomenon discerned by philosophers; however, it is intuitive enough for us to think of it on the well-known Kripke-Putnam model for *water*, hypothesizing that it targets a *real* psychological kind with an explanatory real essence. Following Putnam,² I’ll assume that central instances of “IEM” are to be identified for theoretical debate roughly by Shoemaker’s characterization in the quotations above, shared by philosophers who otherwise defend different accounts of the phenomenon.

I’ll elaborate on Shoemaker’s characterization by relying on Seeger’s (2015b) definition of *error through misidentification*, “EM” henceforth. The *source object* in a judgment of the form *a is F* is *the object from which*

the predication content *F* derives – the one that epistemically grounds it, if any. The *target object* is the object *a* to which that content is ascribed, if any. The definition of EM, then, is this:

EMA judgment *a is F* is in error through misidentification iff the source object is different from the target object.³

The core idea – generalizing what Pryor (1999, p. 296) says about memory – is that in *EM* cases we can tease apart, and retain, the justification for the existential claim, *that something is F*; it is independent from the justification for the singular claim, *that a is F*, which might be false or unjustified. In *IEM* cases, in contrast, the evidence gives us an indissoluble “package deal” – indicating both that something *is F*, and that *it is a which is F* together with it.⁴ Shoemaker articulates and illustrates this “package deal” metaphor thus: “[...] in being aware that one feels pain one is, tautologically, aware, not simply that the attribute *feel(s) pain* is instantiated, but that it is instantiated *in oneself*” (Shoemaker 1968, 563f.; his emphases).

If a judgment made on grounds *E* is liable to EM, I will say that it is *vulnerable to error through misidentification* (“VEM” henceforth, also for *vulnerability* . . .), relative to *E*; otherwise, it is *IEM* relative to *E*. Shoemaker (1970) distinguishes two kinds of *IEM* depending on how strongly we interpret these modalities. If it is just conditions in the actual and nearby worlds that rule out the possibility that the mistake specified above will ever happen, then the judgment is *de facto IEM* relative to *E*. If the conditions in every logically possible world rule out the mistake, then the judgment is *logically IEM* relative to *E*.

Recanati (2007, 2009, 2012) argues for a Lewisian account of “basic” or “implicit” *de se* thoughts, on which – along the Wittgensteinian lines argued for by Anscombe (1975) – the first-person doesn’t make a referential contribution to such thoughts. He contends that this provides a good theoretical explanation of the *IEM* datum: no reference, hence no need for a potentially misguided identification. Recanati takes his proposal to be a particular case of Evans’s (1982, pp. 180–181) “Simple Account” of *IEM* that Coliva also adopts, on which VEM judgments depend on an “identification” of the referent, while *IEM* judgments are identification-free.⁵

I have argued however that both the *no-reference* view and the more general *Simple Account* fail to explain *IEM*, as characterized above (García-Carpintero 2015, 2016, 2017, 2018). A main argument against the first is that there are examples involving indexicals other than “I” for which the view that they don’t refer is implausible, which nonetheless appear to be *IEM*: “you are standing very close”, “he is a long way off”, both based on perception, Wright (2012); “this keyboard is black”, Peacocke (2008), also based on perception.⁶ These cases also challenge the more general Simple Account because they don’t appear to be “identification-free” but

presuppose an “identification” of the referents. There also are judgments that don’t appear to rely on any identification that seem nonetheless to be VEM as defined above: Pryor’s (1999) cases of *wh-misidentification*, in which the subject makes a justified existential judgment (*someone drank my beer*), without committing to there being a unique witness, and goes on to misidentify one (*Alex drank my beer*).

In the initial quotation Shoemaker (1970) takes the memory-based claim “I saw John yesterday” to be IEM with respect to “I”; but he argued that it is merely *de facto* IEM. Evans (1982) contested this view.⁷ The way I understand Campbell, Sugden (1999), he argues that cases of thought insertion show that self-conscious thoughts are also not logically IEM, leaving open that they may be *de facto* IEM. The early debate about memory concerned science-fiction cases of brain-splitting and brain-transplants. Recent discussions depart instead from the “reconstructive” character of episodic memory (Section 2), illustrated by the fact that some of our (episodic) memories are “observer” or “third-person”, as opposed to “field” or “first-person” (Rice, 2010). In the latter, we remember ourselves occupying the “origin of perspective” from which we perceived the apparently remembered scene. In the former, we recall ourselves as a participant in that scene. In a paper whose setup overlaps with this (García-Carpintero, *in press*) I argue that this reconstructive character of memory offers ordinary examples that vindicate the Shoemaker line that memory claims are only *de facto* IEM. Here I’ll argue that, once we make distinctions that Campbell advances on how the “owner” of the attitude is to be understood, cases of thought insertion – and *auditory verbal hallucinations*, I’ll lump them together, cf. Sousa and Swiney (2013, p. 648), Langland-Hassan (2016, p. 676) – help us to make a compelling case for the merely *de facto* character of the IEM of some self-conscious thoughts. There are analogies between observer memories and schizophrenic symptoms relevant here: both present their targets from “external” perspectives that make it coherent to question self-identifying with them.

The hypothesis that IEM is a natural psychological kind whose pre-theoretical makeup is best captured by the characterization above is crucial for my argument. Theoretical proposals like the no-reference or the more general Simple Account, I argued, fail to explain the phenomenon and thus do not aptly define the kind. I favor instead the view that Coliva & Palmira (this volume-b, Section 1) aptly call *Metasemantic* account – cf. García-Carpintero (2015, 2016, 2017, 2018), Palmira (2020, 2022), Verdejo (2021) for related but diverse elaborations. Let me just sketch here the guiding idea. Consider *Julius invented the zip*, assuming that the referent of “Julius” is stipulated to be *the inventor of the zip*. This looks IEM on our characterization of the phenomenon; indeed, *logically* IEM: if in fact somebody invented the zip, it cannot be other than Julius. Here *being the inventor of the zip* plays a metasemantic role

in fixing which *Julius* is at stake. Similarly, in intuitively IEM demonstrative examples (“you are standing very close”, “he is a long way off”, “this keyboard is black”), material derived from perception that “entails” the ascribed features (*standing close, being a long way off, being black*) helps fixing the referent.⁸

2. A model: observer memories and the IEM status of episodic memory

As said above, Shoemaker (1970) argues that the memory-based judgment “I saw John yesterday” is only *de facto* IEM. Evans (1982) challenged Shoemaker’s views. The debate was conducted based on weird science-fiction cases involving brain-splitting and brain-transplants.⁹ Pryor (1999, pp. 290–297) defends Shoemaker against Evans’s arguments – for the mere *de facto* IEM of perception (*I see John*) also by appealing to mundane – hence more compelling – situations involving mirrors and changes in point of view (*ibid.*, 297). In this section, I’ll outline the argument I have recently given for a Shoemakerian take on personal memory, which will offer us a model for the parallel argument on self-ascriptions of conscious thoughts in Section 4.

I’ll first summarize the gist of Evans’s discussion, because the main critical consideration in debates on thoughts we’ll confront are analogous. Remember the intuitive characterization we offered in Section 1 of EM and its possibility or otherwise, VEM and IEM. To show that a source of evidence (perception, memory) at most warrants *de facto* IEM, we must describe a situation (which may only obtain in remote possible worlds; the error need not occur in nearby worlds given that we grant the judgment its status as *de facto* IEM) that defeats justification for the belief *a is F*. Some related justification exists, but it comes from a source other than *a*: the judgment *a is F* is defeated, perhaps false, while its existential “part” (entailment) *that someone is F* remains justified. We just saw Campbell arguing along these lines to dispute the IEM of “that chair is yellow” in the discussion summarized in the last paragraph of the previous section.

Now, Pryor insists that the defeater to be provided for *a is F* must be *undercutting* but not *additive*. An undercutting defeater for *p* undermines the justification for believing it; and *additive* defeater gives, in addition, positive evidence for *not-p* (Pryor 1999, p. 284; cf. also Coliva & Palmira (this volume-a, Section 2). If additive defeaters are allowed, there would not be *any* cases of IEM, not even *I am in pain* when based on introspection. Smith (2006, p. 279) argues for this skeptical result, relying on additive defeaters: “suppose you experience what you take to be a pain. Now suppose someone reliably tells you that you are not really experiencing a pain but an itch, and that this has been caused by someone else’s suffering a real pain. Whenever they experience a pain, they press a button which causes you to feel a pain-like itch.” This illustrates why additive defeaters

should be disallowed. To show that a judgment is VEM it must be established that the justification for *a is F* that was *already in place before exposure to the defeater* still justifies the existential entailment, *someone is F*, after the defeat of *a is F*.¹⁰

Intuitively, it is not possible to undercut the introspective justification for *I am in pain* without also defeating *someone is in pain*. Smith's story only offers *new* justification for the existential claim, sneaking it in with the defeater – it tells us that when we think we feel pain, someone else does instead. This doesn't show that the initial justification for *I am in pain* given by the subject's nociceptors provides a justification for *someone is in pain* that can survive its defeat. Imagine that we are just told that data from a brain scan suggests that what we feel is an itch. If this defeats the singular claim, it defeats with it its existential "part". We are left with no good reason to think that the pain we thought we were feeling is in fact someone else's.¹¹

The gist of Evans's complaints about Shoemaker's considerations for the merely *de facto* IEM character of memory is that his cases are similarly inappropriate. Evans (1982, pp. 144–145) argues that, if we have apt (non-additive) reasons to disbelieve our mnemonic experiences, it would be absurd to still uphold that we were experiencing the apparent memories of *someone else*. The proper conclusion should rather be that we are more prone to memory illusions than we thought. Only by smuggling into the defeater some additional information – some story to the effect that we are experiencing the true memories of someone else, like the one Deckard tells Rachael in *Blade Runner*, fn. 9 – could we stick to the existential "part" of the original claim. In my view, Pryor (1999, pp. 294–296) addresses Evans's points convincingly enough. But more ordinary cases – like one involving mirrors (ibid., 297) that Pryor devises to show the merely *de facto* character of the IEM of perception – would be helpful. I'll summarize my argument (García-Carpintero, *in press*) that recent findings about the reconstructive character of memory that observer memories illustrate offer us precisely that regarding the IEM of memory.¹²

In the paper that triggered current research on the topic, the contrast is described as follows:

In some memories one seems to have the position of an onlooker or observer, looking at the situation from an external vantage point and seeing oneself "from the outside."

In other memories the scene appears from one's own position; one seems to have roughly the field of view that was available in the original situation and one does not "see oneself" (Nigro & Neisser, 1983, pp. 467–468)

In this characterization Nigro and Neisser assume that the perspective in observer memories is also that of the observed subject ("one seems to have the position of an onlooker"). Observer memories would thus have the paradoxical character of "out of body" experiences, in which one

represents oneself as occupying two locations at once. Lin and Dranseika (2021) report experimental data in which a large proportion of subjects (albeit not all) describe their observer *imaginings* in that way. The characterization that my own phenomenology supports is more guarded: observer memories present us as a participant in an event visualized from a given perspective, leaving it open whether it is occupied (and, if so, by the subject or someone else); it may just be that of a device like a camera.¹³ In field memories in contrast we locate ourselves at the origin of the visualizing perspective – the one we should have had at the originating perceptual experience. Rice (2010) offers a helpful review of the empirical literature.

Now, some philosophers claim that observer “memories” are not really memories; Vendler (1979) is an earlier example, and Fernández (2021) – as García-Carpintero (in press) argues – a sophisticated variation. A main intuitive motivation for the view comes from “preservative” intuitions about memory. Correct episodic memories shouldn’t just represent the same events as the perceptual experiences from which they come through some causal relation. All features of those perceptual experiences they reproduce should be accurate, even if not all are represented in full detail; this is because perceptual experiences leave “traces” reproducing aspects of their contents, retrieved later in recollection episodes. But this view has been conclusively empirically refuted by now. Episodic memory is to a large extent *reconstructive*; this still leaves open the possibility that a causal condition distinguishes them from imaginings, against the corollary that some philosophers derive from the downfall of the preservative view (cf. Robins, 2020). The view that observer memories are not memories is unduly revisionist (Sant’Anna, 2018).

How does granting observer memories the status of genuine episodic memories help to show that they are not logically IEM? I’ll outline the train of thought that the already mentioned memory paper develops. Consider a story involving disputed memories that (unlike its author) I think suggests a plausible train of thought to conclude that memory is not logically IEM:

I am reminiscing with my colleagues about a philosophical meeting in which I made a brilliant objection to the speaker’s thesis. However, none of us can remember any occasion on which I made that point. Then someone recalls Andrew Brennan making the very same point in a discussion which bears striking similarities to the one I have described. I am persuaded that this indeed was the discussion in question, and that I have misidentified the person who made the objection. “I made a brilliant objection” thus seems to be a mis-remembering – a personal memory-judgment with an incorrect detail, mistaken due to an error in identification. It seems natural to say that my reminiscence was based on information that was garbled, resulting in

a misidentification of the maker of the objection as myself. Thus memory-judgments are not guaranteed IEM (Hamilton, 2007, pp. 413–414)

A possible elaboration of how exactly the information was “garbled” goes as follows. A perceptual experience correctly representing the colleague raising the brilliant objection somehow leads to an erroneous memory of oneself raising it, perhaps an observer one.¹⁴ This speculation is not meant as a correct account of actual cases, which I am in no position to offer. For my purposes I only need the claim that it states a coherent *possibility*. Just before critically addressing Evans’s arguments against the view that Shoemaker’s cases establish that memory is only *de facto* IEM, Pryor (1999, p. 290) confesses that he “doesn’t have much positive argument to present in support of the claim” that the subject’s defeated apparent memories still offer her knowledge of the existential claim; this “just seems intuitively plausible to me”, Pryor says (*ibid.*). We can do better. When the subject’s role in the recollected event is disputed, she still may be confident about aspects of the recalled information – say, true details about the room, the location of the person raising the objection that the subject wrongly identified as herself and other colleagues, the identity of the speaker to whom the question was posed, the contents of her talk, her answer to the question, and so on. Facts such as the subject’s phenomenally grounded confidence in such details, their reliability, safety, or their counterfactual dependence on what transpired at the meeting may suffice (given the right epistemology of memory, whether dogmatist, reliabilist, or whatever) to preserve the subject’s justification for the existential claim: ok, I didn’t ask the question, but someone surely did.¹⁵

The *metasemantic* account of IEM underwrites counterexamples to the logical IEM of memories like this. It may well be that self-identification with a character in episodic memories – be they *field* or *observer* – is, in normal conditions, part of the reference-fixing information for the referent of “I”.¹⁶ This would explain on the metasemantic account why these attitudes are *de facto* IEM.¹⁷ The only logically non-negotiable aspect of that reference-fixing is the description provided by the token-reflexive self-reference rule. Hence, when considering nonstandard conditions like those in the example, the subject might consistently retreat to the option that he can just be identified as the experiencer of the relevant mnemonic episode; it is left as an open question whether he really had the properties ascribed to the individual represented in them as himself, while the justification of their existential generalization is preserved.

3. Varieties of self-conscious thoughts

I move on now to examine debates on whether cases of thought insertion help to prove that self-ascriptions of conscious thoughts are not logically IEM, and then in Section 4 I make my own case for it. I alluded earlier to one of the examples commonly discussed in the literature, of a subject who declares experiencing as somehow not his own the judgment *I should murder Lissi*.¹⁸ This certainly sounds *prima facie* bizarre. Most participants in the debate are moved however by Campbell's desideratum (2002, p. 39) that delusions be explained "as broadly rational responses to highly unusual experiences". Of course, this runs a contrasting risk that Coliva (2002a, p. 45) and Gunn (2016, pp. 570–571) voice – overlooking "the bizarre nature of the phenomenon", as the latter puts it. I'll indicate below how my proposal meets the concern.¹⁹

As Campbell puts it, researchers have tried to achieve this goal by finding "some structure in our ordinary notion of the ownership of a thought which we might not otherwise have suspected", Campbell, Sugden (1999, p. 610). The suggested "structure" distinguishes the *authorship* of the relevant thoughts from their *ownership* (cf. also Stephens & Graham, 1994). The latter can be elucidated in terms of a notion with a long pedigree in philosophy that has only recently become a critical target for current research; García-Carpintero and Guillot (2023) review it in an introduction to a recent compilation on the topic. It goes by many different labels including *inner awareness*, *pre-reflective self-consciousness* or *for-me-ness*; here I'll follow them in calling it *subjectivity*. Minimally this is a feature that distinguishes our conscious thoughts from both the thoughts we ascribe to others, and our unconscious thoughts. On a most deflationary way of understanding it, it is just the disposition (whatever its basis) to self-ascribe the thoughts in introspection – here a self-conscious metarepresentational judgment. Subjectivity thus marks the ontological point that conscious thoughts do not float free from a subject. More substantive accounts specify the basis for that disposition as a distinct phenomenal feature common to all introspectable thoughts, with its own unique phenomenal character unlike that of *qualia* differentiating particular conscious thoughts. The access that the subject has to it is analyzed as either a nonrepresentational relation of *acquaintance* with the conscious thought, and perhaps through it with its subject, or in representational terms instead. I won't go further into this here; see García-Carpintero and Guillot (2023) for further details and references.

Some accounts explain thought insertion in that the relevant thoughts lack subjectivity. Like Henriksen et al. (2019) and Mathieson (2023), p. I find it difficult to make sense of this. Subjects "cannot lack subjectivity in the minimal sense I endorse, since the very basis for patients' complaints that thoughts are "in them" without being "theirs" relies on some (at least

minimal) sense of subjectivity or ownership being retained”, Mathieson (2023, fn. 6; cp. fn. 9). These authors, however, advance the related account that subjectivity is somehow “disturbed in schizophrenia spectrum disorders” (Henriksen et al., 2019, p. 7; Mathieson, 2023, Section 5). Such “disturbance” is described as an “increasingly felt distance between the experiencer and their experiences” (Mathieson, 2023, Section 5) that patients report. I find this view unstable. Both on the minimal dispositional account of subjectivity, and in the more robust phenomenal-trait view that they hold, subjectivity doesn’t seem gradable. I cannot hence see how the “disturbance” of an “increasing self-alienation” can be a “deficit”, presumably taking away a measure of it.²⁰ As far as I can tell, the reported “felt distance” and “self-alienation” just amount to the tension between the patients’ simultaneous self-ascription and denial thereof that we are trying to understand.²¹

As Campbell (2002) complains, Coliva’s objection to his argument that thought insertion undermines the IEM of introspective self-ascriptions depends on her refusal to countenance any other relevant sense of self-ascription than “ownership” in the subjectivity sense in which this amounts to “the possibility of self-ascription of it by me” (Campbell, 2002, p. 35). She thus must find patients’ disclaimers incoherent (Coliva, 2002a, p. 45). Campbell’s view is that it is the *authorship* sense that the thought is “generated by me” (Campbell, 2002, p. 36) that provides the explanation: “the content of the schizophrenic’s illusion is that he has first-person knowledge of token thoughts which were formed by someone else. And there is no immediate contradiction in that” (Campbell & Sugden, 1999, p. 620).

There are however two different strands in this “authorship” sense that Campbell does not clearly separate (Seeger, 2015b, p. 853, fn. 15), a notion of *agency* – being a significant link in the causal chain producing the thought, which thus “causally originated in the appropriate way within the subject” (Seeger, 2015b, p. 844) – and a different (as I’ll call it) *committed authorship* or just *commitment* sense in which the thoughts express and affect the subject’s “long-standing beliefs and desires” Campbell and Sugden (1999, p. 621). Seeger (2015b) holds that it is the first that helps to interpret claims of thought insertion; Bortolotti and Broome (2009) defend the commitment view that it consists in “failing to ascribe to oneself a thought that is accessed first-personally and in failing to either endorse the content of that thought with reasons or manifest commitment towards it in behaviour” (ibid., 206).²² They argue that agency views fail to distinguish thought insertion from “unsolicited” mental states such as tunes that come to mind unbidden (ibid., 219).

Seeger’s account, however, assumes a broad notion of agency that encompasses but goes beyond intentional causation, including anything originated “within the subject” like such unbidden imaginings – cf. Sousa and Swiney

(2013, p. 641), Langland-Hassan (2016, pp. 678–679). Seeger argues that it is in fact commitment accounts that fail to distinguish inserted from unsolicited thoughts, when the latter don't conform to the subject's long-standing dispositions, and thus she wouldn't endorse them (Parrott, 2017, pp. 42–45). Many common examples illustrate this point. Schizophrenic patients frequently disclaim imperatival thoughts like “Kill God!” that they nonetheless own, experienced as addressed to them; indeed, the example we have been considering so far might be a case in point, because the “electrically given thought that I should murder Lissi” might have been experienced “uttered” in auditory imagination as “Kill Lissi!”. Consider this passage in an article on Kip Kinkel, an early school shooter:

Mental-health professionals refer to voices that order an individual to act in a certain way as “command hallucinations.” Kip told a psychologist that, when he and his father had returned home that day, he heard voices saying, “‘Get your gun. Shoot him. Shoot him.’ So I did. I had no choice.” After he shot his father, he said, the voices continued: “They told me to kill Mom because I'd already killed Dad. ‘No choice, do it!’ they said.” The next morning, Kip said, the voices told him, “Go to school and kill everybody.” (Gonnerman, 2023)

Clearly, although we appraise orders that we experience as addressed to us, we in no way take responsibility for them, or seek to find them coherent with our rational perspective (Sousa & Swiney, 2013, p. 643); promoters of the commitment account should at least admit a limited application for it.²³ It may in fact be wise to adopt a pluralist view, allowing that in some cases it is agency in Seeger's broad sense that is disclaimed, in others commitment.²⁴ To properly describe all cases we thus need the three senses of ownership – whether subjects are, or take themselves to be, either *experiencers* of the thoughts' subjectivity, their *agents/producers*, or their committed *authors*. We should also take into consideration the *force* or *mode* of the mental states, because which sort of ownership is at stake may crucially depend on this: “[...] it is plausible to think that thoughts are not identified uniquely on the basis of their propositional content, but also on the basis of the attitude that the subject has towards that content” (Bortolotti & Broome, 2009, p. 221). We need, as Sousa and Swiney (2013) superbly put it,

to draw a clear distinction between different types of thought [...] thoughts may differ in terms of content (whether they are about vacation or work), coding format (whether they are in a propositional language of thought or in imagery), or “attitude” (if in a propositional language of thought, whether they are beliefs or desires; if in imagery like inner speech, whether they are assertions or commands) [...] one may say to oneself in inner speech “finish this article!” This thought has a certain content (about finishing an article), coding format (verbal imagery), and attitude (it is a command) (Sousa & Swiney, 2013, p. 645)

The examples we have considered so far presume that the relevant thoughts come in *inner speech*, “the phenomenon of speaking to oneself silently” (Langland-Hassan, 2016, p. 676). There is an ongoing controversy about the nature of inner speech. In a discussion of auditory verbal hallucinations including thought-insertion, Gregory (2016) notes that subjects may experience hallucinations not “in” their own voice (cf. also Wu, 2012, pp. 95–96); our imperatival cases might be actual instances. Gregory argues that these should not be conceptualized as *inner speech* proper, but rather as auditory imaginings. He is assuming here a view that he defends in the paper, which more recently he presents as the “consensus [...] among philosophers working on the topic [...] that inner speech is a kind of actual speech” (Gregory, 2022, Section 1; cf. e.g., Kompa 2023).²⁵ Although Gregory himself now has doubts, I’ll assume this “consensus”.

Sousa and Swiney (2013) make the quoted point in support of their compelling criticism of Fernández’s (2010) claim that inserted thoughts are in all cases beliefs – or, rather, judgments (assertions in inner speech); beliefs are dispositional states, while inserted thoughts are occurrent eventualities.²⁶ In brief, there is no good reason to rack imperatival examples, say, on that Procrustean bed. Currie’s (2000) view that they are imaginings is a better fit. His account of thought-insertion, however, requires that all cases be imaginings, and that is another Procrustean bed we should avoid.²⁷ In a clear-cut sense that semanticists have elucidated (Charlow, 2014), acceptance of an order understood as addressed to us (*Kill Lissi!*) entails a commitment to a normative belief, *I should kill Lissi*. Subjects might not be really imagining this but might be assertorically committing to it nonetheless; this is meant as a claim about the nature of the relevant thoughts, not just about how subjects experience them (cf. Seeger, 2015b, pp. 839–840).²⁸

Let me sum up the suggestion about the proper way of making sufficiently intelligible claims of thought insertion that I derive from the previous discussion, before moving on to our issue of the IEM of thought. We must be clear about the kind of thought at stake – as Sousa and Swiney put it, its format, content, and force; what it really is, and how subjects experience it. Perhaps it is an imagining, as in the previous imperatival cases, which subjects experience as conveyed to them in a peculiar way (“electrically” given). In that case, they are really their producers and not just their experiencers, but they disclaim the former; some of the subpersonal accounts that have been envisaged (cf. Langlan-Hassan, 2016) might explain this experience. It may also be the expression of a normative belief that they really hold in accepting the command, and then they don’t deny authorship; it is their “ultimate” agency, or relevant origin that they disclaim.²⁹

We thus make sense of claims of thought insertion along the lines envisaged by Campbell (“there is no immediate contradiction” in their

reports, Campbell & Sugden, 1999, p. 620) and others. In response to Coliva's (2002a, p. 45) and Gunn's (2016, pp. 570–571) sensible complaint, this however doesn't make subjects' episodes more rational than it should. As Roessler (2013) suggests, we may grant that "patients' tendency to attribute the thinking in question to other individuals and to invoke abstruse mechanisms of "transmission," or even just their apparent inability to appreciate the implausibility, all things considered, of their denial of "agentive ownership," reflects a disorder of rationality" (ibid., 659); cf. also Campbell and Sugden (1999, pp. 620–622). Their denials are certainly *bizarre* by the DSM 5 TR (2022, 101) criterion that they are "clearly implausible and not understandable to same culture peers and do not derive from ordinary life experiences".

4. Self-ascriptions of conscious thought are not always IEM

The account in Section 3 affords what we need to argue that self-ascriptions of thoughts are not logically IEM. Our main ground is the tripartite distinction of senses of ownership, applied to thoughts properly individuated by their type, content and format. That move was a crucial part also in the argument outlined in Section 2 for memory. There we distinguished mnemonic experiences from memories proper. It was not the former whose self-ascription was argued not to be IEM, but the latter. It wouldn't do to limit the discussion to the IEM of mnemonic experiences, which as far as I know nobody denies; the claim confronting Evans and Shoemaker is not whether they are IEM, but whether the memories they underwrite, and judgments based on them, are. Mnemonic experiences are needed not to beg the question: the debate concerns whether some of them may be erroneous memories, the mistake in question constituting a case of EM.

It is similar here. For all I can tell, the self-ascription of experiencing an auditory image of "Kill Mom" is logically IEM. In ordinary circumstances, the subject might take herself to be imagining someone making the command it expresses, herself or someone else, with a voice "sounding" like her own or otherwise. She can also take it as a command she addresses to herself in inner speech. In both cases she would be, and would feel herself to be, the agent and the author – of the imagining in the first case, the command in the second. The schizophrenic subject interprets it as a sort of "perception" of a command, mediated by unknown telepathic processes, whose agent and author are someone else; someone authoritative enough for him to endorse the normative judgment that acceptance of the command entails, *I should kill Mom*.

I take this to suggest an epistemic possibility. A non-schizophrenic subject experiences a "Kill Mom" episode as an order she addresses to herself – to put it in the shape of an explicit performative, *I am hereby commanding*

myself to kill Mom, which we can also take as the introspection-based self-ascription at stake. Unknown telepathic processes like those that schizophrenic patients fabricate are available in the envisaged situation – some form of e-mail system directly communicating minds.³⁰ Finally, the circumstances make it reasonable to consider the command authoritative – Mom has been grossly misbehaving lately. Under those conditions, the subject might sensibly worry whether the command ultimately truly originates in her, i.e., whether she is not in fact receiving it from someone else: *ok, someone is thereby commanding me to kill Mom, but is it me?* Campbell might have been contemplating something like this, but he doesn't say. In any case, it is a relevant epistemic possibility that shows that self-ascriptions of thoughts such as intentions or self-commands are not logically IEM.

In the literature I know, Roessler (2013, pp. 662–664) comes closest to considering this argument:

If you were to enjoy an experience with the distinctive phenomenology of inner speech without any awareness of performing an act of inner speech, you would be introspectively aware of an episode of thinking without being aware that it is you who is doing the thinking. You might then consistently, and intelligibly, affirm, say, that the thought “Kill God” is being entertained in your mind while denying that it is you who is thinking “Kill God.” (ibid., 662)

He then rejects that the condition envisaged in the first sentence – analogous to the one my argument features – is coherent. If I understand him correctly, he argues on Anscombian grounds that imagining oneself saying something is imagining an intentional activity, necessarily performed by oneself *as such, de se*; while imagining someone making a speech act is something else entirely: it is imagining that one comes to know by listening. It doesn't make sense to contemplate that one has mistaken one with the other. But this just question-beggingly denies that the envisaged situation is possible. He acknowledges that “[s]ometimes it is far from clear whether one imagines saying something or imagines listening to someone else saying something” (ibid., 664). This is all I need; but he just goes on to reiterate his Anscombe-based dismissal. But sometimes we are not sure whether we are perceiving bells tolling or imagining it instead; my own phenomenology establishes this, and it is the standard interpretation of Perky's (1910) famous experiment, cf. Currie (2000, p. 180). These are not the cases we are discussing, but it questions Roessler's Anscombian point. For if it is an imagining, it is something we do, *de se*: while if it is a perceptual experience, it is not.³¹

Parrott (2017, pp. 54–58) also considers an argument like mine, which he relates to the memory case roughly along the lines I have. He contends that, while in the memory case we can accept that the envisaged situation that the mnemonic experience misidentifies the

subject is a *nearby* possibility (to be precise, a remote possibility such that the relevant condition is *there* a sufficiently nearby, accessible one), the corresponding situation for thoughts' self-ascriptions is not: "There is no nearby world in which other people insert thoughts into peoples' minds and neither are experiences of thought insertion explicitly defined in relation to some other person's thinking", (ibid., 57). But the Kip Kinkel case referenced in the previous section is explicitly defined in relation to some other person's thinking – the voice to whom the adolescent grants authority is clearly that of a person, no matter how otherwise indefinite; the same might obtain in other cases in the literature, like the "Kill Lissi" one. For all Parrott says, thus, that situation obtains in a world accessible to a possible one – although one remote from the actual world.

Hu (2017) provides another argument to establish that self-ascriptions of conscious thoughts are not logically IEM, which Palmira (2022) questions with considerations I find compelling. Our arguments are very different; so much so that, as far as I can tell, Palmira's points don't apply to mine. Hu aims to show that schizophrenic patients' claims of thought insertion *by themselves* establish the non-IEM claim. What I have argued is that cases of thought insertion help to devise epistemic possibilities that show the same about the self-ascription of conscious thoughts by *non-schizophrenic* subjects.

Let me sum up. In the memory case, I argue that new developments in the understanding of memory – pointing to their reconstructive character, which observer memories illustrate – help to devise epistemic possibilities that establish the at most *de facto* character of the IEM of memory-based judgment. The fact that in observer memories we represent us "externally" as a participant in the remembered scene allows us to contemplate fully distancing ourselves from that individual, imagining that he was in fact someone else whom we misidentified with us. Similarly, the fact that schizophrenic patients represent "externally" the agent of thoughts such as orders that they in fact give themselves, an experience as such, by disclaiming to be them, allows us to envisage situations in which we may conclude that the true agent of orders we experienced as giving to us was in fact someone else. We thus establish that the IEM of memory judgments and self-ascriptions of conscious thoughts as agent or author is only *de facto*.

My argument addresses self-conscious thoughts – intentions, self-commands, judgments – whose self-ascribing subjects represent themselves not just as their experiencers, but also as their agents and committed authors. I made a related point about the memory case, to wit, that we must target the committal case of memories and memory-based judgments, which is what Evans and Shoemaker were discussing. Non-committal mnemonic experiences allow us to describe the relevant possibilities without

begging any question, but their IEM status is not under debate. Palmira (2020) develops Coliva’s “introspectionist” view, by articulating the self-reference rule exclusively for self-ascriptions involving the subject’s *experiencer* role. Palmira can thus uphold Coliva’s view that “if one is first-personally aware of an occurrent conscious thought, then it must be one’s own, and this is a conceptual point” (Sollberger, 2014, 601).

I of course grant this, but, as in the memory case, I point out that it is not just thoughts *as experienced*, but also *as produced* and *endorsed* that *prima facie* appear to be logically IEM and have been claimed to be so in the philosophical literature.³² Again, taking into consideration self-ascriptions of conscious thoughts as their experiencer – whose status as logically IEM, as said, nobody should dispute – is required to investigate the debated cases of the IEM status of their self-ascription as agent and author without begging any question. But what Campbell wanted to show is that the phenomenon of thought insertion helps to establish that the latter are not IEM. I think this is what is ultimately at stake, and I have presented a line of thought that elaborates on the points that Campbell left undeveloped.

Notes

1. Wiseman (2019) raises interpretive issues about Shoemaker’s “elucidation”; Palmira (2022) compellingly addresses the philosophical implications she derives from them.
2. Putnam held that *being H₂O* defines the real essence of the kind designated by “water” in its “predominant sense” (Putnam, 1975, p. 239); he explicitly allows that “water” has another sense defined by superficial traits, cf. Tobia et al. (2020) for X- ϕ validation. Like me, Campbell (2002, p. 35) distinguishes IEM as a pre-theoretical *datum*, from the philosophical theories that aim to account for it; cf. also Recanati (this volume, Section 1–2). Campbell’s indications on what he takes to be the datum (Campbell, 1999, p. 89) agree with the proposal I’ll go on to make in the main text.
3. This slightly modifies Seeger’s (2015b, p. 2) account, aiming to cover cases of Pryor’s (1999) “wh-immunity”, see below. McGlynn (2021, p. 2305) offers a definition that I take to develop this core idea; for my purposes a simpler one is good enough. Recanati (this volume, Section 1) also offers a good, intuitive similar characterization. Cf. Morgan (2019, pp. 446–447) and Palmira (2020, Section 2) for related notions and McGlynn (2021, pp. 2309–2310) for a defense of the specifics of his own.
4. Wright (1998, p. 19), Campbell (1999, p. 89) and Prosser (2012, pp. 161–162) support this core idea; cf. Coliva & Palmira (this volume-a, Section 2) for more details on different views on the phenomenon. Pryor’s “package deal” metaphor parallels Evans’s (1982, p. 182) “no gap” one.
5. Cf. Coliva and Palmira (this volume-b, Section 1) and Recanati (2024, Section 2) for elaboration.
6. To elaborate: “this keyboard is black” appears to be IEM because, if the existential “part” of the judgment is true and justified (i.e., something in the immediate vicinity of the speaker, the *source*, is black), it doesn’t seem to be possible that it is not the referent of “this keyboard”. The intuitive reason – which the Metasemantic account

I outline below articulates – is that the referent of the demonstrative “this keyboard” is in part perceptually fixed in the relevant context as the *black* object before the speaker. Formally and philosophically well-developed versions of the no-reference view have been advanced, cf. Ninan (2013), Bochner (2023), p. I don’t think these models vindicate the Simple Account, because they still presuppose forms of “identification” consistent with the Metasemantic view; but I cannot develop this point here.

7. Cf. Coliva and Palmira ([this volume-a, Section 1](#)) for elaboration on the debate.
8. Campbell (1997, p. 70) argues against a view like the one just outlined with the following example: “if you judge, ‘that chair is yellow’, it may be that you thereby know of something that it is yellow, but that thing is not the chair, if, for instance, the chair is transparent and set against a yellow background. If your judgment is mistaken, you can rectify it by retreating to the more cautious, ‘At any rate, something is yellow’.” García-Carpintero (2018) retorts that this just shows that the IEM in these cases is merely *de facto*.
9. Perhaps they are not that weird. Folk viewers of *Blade Runner* (Scott, 1982) don’t manifest any difficulty in accepting that this admission by Rachael presents a coherent possibility: “I remember [music] lessons, but I don’t know if it is me or Tyrell’s niece” – Deckard has just told her that her private memories were “implants” from those of Tyrell’s niece.
10. “IEM arises only if the *original* justification cannot be cited as support for ‘Someone is F’ when the assertion is doubted for any reason at all” (Hamilton 2007, p. 411), my emphasis. This is the *desideratum* that Palmira (2020, p. 3840) calls “Preservation of Grounds”; cf. also Coliva and Palmira ([this volume-a, Section 2](#)).
11. This is Shoemaker’s (1968, 563f.) “package deal” intuitively correct point, quoted in [Section 1](#) “[...] in being aware that one feels pain one is, tautologically, aware, not simply that the attribute *feel(s) pain* is instantiated, but that it is instantiated *in oneself*”.
12. I’ll avoid henceforth the circumlocutions required by the factivity of “perception” and “memory” by using those terms for perceptual and mnemonic *experiences* or seemings. I hope no misunderstanding ensues. I’ll come back to the importance of the distinction in [Section 4](#).
13. Cf. Perrin and McCarroll (2023, p. 306) for a characterization congenial with this view.
14. I mostly agree with Teroni’s (2024) account of how memories (mnemonic experiences really) are identified by the subject, which assumes as I do that this is a “personal-level” (as opposed to *subpersonal*) matter. It allows for mistakes both of omission (failing to realize that an attitude is a mnemonic experience) and commission (mistaking as one something that is not, perhaps an imagining), but it assumes that typically we are reliable. I also agree with him that, given sensible constraints he sets up, accounts like Fernández’s (2024) in terms of the contents of memories are inadequate; adequate accounts should rely instead on traits of the attitude type. Teroni offers a compelling one on which it features a “feeling of familiarity”.
15. Hamilton argues that story cannot be taken as I do, by invoking Evans’s consideration outlined above to dismiss my Shoemakerian conclusion: without smuggling information into the defeater, making it *additive* instead of *undercutting*, “the idea that, if it was not myself, then it must have been someone else who made the objection, would just be a stab in the dark” (Hamilton 2007, p. 414). García-Carpintero ([in press](#)) argues that his considerations assume preservationist intuitions and disregard thereby observer memories.

16. The same may apply to bodily self-ascriptions, including pain if we are prepared to question the logical status of Shoemaker's "package deal" view about them, fn. 11.
17. Even if the story as I imagined it was not just a coherent counterfactual possibility, but an actual process, this wouldn't establish that memory is not *de facto* IEM. This depends on the proper way of characterizing that notion, and the epistemically relevant facts. There might be, as envisaged, good reasons for the view that, in clear-cut situations constituting an epistemic default, memories (and bodily proprioception-based self-ascriptions) must be reliable. Recanati (2024, Section 3–4) develops an account I sympathize with.
18. The full report was "One evening the thought was given to me electrically that I should murder Lissi" (Mullins & Spence 2003, p. 295). Partly in response to sensible complaints by Gunn (2016) and Henriksen et al. (2019) that philosophers repeatedly recycle the same examples, I'll offer new ones, see the Kip Kinkel case. I am afraid that I don't comply with their advice that to address these issues we obtain "a solid, clinical grasp of the phenomenon of thought insertion" (Henriksen et al. 2019, p. 3). I agree that ideally one should; I'll try to limit my claims to what can be justified by standard philosophical methodology and theories of contentful attitudes.
19. Cf. Roessler's (2013, pp. 666–668) nuanced discussion. Verdejo (2023) argues for the bold view that "in the light of their experiences, people undergoing thought insertion express, or potentially express, fully rational judgments concerning the ownership of their thoughts" (ibid., Section 1). Even if relativized to subjects' experiences, the "fully rational" appraisal sounds a bit overblown given how confusing patients' reports are, and the little we can in fact know about their experiences.
20. Cf. Rothenfluch (2021, p. 166 fn). Defenders of this sort of account might instead appeal to "additions" to the experience of subjectivity, as opposed to deficits; cf. Parrott (2017, Section 2). Mathieson (2023, Section 5) doubts that this might be the whole story.
21. Gunn (2016, p. 572) argues that "[t]he subjective 'feel' of the experience of one's own mental activity is how one knows that one is the agent, author and owner of a thought. If one does not recognize that a thought is one's own (for whatever reason) then personal ownership is lacking". By "personal ownership" she is clear that she means subjectivity (ibid.). The argument begs the question. As we have seen, subjects typically acknowledge a sense in which the thoughts they disown are still theirs ("the thought was *given to me* electrically . . .", fn. 18, my emphasis). It is a theoretical question to articulate in what sense subjects deny ownership; we cannot just declare that patients' denials already establish that subjectivity is missing or at least diminished (Sousa & Swiney, 2013, p. 643). Seeger (2013, p. 262) points out the same questionable move in Martin and Pacherie (2013, p. 113).
22. Rothenfluch (2021) offers a recent variation on this sort of account, on which "the experience that an introspectively accessed thought is not one's own consists in the feeling of rational indifference toward the thought. Patients have this experience because they lack an expectation of rational authority" (ibid., 168) that they should have, given that "inserted thoughts are of the type that we would ordinarily expect to be rationally justified" (ibid., 169).
23. Rothenfluch (2021, p. 169, fn.) should grant this, for her account of spontaneous thoughts as nonetheless expected to be under our rational control (in that we might *discard* them) doesn't apply to orders felt as inserted. The problem they pose for her account is that patients' "rational indifference" toward them is in such cases entirely apt. Also, patients accept the commands as authoritative and are thereby committed to the ensuing normative judgments, *I should kill Lissi/I should kill Dad*. They thus

- fully meet any “expectation of rational authority” vis-à-vis them; but they still disclaim them as not their own.
24. Proponents of commitment accounts are indeed right that there are many cases in which what patients report is lack of commitment, cf. Gunn (2016, p. 563) for illustration.
 25. Roessler (2016, Section 2) has an illuminating discussion of the debate confronting this view and its competitor, that inner speech is just the imaginative representation of speech.
 26. Roessler (2013, p. 661) claims that “in the familiar sense, the term [*authorship*] applies to types, not tokens. It is books that have authors, not individual copies of books”. But artists also author paintings, sculptures and performances, in addition to books and symphonies.
 27. Currie’s account has it that subjects miscategorize their imaginings as beliefs because imaginings are actions and the subpersonal mechanism underlying action awareness (which, like Campbell, Sugden (1999, pp. 611–614), he identifies with Frith’s *forward-efference/comparator model*) fails in them. If they don’t endorse those beliefs, they explain their presence as having been produced by others. This is on his view how the delusions can be seen as “broadly rational responses to highly unusual experiences” (Campbell, 2002, p. 39). Langland-Hassan (2016) discusses objections to Frith’s model and offers alternative accounts that might underwrite similar explanations.
 28. Patel (2023) defends the *prima facie* hugely implausible “antirealist” view that “sufferers of thought insertion do not have a thought, but, as with dreams, merely simulate having a thought inserted into their minds” (ibid., Section 1). But dreams are categorized as either imaginings or hallucinations (Gregory 2023), and hence thoughts in a generic sense; and accounts of pretense and simulation explain them as resulting from imaginings (Picciuto & Carruthers, 2016). Patel in fact argues for the more nuanced claim “that thoughts corresponding to first-person descriptions of thought insertion are not parts of thought insertion episodes. . . . the thought corresponding to the description is the thought *Kill god*, and so the anti-realist is only denying that this thought is a part of the episode”, (ibid., Section 3). To the extent that I understand it, this more nuanced claim still seems wrong to me. Their reports offer good reasons, I take it, to think that patients carry out acts of propositional imagining that someone tells them to kill God (Mom, Lissi . . .); that they sometimes experience themselves as receiving such orders from a source invested with the required authority, judge that they should act on them, and, tragically, do so.
 29. On Seeger’s (2015b, p. 847) account of agency, subjects would be the producers also in this case, but I take them to be assuming a concept of agency on which they are not the (ultimate, truly significant) agents – cf. Vosgerau and Voss (2014, pp. 539–541), who offer a nuanced account of what I am calling *agency* in terms of causal and control features; cp. Seeger’s (2015b, pp. 849–850).
 30. This condition is meant to block the objection that the possibility I am envisaging inaptly characterizes an additive defeater (see the discussion above in §2), along the lines of Pryor’s (1998, 296) response to Evans. The defeater I consider doesn’t bring with it the notion that someone else authored the command; it is a general background condition that supports it. It is thus as with standard cases establishing that a self-ascription is VEM, for instance the Perky (1910) inspired cases I describe immediately below.
 31. I assume that the envisaged situations are remote possibilities, leaving open the view that self-ascriptions of intentions or self-commands are *de facto* IEM. A reviewer

worries that if they are possible an excessive skepticism about our own self-knowledge follows, threatening our rationality. I addressed this worry in fn. 17. Consider the Perky-inspired more ordinary cases mentioned in response to Roessler's Anscombian argument. My own experience shows them to be real. However, I think a sensible, fallibilist epistemology would nonetheless allow that we can typically tell by introspection whether we are perceptually experiencing or imagining.

32. Verdejo (2023) invokes his “full rationality” claim (fn. 19) to argue for the philosophical significance of the *agent-and-author* based self-reference rule, not just the *experiencer* rule.

Disclosure statement

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

Funding

Financial support for my work was provided by the MICIU/AEI/ 10.13039/501100011033, research projects [PID2020-119588GB-I00], [CEX2021-001169-M], and through the award “ICREA Academia” for excellence in research, 2018, funded by the Generalitat de Catalunya. Thanks to Jordi Fernández, Michele Palmira, Víctor Verdejo and two anonymous reviewers for very useful comments and suggestions, and to Michael Maudsley for the grammatical revision.

References

- Anscombe, G. E. M. (1975). The first person. In D. G. Samuel (Ed.), *Mind and language* (pp. 45–65). Clarendon Press.
- Bochner, G. (2023). Fixing internalism about perceptual content. *Philosophical Explorations*, 26(3), 404–419. <https://doi.org/10.1080/13869795.2023.2209577>
- Bortolotti, L., & Broome, M. (2009). 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. (1997). Sense, reference and selective attention. *Proceedings of the Aristotelian Society, sup. vol.* (Vol. 71. pp. 55–98).
- Campbell, J. (1999). 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. (2002). The ownership of thoughts. *Philosophy, Psychiatry and Philosophy*, 9(1), 35–39. <https://doi.org/10.1353/ppp.2003.0001>
- Campbell, J., & Sugden, S. J. B. (1999). Schizophrenia, the space of reasons and thinking as a motor process. *The Monist*, 8(4), 609–625. <https://doi.org/10.5840/monist199982426>
- Charlow, N. (2014). The meaning of imperatives. *Philosophy Compass*, 9(8), 540–555. <https://doi.org/10.1111/phc3.12151>
- Coliva, A. (2002a). On what there really is to our notion of ownership of a thought. *Philosophy, Psychiatry and Philosophy*, 9(1), 41–46. <https://doi.org/10.1353/ppp.2003.0005>

- Coliva, A. (2002b). Thought insertion and immunity to error through misidentification. *Philosophy, Psychiatry and Philosophy*, 9(1), 27–34. <https://doi.org/10.1353/ppp.2003.0004>
- Coliva, A., & Palmira, M. (this volume-a): Immunity to error through misidentification I: Nature and scope. *Philosophical Psychology*.
- Coliva, A., & Palmira, M. (this volume-b): Immunity to error through misidentification II: Explanation and significance. *Philosophical Psychology*.
- Currie, G. (2000). Imagination, delusion and hallucinations. *Mind and Language*, 15(1), 168–183. <https://doi.org/10.1111/1468-0017.00128>
- Evans, G. (1982). *The varieties of reference*. Clarendon Press.
- Fernandez, J. (2024). Memory and self-reference. *International Journal of Philosophical Studies*, 32(1), 59–77. <https://doi.org/10.1080/09672559.2024.2334461>
- Fernández, J. (2010). Thought insertion and self-knowledge. *Mind and Language*, 25(1), 66–88. <https://doi.org/10.1111/j.1468-0017.2009.01381.x>
- Fernández, J. (2021). Observer memory and immunity to error through misidentification. *Synthese*, 198(1), 641–660. <https://doi.org/10.1007/s11229-018-02050-3>
- García-Carpintero, M. (2015). De Se Thought. In S. Golberg (Ed.), *Oxford handbooks Online*. <https://doi.org/10.1093/oxfordhb/9780199935314.013.61>
- García-Carpintero, M. (2016). Token-reflexive Presuppositions and the De Se. In M. García-Carpintero & S. Torre (Eds.), *About oneself* (pp. 179–199). Oxford University Press.
- García-Carpintero, M. (2017). The philosophical significance of the De Se. *Inquiry: A Journal of Medical Care Organization, Provision and Financing*, 60(3), 253–276. <https://doi.org/10.1080/0020174X.2017.1262003>
- García-Carpintero, M. (2018). De Se Thoughts and immunity to error through misidentification. *Synthese*, 195(8), 3311–3333. <https://doi.org/10.1007/s11229-015-0817-y>
- García-Carpintero, M. (in press). Memory-based reference and immunity to error through misidentification.
- García-Carpintero, M., & Guillot, M. (2023). Introduction: Views about self-experience. In M. García-Carpintero & M. Guillot (Eds.), *Self-experience: Essays on inner awareness* (pp. 1–23). Oxford University Press.
- Gonnerman, J. (2023). *The aftermath*. New Yorker.
- Gregory, D. (2016). Inner speech, imagined speech, and auditory verbal hallucinations. *Review of Philosophy and Psychology*, 7(3), 653–673. <https://doi.org/10.1007/s13164-015-0274-z>
- Gregory, D. (2022). How not to decide whether inner speech is speech: Two common mistakes. *Phenomenology and the Cognitive Sciences Online First*, 23(2), 231–252. <https://doi.org/10.1007/s11097-022-09814-w>
- Gregory, D. (2023). Imagining a way out of dream skepticism. *Erkenntnis*. <https://doi.org/10.1007/s10670-023-00662-x>
- Gunn, R. (2016). On thought insertion. *Review of Philosophy and Psychology*, 7(3), 559–575. <https://doi.org/10.1007/s13164-015-0271-2>
- Hamilton, A. (2007). Memory and self-consciousness: Immunity to error through misidentification. *Synthese*, 171(3), 409–417. <https://doi.org/10.1007/s11229-008-9318-6>
- Henriksen, M. G., Parnas, J., & Zahavi, D. (2019). Thought insertion and disturbed for-meness (minimal selfhood) in schizophrenia. *Consciousness and Cognition*, 74, 102770. <https://doi.org/10.1016/j.concog.2019.102770>
- Hu, I. (2017). The epistemology of immunity to error through misidentification. *The Journal of Philosophy*, 114(3), 113–133. <https://doi.org/10.5840/jphil2017114310>

- Kompa, N. (2023). Inner speech and ‘pure’ thought – do we think in language? *Review of Philosophy and Psychology*. <https://doi.org/10.1007/s13164-023-00678-w>
- Langland-Hassan, P. (2016). Hearing a voice as one’s own: Two views of inner speech self-monitoring deficits in schizophrenia. *Review of Philosophy and Psychology*, 7(3), 675–699. <https://doi.org/10.1007/s13164-015-0250-7>
- Lin, Y.-T., & Dranseika, V. (2021). The variety and limits of self-experience and identification in imagination. *Synthese*, 199(3–4), 9897–9926. <https://doi.org/10.1007/s11229-021-03230-4>
- Martin, J.-R., & Pacherie, E. (2013). Out of nowhere: Thought insertion, ownership and context-integration. *Consciousness and Cognition*, 22(1), 111–122. <https://doi.org/10.1016/j.concog.2012.11.012>
- Mathieson, D. (2023). Schizophrenic thought insertion and self-experience. *Review of Philosophy and Psychology Online First*. <https://doi.org/10.1007/s13164-023-00680-2>
- McGlynn, A. (2021). Immunity to wh-misidentification. *Synthese*, 199(1–2), 2293–2313. <https://doi.org/10.1007/s11229-020-02885-9>
- Morgan, D. (2019). Thinking about the body as subject. *Canadian Journal of Philosophy*, 49(4), 435–457. <https://doi.org/10.1080/00455091.2018.1482432>
- Mullins, S., & Spence, S. (2003). Re-examining thought insertion. *British Journal of Psychiatry*, 182(4), 293–298. <https://doi.org/10.1192/bjp.182.4.293>
- Nigro, G., & Neisser, U. (1983). Point of view in personal memories. *Cognitive Psychology*, 15(4), 467–482. [https://doi.org/10.1016/0010-0285\(83\)90016-6](https://doi.org/10.1016/0010-0285(83)90016-6)
- Ninan, D. (2013). Self-location and other-location. *Philosophy and Phenomenological Research*, 87(2), 301–331. <https://doi.org/10.1111/phpr.12051>
- Palmira, M. (2020). Immunity, thought insertion, and the first-person concept. *Philosophical Studies*, 177(12), 3833–3860. <https://doi.org/10.1007/s11098-019-01411-z>
- 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>
- Parrott, M. (2017). Subjective misidentification and thought insertion. *Mind and Language*, 32(1), 39–64. <https://doi.org/10.1111/mila.12132>
- Patel, S. (2023). Thought insertion without thought. *Review of Philosophy and Psychology*. <https://doi.org/10.1007/s13164-023-00689-7>
- Peacocke, C. (2008). *Truly understood*. Oxford University Press.
- Perky, C. W. (1910). An experimental study of imagination. *The American Journal of Psychology*, 21(3), 422–452. <https://doi.org/10.2307/1413350>
- Perrin, D., & McCarroll, C. J. (2023). Immunity to error through misidentification in observer memories: A moderate separatist account. *Philosophy and Phenomenological Research*, 106(2), 299–323. <https://doi.org/10.1111/phpr.12850>
- Picciuto, E., & Carruthers, P. (2016). Imagination and pretense. In A. Kind (Ed.), *The Routledge handbook of the imagination* (pp. 314–325). Routledge.
- Prosser, S. (2012). Sources of immunity to error through misidentification. In S. Prosser & F. Recanati (Eds.), *Immunity to error through misidentification: New essays* (pp. 158–179). Cambridge University Press.
- Pryor, J. (1999). Immunity to error through misidentification. *Philosophical Topics*, 26(1), 271–304. <https://doi.org/10.5840/philtopics1999261/246>
- Putnam, H., (Ed.). (1975). The meaning of ‘meaning’. In *Philosophical papers* (Vol. 2, pp. 215–271). Cambridge U.P.
- Recanati, F. (2007). *Perspectival thought*. Oxford University Press.
- Recanati, F. (2009). *De re* and *De se*. *Dialectica*, 63(3), 249–269. <https://doi.org/10.1111/j.1746-8361.2009.01194.x>

- Recanati, F. (2012). Immunity to error through misidentification: What it is and where it comes from. In S. Prosser & F. Recanati (Eds.), *Immunity to error through misidentification: New essays* (pp. 180–201). Cambridge University Press.
- Recanati, F. (2024). IEM Explained. *Philosophical Psychology*.
- Rice, H. J. (2010). Seeing where we're at: A review of visual perspective and memory retrieval. In J. H. Mace (Ed.), *The act of remembering: Toward an understanding of how we recall the past* (pp. 228–258). Wiley-Blackwell.
- Robins, S. (2020). Defending discontinuism, naturally. *Review of Philosophy and Psychology*, 11(2), 469–486. <https://doi.org/10.1007/s13164-020-00462-0>
- Roessler, J. (2013). Thought insertion, self-awareness, and rationality. In K. W. M. Fulford, M. Davies, & R. Gipps (Eds.), *The Oxford Handbook of Philosophy and psychiatry*. (pp. 658–672). Oxford University Press.
- Roessler, J. (2016). Thinking, inner speech, and self-awareness. *Review of Philosophy and Psychology*, 7(3), 541–557. <https://doi.org/10.1007/s13164-015-0267-y>
- Rothenfluch, S. (2021). A modified self-knowledge Model of thought insertion. *Review of Philosophy and Psychology*, 11(1), 157–181. <https://doi.org/10.1007/s13164-019-00449-6>
- Sant'anna, A. (2018). Episodic memory as a propositional attitude: A critical perspective. *Frontiers in Psychology*, 9, 1220. <https://doi.org/10.3389/fpsyg.2018.01220>
- Seeger, M. (2013). Commentary on Martin & Pacherie. Out of nowhere: Thought insertion, ownership and context-integration. *Consciousness and Cognition*, 22(1), 261–263. <https://doi.org/10.1016/j.concog.2013.01.003>
- Seeger, M. (2015a). Authorship of thoughts in thought insertion: What is it for a thought to be one's own? *Philosophical Psychology*, 28(6), 837–855. <https://doi.org/10.1080/09515089.2014.942897>
- Seeger, M. (2015b). Immunity and self-awareness. *Philosophers' Imprint*, 15(17). <http://hdl.handle.net/2027/spo.3521354.0015.017>
- Shoemaker, S. (1968). Self-knowledge and self-awareness. *Journal of Philosophy LXVC*(19), 555–567. <https://doi.org/10.2307/2024121>
- Shoemaker, S. (1970). Persons and their pasts. *American Philosophical Quarterly*, 7, 269–285.
- Sollberger, M. (2014). Making sense of an endorsement Model of thought-insertion. *Mind & Language*, 29(5), 590–612.
- Sousa, P., & Swiney, L. (2013). Thought insertion: Abnormal sense of thought agency or thought endorsement? *Phenomenology and the Cognitive Sciences*, 12(4), 637–654. <https://doi.org/10.1007/s11097-011-9225-z>
- Stephens, G. L., & Graham, G. (1994). Self-consciousness, Mental Agency, and the clinical psychopathology of thought insertion. *Philosophy, Psychiatry, & Psychology*, 1, 1–10.
- Teroni, F. (2024). Memory identification and its failures. *Philosophy and the Mind Sciences*, 5, 28. <https://doi.org/10.33735/phimisci.2024.10386>
- Tobia, K., Newman, G., & Knobe, J. (2020). Water is and is not H₂O. *Mind and Language*, 35(2), 183–208. <https://doi.org/10.1111/mila.12234>
- Vendler, Z. (1979). Vicarious experience. *Revue de Métaphysique et de Morale*, 84(2), 161–173.
- Verdejo, V. (2021). Perspectives on de se immunity. *Synthese*, 198(10), 10089–10107. <https://doi.org/10.1007/s11229-021-03122-7>
- Verdejo, V. (2023). On the rationality of thought-insertion judgements. *Philosophical Psychology*, 1–24. <https://doi.org/10.1080/09515089.2023.2296580>
- Vosgerau, G., & Voss, M. (2014). Authorship and control over thoughts. *Mind and Language*, 29(5), 534–565. <https://doi.org/10.1111/mila.12065>

- Wiseman, R. (2019). The misidentification of immunity to error through misidentification. *The Journal of Philosophy*, 116(12), 663–677. <https://doi.org/10.5840/jphil20191161242>
- Wright, C. (1998). Self-knowledge: The wittgensteinian legacy. In C. Wright, B. C. Smith, & C. McDonald (Eds.), *Knowing our own minds* (pp. 13–45). Clarendon Press.
- Wright, C. (2012). Reflections on François Recanati’s ‘immunity to error through misidentification: What it is and where it comes from’. In S. Prosser & F. Recanati (Eds.), *Immunity to error through misidentification: New essays* (pp. 247–280). Cambridge University Press.
- Wu, W. (2012). Explaining schizophrenia: Auditory verbal hallucination and self-monitoring. *Mind & Language*, 27(1), 86–107. <https://doi.org/10.1111/j.1468-0017.2011.01436.x>