THE *SELF* FILE AND THE WITTGENSTEINIAN CHALLENGE

Michele Palmira
(Université Computense de Madrid)

Résumé
L'article propose un examen du dossier EGO en abordant un défi, qui peut être tiré des travaux de Ludwig Wittgenstein, selon lequel la distinction entre les utilisations de « je » en tant qu'objet et en tant que sujet est bonne, et qu'elle peut être rendue plus précise en affirmant que les utilisations de « je » en tant que sujet sont immunisées aux erreurs d'mauvaise identification, alors que les utilisations de « je » en tant qu'objet ne le sont pas. Dans cet article, je reconstruis ce défi et l'utilise comme un miroir pour éclairer deux questions importantes concernant la première personne : l'hypothèse selon laquelle « je » obéit à une règle de référence réflexive du penseur, et le rôle que joue la connaissance de soi dans une explication de l'immunité de la première personne.

Abstract
The paper offers an examination of the SELF file by addressing a challenge, which can be elicited from the works of Ludwig Wittgenstein, that the distinction between uses of 'I' as object and as subject is a good one, and that it can be made more precise via the claim that uses of 'I' as subject are *immune to error through misidentification*, whereas uses of 'I' as object are not. In this article I reconstruct that challenge and use it as a foil against which to illuminate two important issues regarding the first person: the hypothesis that 'I' obeys a thinker-reflexive rule of reference, and the role self-knowledge plays in an account of first-person immunity.

1. Introduction

Over the years, François Recanati has developed a rich framework for theorising about the nature of our thoughts which hinges on the notion of a *mental file* (see e.g. Recanati 2009, 2010, 2012a, 2014, 2017). Mental files are mental representations whose main function is to enable the collection and storage of information about the specific object they refer to. The files' reference is not determined via the information stored into the file: on Recanati's own file-theoretic framework, the reference of mental files is fixed via epistemically rewarding relations the thinker bears to the object of their thought, thereby making mental files nondescriptive modes of presentation, or *singular concepts*, of an object.
Within the class of singular thoughts, first-person thoughts – namely thoughts that I may express by using the first-person pronoun, such as 'I was born in Taranto', 'I'm thinking that it will rain', 'My legs are crossed', 'I think therefore I am' – have traditionally taken the pride of place in philosophy. A good chunk of the contemporary literature on the specialness of first-person thought rests on the distinction, which has been long credited to Ludwig Wittgenstein, between two kinds of use of the first-person indexical 'I': on the one hand, there are uses of 'I' as "*object*", which involve an identification of the person who uses it; on the other hand, there are uses of 'I' as "*subject*", which involve no identification at all[1]. Orthodoxy has it that the distinction is a good one, and that it can be made more precise via the claim that uses of 'I' as

[1] See Wittgenstein 1958, 66-67.

subject are *immune to error through misidentification* (henceforth IEM), whereas uses of 'I' as object are not[2].
However, it has been recently argued (Wiseman 2019) that textual evidence shows Wittgenstein never to have endorsed the distinction between two kinds of use of 'I'. In fact, quite the opposite is true: Wittgenstein maintains that, though *prima facie* plausible, such a distinction is heavily burdened with the commitment to a Cartesian metaphysics of the self. This interpretative corrective not only has a significant bearing on Wittgenstein scholarship. For Wittgenstein's remarks in the *Blue Book* also – and, to my mind, more importantly – raise an important challenge to the widespread IEM-based approach to the first person. That challenge, I believe, is an helpful foil against which to illuminate three important issues regarding the first person: the hypothesis that the SELF file obeys a token-reflexive rule of reference and the role self-knowledge plays in an account of first-person IEM.

2. The Wittgensteinian challenge

As correctly pointed out by Rachel Wiseman (2019, 667-668), Wittgenstein considers the possibility of there being distinct uses of 'I' that capture the (alleged) difference between reports of the physical states of one's body ("My arm is broken" [Wittgenstein 1958, 66]) on the one hand, and reports of one's own personal experiences ("I see so-and-so" [Wittgenstein 1958, 66]) on the other. In a key passage of the *Blue Book*, Wittgenstein writes:

> "We feel then that in the cases in which "I" is used as subject, we don't use it because we recognize a particular person by his bodily characteristics; and this creates the illusion that we use this word to refer to something bodiless which, however, has its seat in our body." (Wittgenstein 1958, 69)

The line of thought Wittgenstein is probing seems to be this: If uses of 'I' as subject are somehow special, this must be so since those uses do not involve an identification of the user by means of their bodily features. For when we think thoughts such as 'My arm is broken or 'I have grown six inches' we do identify ourselves via the bodily properties we have, something we don't do when we think 'I see so-and-so' or 'I am thinking that it will rain'. But if our uses of 'I' as subject do not involve any identification of a particular subject via their bodily features, then we must be using 'I' to refer to something which does not have bodily features. So, our uses of 'I' as subject refer only if they refer to a purely mental entity. However, Wittgenstein claims, this is an "illusion" (Wittgenstein 1958, 69). For the only way to justify the idea that uses of 'I' as subject pick out a mental entity is by maintaining "that this body is now the seat of that which really lives" (Wittgenstein 1958, 66). Wittgenstein regards this claim as "senseless", in that "this is not to state anything which in the ordinary sense is a matter of experience" (Wittgenstein 1958, 66).
Wiseman (2019, 664) claims that these remarks contain a challenge to the orthodox view that certain uses of 'I' are special since they give rise to IEM self-ascriptions. Let me first clarify the notions at stake. An error through misidentification occurs when a subject S has certain grounds G that warrant them to judge 'Someone is F'; a subject S is warranted to judge "a is

[2] IEM orthodoxy originates in Shoemaker (1968). For various authors following Shoemaker's lead, see the passages and references quoted in Wiseman (2019, 664 fn. 7, 665-6). Let me clarify that, henceforth, I will use "'I'" to refer to the first-person concept in thought and not to the first-person pronoun in language.

(not) F' partly on the basis of G; and yet, unbeknownst to S, *a* is not (is) F[3]. Consider this case, due to Coliva (2006, 403):

> "(HAT) Walking in the park, across the pond I see a woman wearing an enormous bright red hat. Mistakenly taking that woman for my aunt Lillian, I form the judgement 'Aunt Lillian is wearing an extraordinary hat'."

In (HAT), the visual experience warrants me to judge "that woman is wearing an extraordinary hat". However, I misidentify that woman for aunt Lillian. Now, suppose that my eye doctor comes along and tells me that my sight does not work very well when I have to recognise familiar people, even relatives, at fairly long distances. This, surely, defeats my grounds for the judgement 'Aunt Lillian is wearing an extraordinary hat'. However, this piece of information in no way challenges the fact that my visual experiences afford me with a warrant for 'That woman is wearing an extraordinary hat' (if the hat is indeed extraordinary). Now, a certain judgement is immune to this kind of error when grounds of one judgement do not leave any justificatory gap between 'Someone is F' and 'a is (not) F'[4]. Focusing on the first-person concept, we can define the notion of immunity to error through misidentification accordingly:

> S's judgement 'I am (not) F' made on grounds G is immune to an error through misidentification relative to the concept 'I' just in case it is impossible that G warrant S to judge 'Someone is F' without warranting S to judge 'I am (not) F'.

We can now reconstruct what I'll call "the Wittgensteinian challenge" in a premise-conclusion form:

> (1) Certain uses of 'I' are special because they give rise to self-ascriptions that are IEM.
> (2) If certain uses of 'I' are special because they give rise to self-ascriptions that are IEM, then these uses pick out a mental entity.
> (3) It's not the case that uses of 'I' pick out a mental entity.
> Therefore:
> (4) It's not the case that uses of 'I' as subject are special because they give rise to self-ascriptions that are IEM.

To appreciate the significance of the Wittgensteinian challenge, let us note with Christopher Peacocke that:

> "Philosophical problems about the self and the first person provide a salient illustration of the challenge of integrating the epistemology and the metaphysics of a domain. There has been a persistent impulse amongst thinkers about the self to postulate a transcendental subject of experience and thought. It is an impulse to which Kant, Schopenhauer, Husserl and the early Wittgenstein all yielded. The impulse results from a combination of genuine insight and genuine error. The

[3] This formulation captures, I believe, the core of the phenomenon without going through the rather complex characterisation offered by Pryor 1999. See García-Carpintero (2018), Hu (2017) and McGlynn (2016) for critical discussions of Pryor's own formulation, and for other attempts at simplifying it.
[4] This idea is usually further unpacked by saying that it is not possible for S to acquire defeating evidence that undercuts S's grounds qua warrant for 'I am [not] F', but leaves them intact *qua* warrant for 'Someone is F'. See García-Carpintero 2018, 3320; Hu 2017, 118-119; McGlynn 2016, 41; and Pryor 1999, 284.

> insight consists in the appreciation that there is an Integration Challenge which calls for philosophical solution. The error consists in trying, in this domain, to achieve integration by postulating an exotic domain of the transcendent, rather than by revising and deepening one's epistemology." (Peacocke 1999, 263)

In the *Blue Book* Wittgenstein not only rejects the existence of a transcendental subject of experience and thought but also raises a challenge to the effect that a vindication of the seemingly special epistemic features of the first person does require postulating such an exotic object of reference for 'I'. Sceptics about the specialness of the first person have therefore found in Wittgenstein an unexpected ally,[5] one who is sceptical about the possibility of meeting the integration challenge about the self.
Wiseman describes the task that defenders of orthodoxy have to face somewhat vividly:

> "The orthodox view yokes self-consciousness to IEM, and so the former comes to be framed in terms of a special epistemological capacity or a special form of representation; the philosophical trick is to capture this in a way that does not introduce a special object—namely the seat of consciousness or that which really lives." (Wiseman 2019, 676)

In §§ 3-4 of the article I articulate my response to the Wittgensteinian challenge. My aim is to reject premise (2) by arguing that there's an explanation of the IEM status of introspection-based self-ascriptions that operates only at an epistemological level. The explanation I offer relies on a novel understanding of the thinker-reflexive nature of first-person thought which crucially appeals to the role that introspection plays in fixing the reference of 'I'. I show how this view fits nicely within Recanati's file-theoretic framework. In §5 I respond to some objections. In §6 I offer some concluding thoughts.

3. The pattern of reference of the Self file – first pass

There are two claims that philosophers often make about first-person thought. The first is that first-person thoughts are the thoughts we customarily express via the first-person pronoun 'I'. The second is that first-person thoughts are thinker-reflexive, i.e. they are about the thinker of the thought. Recanati (2014, 506-8) observes that both claims fall short of answering the question of why first-person thoughts are about oneself. For one, certain thoughts can be accidentally about myself although they don't count intuitively as first-person thoughts, when for instance I think "The man who has a stain on his shirt has grey hairs" without realising that I am that man. For another, a purely linguistic characterisation of first-person thought sends us directly back to the conventional meaning – what Kaplan calls the *character* – of the first-person pronoun, stated via the token-reflexive rule that each utterance of the pronoun refers to the utterer of that linguistic item. The pattern of reference of thoughts, however, isn't governed by conventions (see also Morgan 2015, 1801). So, a purely linguistic characterisation of first-person thought is also doomed.
I agree with Recanati's criticism here. And I also agree with his contention that the idea that the pattern of reference of first-person thought is governed by the token-reflexive rule harbours a grain of truth that holds the key to a correct understanding of the nature of the SELF file. Starting from the linguistic case, Recanati (2014: 508) observes that while the token of the expression "The man who has a stain on his shirt has grey hairs" picks out myself as its referential content, this doesn't mean that the type of the expression "The man…" is supposed

[5] See Cappelen and Dever 2013.

to refer to the speaker. By contrast, the reflexivity of the first-person pronoun "I" is encoded in the meaning of the type of the expression. The token-reflexive rule for the SELF file holds a grain of truth precisely because it captures the same kind of non-accidental reflexivity that is captured by the token-reflexive rule for the first-person pronoun. However, as has emerged previously, while the plausibility of the token-reflexive rule for the first-person pronoun hinges on the existence of linguistic conventions, the same cannot be said about the corresponding token-reflexive rule at the level of thought. So, as Recanati puts it (2014, 508): "What we need is a property of the type that plays the same role as conventional meaning plays in the language case".

Again, I concur with Recanati. The challenge of specifying the target property must be further sharpened though. To see why, note that reading through the literature reveals that the token-reflexive rule for 'I' has been stated in a number of different ways:

> "The reference of the first-person is fixed by the simple rule that any token of 'I' refers to whoever produced it." (Campbell 1999, 621)
> "Only I can have a thought about myself by correctly presupposing that my thought is about the owner of the very thought of which this presupposition is an ancillary constituent." (García-Carpintero 2016, 194)
> "[A]ny token first-person thought will be about the subject whose thought it is." (Morgan 2015, 1801)
> "'I' refers to whoever has the control over its production." (O'Brien 2007, 68)
> "∀x ∀event of thinking ϑ: 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, 83)

While García-Carpintero (2016) and Morgan (2015) use the somewhat neutral terms "owner" and "subject" of the thought, Campbell (1999), O'Brien (2007) and Peacocke (2014) choose the seemingly more committal terms "producer" (of the thought) or "agent" (of the event of thinking). Is this a mere labelling choice, or should we pay close attention to how to formulate the reference rule for 'I' in order to find the target property that contributes to explaining why the token-reflexive rule for the SELF file is correct?

Clearly those distinctions *do not* make a difference for the extensional adequacy of the rule. As a matter of psychological fact, if x is thinking T, x is also producing it. We can also grant that all those formulations of the rule are intensionally adequate: if x is thinking and producing T with respect to a given context c and the world of c, x is also thinking and producing T with respect to c and any other world.

However, a reference rule for a concept might also serve the purpose of individuating the target concept. The standard move is to take a reference rule to underwrite the canonical patterns of use of the concept – to be specified in terms of suitable primitive introduction and elimination rules – that one must have an implicit grasp of and be disposed to follow in order to count as possessing the concept and be a competent user of it.[6] Take, for instance, the concept of conjunction '&'. Its rule of reference tells us that a conjunctive thought 'A&B' is true just in case 'A' is true and 'B' is true. This suggests if having the concept '&' involves being disposed to both infer 'A&B' from the truth of 'A' and the truth 'B', and to infer the truth of 'A' and the truth of 'B' from the truth of 'A&B'. I submit that if we operate under this influential approach to concepts, the way in which we articulate the notion of being the owner (thinker) of a thought will make a difference. Let me illustrate this point by harnessing a line of reasoning that plays a prominent role in debates about the first person.

[6] See Peacocke (1992, 2008) for a full articulation and defence of this kind of theory of concepts.

John Campbell (1999, 2002) has famously argued that the best explanation of the phenomenon of thought insertion, a delusion associated with schizophrenia whose hallmark is that subjects report that the thoughts they have introspective access to are not their own, requires making room for two distinct notions of an owner (thinker) of a thought: the owner *qua* "author" and the owner *qua* "recipient" of a thought. In order to count as the author-owner of a thought, the thought "must have been generated by me" (Campbell 2002, 36). That is, if I'm the author-owner of a thought T, the existence of T must be causally explainable fully by appealing to my occurrent and dispositional mental states[7]. In order to count as the recipient-owner of a thought, by contrast, what matters is "the possibility of self-ascription of it by me" (Campbell 2002, 35). That is, to be the recipient-owner of T is for T to be self-ascribed by me on the basis of my introspective awareness of it. Note that Campbell's, O'Brien's and Peacocke's formulations of the token-reflexive rule, given their emphasis on the thinker's production of the thought, seem to fall on the author-ownership side of the author/recipient distinction.

According to Campbell's model of thought insertion, a deluded subject genuinely self-ascribes recipient-ownership of the thought while, at the same time, denying author-ownership of it. This makes the deluded subject's report rationally intelligible and non-contradictory, despite being false[8]. To generalise: a subject can rationally assent to 'I am the recipient of the thought' while dissenting to 'I am the author of the thought'. This shows that the concept AUTHOR-OWNER and the concept RECIPIENT-OWNER are distinct even if, with respect to given a context c, they pick out the same entity in all possible worlds. Thus, Campbell's distinction does make a *hyperintensional* difference, namely the kind of difference we should pay attention to if we aim to individuate the first-person concept and answer the question of why 'I' refers to what it refers to.

In previous work (Palmira 2020, 2022) I have argued that, once we concede with Campbell the possibility of first-person disowned thoughts, we're forced to adopt a recipient-ownership account of the token-reflexive rule for I. The argument, in a nutshell, is that given that a deluded S could competently and intelligibly deploy the SELF file while denying that they are the author of the thought they are introspectively aware of, their competent use of such a file wouldn't follow the pattern of use established by an author-ownership version of the token-reflexive rule. By contrast, their competent use of the SELF file would accord with the recipient-ownership version of the token-reflexive rule, precisely because in cases of first-person disowned thoughts the subjects do ascribe recipient-ownership of such thoughts. I have further spelled out the notion of recipient-ownership in phenomenological terms, observing that there is something that it is like for a deluded subject to undergo the disowned thought. So, at least minimally, to be the recipient of T is to experience it, and to experience T is to have information about the phenomenal likeness T has for me. This allows us to gloss the notion of recipient-ownership of a thought as follows: if S can self-ascribe T on introspective grounds, S is able to engage in an attentional mental activity,[9] i.e. introspecting T, whereby S gains information about what it is like for them to think T.

In light of the foregoing, I offer the following formulation of the token-reflexive rule for 'I':

> ∀x ∀event of thinking ϑ: if ϑ involves the use of 'I', that use of 'I' refers to x just in case x, upon attending to ϑ, is introspectively aware of the phenomenal character ϑ has for x.

---

[7] This does not mean that one must will one's thoughts in order to be the author-owner of them. As is well-known, it's not up to us to choose our beliefs, but those beliefs can be causally explained by the activity of our minds.

[8] Campbell's model of thought insertion is widely accepted, but I should emphasise that it is by no means sacrosanct. For more detail, I refer the reader to the dispute between Campbell (1999, 2002) and Coliva (2002a, 2002b).

[9] See Siewart (2012) and Giustina (2021) for a defence of the idea that introspection requires attention.

To further spell out such a formulation, I endorse the following thesis about the nature of introspection: an introspective state targeting the phenomenal character of an occurrent thought is partly constituted by the target phenomenal character. In contemporary philosophy of mind, this kind of constitution thesis has been carefully refined and defended by Chalmers (2003), García-Carpintero (2018), Gertler (2001), (2012), Horgan, Tienson and Graham (2006), Horgan and Kriegel (2007), Wright (1998) and is often, though not necessarily, connected to acquaintance views of introspection (see in particular Gertler 2012). While a discussion and novel defence of the constitution thesis cannot be offered in the space of this article, let me note that endorsing this kind of constitution thesis does not require endorsing a Cartesian metaphysics of the self. This ensures that the present account of the pattern of reference of 'I' does not beg the question against the Wittgensteinian challenge[10].
I have articulated an account of the reflexivity of the SELF file which hinges on the idea that competent deployment of such a file comes with one's ability to introspect the phenomenal texture of one's occurrent thought. This gives rise to an *introspectionist* account of the reflexivity of the SELF file. Equipped with it, we can now turn to the Wittgensteinian challenge[11].

4. Meeting the Wittgensteinian challenge

On the view on offer, first-person thought has a distinctive hyperintensional profile since one's introspective awareness of what it is like for one to think a given thought enables one to latch onto oneself qua recipient of the thought. This, to put it in Wiseman's terms, is tantamount to saying that the "special form of representation" (Wiseman 2019, 676) of oneself afforded by first-person thought rests on self-consciousness. Insofar as my argument "does not introduce a special object — namely the seat of consciousness or that which really lives" (Wiseman 2019, 676), the link between first-person thought and self-consciousness can be regarded as a solid tie and not as a suffocating yoke.
Importantly, this puts us in a position to redeem one of Wittgenstein's best-known remarks on the meaning of 'I', which reads as follows:

> "The word "I" does not mean "L.W." even if I am L.W., nor does it mean the same as the expression "the person who is now speaking." But this doesn't mean: that "L.W." and "I" mean different things." (Wittgenstein 1958, 67)

Wiseman (2019, 676-7) takes this passage to contain an important insight, namely that 'I' does a special job in our thought and talk and "to describe that job would be to describe the form of life of creatures with self-consciousness" (Wittgenstein 1958, 67).
On the view I advocate, 'I' does not mean the same as 'NN' since they have different hyperintensional profiles. However, this does not mean that 'NN' and 'I' mean different things, for both files are about the same object, i.e. NN. Moreover, 'I' does not mean the same as the expression 'the thinker of this thought' since the token-reflexive rule does not supply descriptive reflexive content one must entertain in order to think of oneself first-personally.

[10] Importantly, the token-reflexive rule is only reference-fixing and not meaning-giving in Kripke's (1980) sense. That is to say, the SELF file does not contribute a token-reflexive description, i.e. *the subject/thinker/agent of this very thought t*, to the truth-conditions of first-person thoughts, so, having a first-person thought does not involve thinking about oneself as the subject of that thought. So, it's not the case when one has a thought such as 'I am thinking T' one is ipso facto thinking 'That subject who is introspectively aware of thinking T is thinking T', thereby ensuring that first-person thoughts don't turn out to be implausibly trivial. Thanks to an anonymous referee for urging me to clarify this issue.
[11] I borrow this label from Verdejo 2023.

Thus, to describe the job of 'I' in a way that also describes the form of life of creatures with self-consciousness, we should start from the fact that 'I' does the job of enabling one to latch onto oneself qua thinker of the thought thanks to one's introspective awareness of the phenomenal character of the thought one is presently thinking has for one. This tallies quite well with the hypothesis that to be self-conscious is to be aware of the phenomenal texture of one's mental life.

By addressing the Wittgensteinian challenge, I have so far illuminated the thinker-reflexive nature of first-person thought. I turn now to argue that the explanatory connection between first-person thought and self-knowledge also ensures and explains why a given class of self-ascriptions is IEM.

Let us restrict our focus to a class of self-ascriptions, namely introspection-based self-ascriptions of occurrent thoughts, such as 'I'm thinking that it will rain'. These self-ascriptions are IEM: insofar as S's introspective awareness as of T passing through S's mind warrants S to judge 'Someone is F' (where 'F' stands for 'thinking T'), such introspective awareness can't but warrant S to judge 'I am F'. This epistemic dependence is guaranteed to hold in virtue of the fact that S's introspective awareness of T contributes to determining the reference of 'I'. To see why, let us suppose that I in fact have an introspective warrant for thinking 'Someone is thinking T'. How to explain this fact?

If I have any introspective access to T at all, then I am in a position to be aware of what it is like for me to think T. This awareness is exactly what underwrites, according to my formulation of the token-reflexive rule for 'I', first-person thinking: if I have introspective access to T and attend to what it feels for me to think T, since 'I' picks out NN in virtue of the fact that NN is introspectively aware of T's phenomenal character, then I cannot be wrong in ascribing T to me. This ensures that I have a warrant for the self-ascription 'I am thinking T'. From this self-ascription, I can then infer a warrant for the logically weaker existential: 'Someone is thinking T'. This account works at the right (i.e. epistemic) explanatory level, for it singles out a property of the grounds of one's self-ascription. The explanation is *metasemantic*,[12] as the relevant property of the grounds is that they contribute to fixing the reference of 'I'. This contrasts with content-based explanations of IEM to the effect that the IEM status of a self-ascription is explained in terms of the distinctive *content* of the *de se* states (In Recanati's version 2007, 2012 of the content account, it is the selfless content of the experiences that ground the self-ascriptions that explain IEM. For discussion, see Colva and Palmira 2024).

The readers familiar with the debate on IEM will immediately wonder how such an explanation fares vis-à-vis self-ascriptions of inserted thoughts. I address this issue at length in my Palmira 2020, where I argue against the claim, originally put forward by Campbell (1999), that thought insertion presents a counterexample to the immunity of the target self-ascriptions[13]. Instead of rehearsing that argument and to move the debate forward, I'd like to show how the metasemantic nature of my account allows us to fend off a well-known criticism to self-knowledge-based explanations of IEM raised by Shoemaker that has been so far left unaddressed[14]. Shoemaker writes:

> "[I]f the supposition that the perception [of my properties] is by "inner sense" is supposed to preclude the possibility of misidentification, presumably this must be because it guarantees that the perceived self would have a property, namely, the property of being an object of *my* inner sense, which no self other than myself could (logically) have and by which I could infallibly identify it as myself. But, of course, in order to identify a self as myself by its possession of *this* property, I

[12] Other authors defending a metasemantic explanation are García-Carpintero (2018) and Peacocke (2014).
[13] Coliva (2002a, 2002b) also denies the counterexample status to reports of inserted thoughts.
[14] Here "self-knowledge" is synonymous with "self-consciousness".

> would have to know that *I* observe it by inner sense, and *this* self-knowledge, being the ground of my identification of the self as myself, could not itself be grounded on that identification." (Shoemaker 1968, 562-3)

Shoemaker targets specifically inner-sense accounts of self-knowledge, but Gertler (2011, 220) points out we can replace "inner sense" with "introspection by acquaintance" without altering the substance of Shoemaker's line of criticism. In fact, we can replace "inner sense" with any view of introspection which accepts the constitution thesis I endorse – to recall: an introspective state targeting the phenomenal character of an occurrent thought is partly constituted by the target phenomenal character – to raise a Shoemakerian worry: if the introspected state partly constitutes the introspective state, this at most ensures that the subject who is introspecting the thought is identical to the subject who is thinking the thought. Yet, this falls short of ensuring that it is *I* who am the subject of the introspected thought. On the metasemantic view on offer, however, there's an important explanatory connection between the pattern of reference of 'I' and introspection: since 'I' refers to what it refers to partly in virtue of my introspective awareness of what it's like for me to think the target conscious thought, this ensures that the (recipient-)thinker of the introspectively accessed thought is me. This shows that accounts of self-knowledge, such as new acquaintance views of introspection, that subscribe to the abovementioned constitution thesis can successfully figure into an explanation of IEM.

This completes my answer to the Wittgensteinian challenge: Self-consciousness can indeed be framed in terms of a special epistemological capacity to prevent errors through misidentification relative (to certain uses of) the SELF file without requiring any metaphysical extravaganza about the self.

I want to close this section by briefly touching on the idea, defended by Annalisa Coliva in a number of works, (2006, 2012, 2017, but see also Echeverri 2020 and Palmira 2022), that the distinctive feature of 'I' is not that certain uses thereof give rise to self-ascriptions that are IEM, but rather that all competent uses of 'I' enjoy what she calls the *real guarantee*, which is "the idea that any competent use of the first-person pronoun (either in speech or in thought) is such that one can't fail to know that the person one is thinking about (or referring to) when one uses it is oneself" (Coliva 2012, 24, fn. 4, see also Coliva 2003, 429).

One might worry that if Coliva's diagnosis of where the epistemic specialness of the first person lies is correct, I have been barking at the wrong tree. For if the (alleged) specialness of 'I' is to be framed in terms of the real guarantee and not in terms of IEM, the Wittgensteinian challenge should be rather formulated as a challenge to the idea that uses of 'I' as subject are special since they exhibit the real guarantee. Thus, my vindication of the idea that certain uses of 'I' give rise to IEM self-ascriptions would not *eo ipso* amount to a vindication of the specialness of 'I'. This worry, however, can be assuaged if we look closely at Coliva's proposal. To begin with, note that Coliva maintains that the SELF file enjoys the real guarantee since 'I am thinking this thought' should be regarded as a definition of the sense of 'I', as opposed to an identification judgement (Coliva 2003, 2017. See also Echeverri 2020). This is tantamount to saying that Coliva subscribes to the token-reflexive rule for 'I'. Now, if one's use of 'I' in accordance with the token-reflexive rule for 'I' is what guarantees that one knows that the person one is thinking about is oneself, and that rule ‴ I maintain ‴ fixes the reference of 'I' on the basis of one's introspective awareness of the phenomenal likeness one's occurrent thought has for one, it follows the real guarantee is (at least partly) determined by one's introspective awareness of the phenomenal likeness one's occurrent thought has for one.

Furthermore, Coliva (2017, 247) conjectures that the real guarantee and immunity to error through (which-object) misidentification might be very close kins[15]. She ultimately rejects this possibility, claiming that while a perception-based self-ascription such as 'My hair is blowing in the wind' is not IEM in the wh-sense, one's competent use of 'My' guarantees that one knows which person one is. I take this point to be correct but compatible with my answer to the original Wittgensteinian challenge, which maintains that *only* certain uses of 'I', i.e. the uses we make when we self-ascribe occurrent thoughts such as 'I think it'll rain' on introspective grounds, are special since they give rise to IEM self-ascriptions.[16] In all such cases, then, the self-ascriptions are IEM and enjoy the real guarantee. If both epistemic properties are partly explained by one's introspective awareness of what it is like for one to think the thought one is presently thinking, then Coliva's initial conjecture that the real guarantee and the notion of immunity to error through (which-object) misidentification are very close kins can be ultimately vindicated.

As far as I can see, we don't have to take a stand on whether the specialness of 'I' is to be explained by the fact that any competent use thereof possesses the real guarantee, or else by the fact that introspection-based self-ascriptions of thoughts are IEM. For the point that is of import here is that the contribution made by my introspective awareness of the phenomenal likeness my conscious thoughts have for me in fixing the reference of 'I' makes it the case that introspection underwrites both the real guarantee and IEM. This takes us to a more fundamental level of explanation of the specialness of 'I' which affords the means to a satisfactory response to (different versions of) the Wittgensteinian challenge.

5. The pattern of reference of the Self file – second pass

The foregoing discussion shows that the introspectionist account of the reflexivity of the SELF file has a lot to go for it. However, stated as it is, the account has to face some challenges, to which I now turn.

Famously, Gareth Evans (1982, §7) maintained that first-person thought can't be fully accounted for without specifying epistemically substantive ways of referring to oneself. This led Evans to question the explanatory payoffs of the token-reflexive rule (what he calls the "self-reference principle", Evans 1982, 258-61). Evans also argued that other ways of knowing oneself "from the inside" besides introspection, e.g. proprioception and episodic memory, both contribute to determining the pattern of reference of first-person thought and give rise to IEM self-ascriptions. Recanati (2007, 2009, 2012a, 2012b, 2014) pushes both Evansian lines in his body of works on the SELF file and IEM. Now, since my account token-reflexive rule appeals to the role of introspection in fixing the reference of the SELF file, my proposal agrees with Evans and Recanati on the epistemic roots of first-person mental reference. However, the introspectionist account of the reflexivity of the SELF file appears to be in stark contrast with the idea that other ways of knowing oneself from the inside contribute to determining the pattern of reference of first-person thought. More specifically, there's a twofold challenge that the supporter of an introspectionist account of the reflexivity of the SELF file has to face. Víctor Verdejo (2021b, 11) raises the challenge in an especially clear way when he writes:

The problem with this suggestion [i.e. the introspectionist account] is that, on the one hand, it would seem to be chauvinistic with respect to other kinds of evidence that would seem to be respectable candidates to take on the reference-fixing role (such as proprioceptive or

[15] See Pryor (1999) for the first systematic attempt at disentangling two notions of (immunity to) error through misidentification, which he labels "*de re*" and "*which-object*" misidentification. Arguably, immunity to error through "which-object" misidentification was the notion Shoemaker (1968) was after. I refer the reader to Coliva (2006), García-Carpintero (2018), Hu (2017) and McGlynn (2016) for discussion of the distinction.

[16] I should clarify: *de jure* IEM. More on this below.

memory-based evidence) and, more importantly, it would still leave unexplained the connection between reference-fixing for *de se* thought and the grounds for non-IEM *de se* judgments.

Verdejo (2021b) contends that the best way to handle this two-fold challenge is to distinguish between the type-individuating reference rule for the SELF file and various perspectives which the thinker may harness in different tokenings of the file (see Verdejo 2021a, 2021b). On Verdejo's view, what type-individuates the SELF file is the reference rule saying that for *x* to be the referent of such a file is for *x* to be the thinker of a contextually salient event of thinking. By contrast, the various perspectives correspond to different kinds of awareness of being the self-referring *x*: cases in which one is self-aware of being the self-referring x are cases in which one thinks a first-person thought from the inside, i.e. by exploiting proprioceptive, introspective, agentive, mnemonic or kinaesthetic relations to oneself.

I have reservations about Verdejo's multiple-perspective token-reflexive rule for the SELF file. For one, it's unclear what the relation between the perspectives and the rule is. Verdejo writes that the perspectives "inform[s]" (2021b, 14) the rule and that the rule "can be seen as itself displaying a number of perspectives on which a thinker may draw" (2021a, 1705). A natural way to further precisify the "informing" and "displaying" would be in terms of the determinable-determinate relation: the token-reflexive rule is a determinable having the various perspectives as determinates. But if this were so, then the token-reflexive rule would type-individuate the SELF file only when supplied with a perspective. This, however, wouldn't do: Verdejo explicitly admits the existence of non-reference-fixing perspectives which correspond to kinds of awareness of being the self-referring *x* which aren't first-personal, for instance, when I see myself in a mirror and judge 'I look fantastic today'. To fix the problem, one might argue that only first-personal perspectives get to determine the token-reflexive rule, but this looks ad hoc. A similar worry arises if we understand the relation between the token-reflexive rule and the perspectives in terms of grounding, or other relations of constitutive determination.

Alternatively, one might resist the idea of there being any metaphysically tight connection between the token-reflexive rule and the perspectives and hold onto the idea that the token-reflexive rule does type-individuate the SELF file without the aid of multiple perspectives. I doubt that this is a tenable position though. If the above considerations about the possibility of rationally intelligible first-person disowned thoughts are on the right track, the notion of a thinker which also features Verdejo's formulation of the token-reflexive rule is in need of being further clarified by taking into account the difference between author-ownership and recipient-ownership of a thought, for that difference does make a difference as to the hyperintensional profile of the SELF file.[17] So, I don't think there's any hope to come up with a type-individuation of the SELF file that stays neutral on this issue.

Although I find the multiple-perspective token-reflexive rule wanting, the two-fold challenge Verdejo raises against the introspectionist account of the reflexivity of the SELF file must be answered. I begin with the question of why introspection, as opposed to proprioception or memory, does play a special reference-fixing role. My answer rests on the idea that while introspection-based self-ascriptions are *de jure* IEM, proprioception-based and memory-based self-ascriptions are only *de facto* so (see Coliva 2006, García-Carpintero 2018, McGlynn 2016, Pryor 1999, Shoemaker 1970 for further discussion)[18]. So, I agree with the Evans-Recanati

[17] In fact, Verdejo (2023) comes very close to the same conclusion, arguing for the Peacockean agentive understanding of the token-reflexive rule on the grounds of a certain understanding of the phenomenon of thought insertion.

[18] Roughly put, *de facto* IEM self-ascriptions are IEM relative to how things are in the actual world but are vulnerable to EM when we consider different possible worlds in which certain abnormal circumstances occur, whereas *de jure* self-ascriptions are IEM relative to all possible worlds. Following Coliva (2006), we should

point that there are varieties of self-reference, which I take to be unified by the fact that they all rely on evidence that makes one's first-person thought at least *de facto* IEM. That is to say, if one refers to oneself by following a rule whose specification involves evidence that ensures the de facto IEM status of the corresponding judgement, one is thinking a genuine first-person thought. So, reference rules specified in terms of introspective, memory-based, proprioceptive, kinaesthetic evidence all determine different ways of genuine self-reference because they all give rise to *de facto* IEM judgements. However, we must distinguish between fundamental and non-fundamental reference rules. Only the fundamental reference rule of a concept C type-individuates C (See Peacocke 1999, 2008, 2014). I believe the distinctive reflexivity associated with the SELF file reveals that the truly special way of referring to oneself rests on introspective grounds, which ensure that one's self-ascriptions made on such grounds will be *de jure* IEM. This suggests that the fundamental and type-individuating reference rule of the mental 'I' is to be cashed out in introspectionist terms. There's nothing chauvinistic about taking introspection to play a starring role in the understanding of the reflexivity of the SELF file. Of course, different authors have different views about which class of first-person judgements exhibit *de jure* IEM. For instance, Daniel Morgan (2019, 2024) argues that also proprioceptive-based and memory-based judgements are *de jure* IEM. If Morgan's arguments were successful, the introspectionist should acknowledge that there's more than one fundamental reference rule that type-individuates the SELF file. This, however, in no way would undermine the response to the Wittgensteinian challenge offered above.

Let me turn now to the question of what fixes the reference of first-person thought in cases where such thoughts aren't grounded in any of the *de facto* IEM-yielding evidence, for instance when I think 'I look fantastic today' upon looking in a mirror. Recanati (2014, 510) does consider this question and provides what I take to be the right way of answering it:

> "Why is the SELF file hospitable to information gained in other ways than the first person way? Because of the following principle governing files: Two pieces of information (or misinformation) are stored in the same file if they are taken to be about the same object. In this way, pieces of information which are not putative items of first person knowledge may go into the same file as first person thoughts which are putative items of first person knowledge."

The thought here is that when I think 'I look fantastic today' upon looking in a mirror, I exploit an implicit identification judgment of the form 'I = the person who's reflected in the mirror' which makes me take that piece of information to be about the same object I refer to when I refer to myself in a distinctively first-personal way, i.e. by thinking about myself on *de facto* IEM-yielding grounds. This ensures the possibility of non-IEM first-person thought.

6. Conclusions

Wittgenstein's *Blue Book* has been traditionally regarded as the first endorsement and defence, in contemporary analytic philosophy, of the thesis that the first person is special. However, closer inspection reveals that Wittgenstein's work also contains a challenge to such a thesis. In this article I have responded to that challenge by defending an epistemically loaded conception of the reflexivity of the SELF file, one which affords the means to vindicating the epistemic specialness of 'I'. While I agree with Recanati that there are various ways of

---

further distinguish between EM relative to one's grounds, and EM relative to background presuppositions, depending, respectively, on whether a mistaken identification component figures in the grounds for one's judgment, or as one of its background presuppositions. IEM would then depend on the absence of such an identification component either in one's grounds or in the background presuppositions.

referring to ourselves in thought, I have argued that the distinctive reflexivity of the SELF file calls for an explanation which gives priority to the role that introspection plays in fixing the reference of the file. This is a commitment that one may not be willing to take up. However, it remains to be seen whether there exist alternative accounts of the reflexivity of first-person thought which are as equally explanatorily powerful as the one developed in this article.

Bibliography

Campbell, J."Schizophrenia, the Space of Reasons, and Thinking as a Motor Process", *The Monist* 82(4), 1999, p. 609-625.
Campbell, J. "The Ownership of Thoughts", *Philosophy, Psychiatry and Psychology*, 9(1), 2002, p. 35-39.
Cappelen, H.&Dever, J. *The Inessential Indexical*, Oxford: OUP, 2013.
Chalmers, D. "The Content and Epistemology of Phenomenal Belief". In: Q. Smith and A. Jokic (eds.) *Consciousness: New Philosophical Perspectives*, Oxford: OUP, 2003, p. 220-272.
Coliva, A. "Thought Insertion and Immunity to Error through Misidentification", *Philosophy, Psychology and Psychiatry* 9(1), 2002a, p. 27-34.
Coliva, A. "On What There Really Is to Our Notion of Ownership of a Thought", *Philosophy, Psychology and Psychiatry* 9(1), 2002b, p. 41-46.
Coliva, A. "The First Person: Error Through Misidentification, the Split between Speaker's and Semantic Reference, and the Real Guarantee", *Journal of Philosophy* 100(8), 2003, p. 416-431.
Coliva, A. "Error Through Misidentification: Some Varieties", *Journal of Philosophy*, 103(8), 2006, p. 403-425.
Coliva, A. "Stopping Points: 'I', Immunity, and the Real Guarantee", *Inquiry*, 60(3), 2017, p. 233-252.
Coliva, A.&Palmira, M. "Immunity to Error Through Misidentification: Some Trends", *Philosophical Psychology*, 2024.
Echeverri, S. "Guarantee and Reflexivity", *The Journal of Philosophy* 117(9), 2020, p. 473-500.
Evans, G. *The Varieties of Reference*, Oxford: Clarendon, 1982.
García-Carpintero, M. "Token-Reflexive Presuppositions and the De Se", in *About Oneself*, M. García-Carpintero and S. Torre (eds.), Oxford: OUP, 2016, p. 179-199.
García-Carpintero, M. "*De Se* Thoughts and Immunity to Error Through Misidentification", *Synthese* 195, 2018, p. 3311-3333.
Gertler, B. "Introspecting Phenomenal States", *Philosophy and Phenomenological Research*, 63, 2001, p. 305-328.
Gertler, B. *Self-Knowledge*, London: Routledge, 2011.
Gertler, B. "Renewed Acquaintance", in D. Smithies and D. Stoljar (eds.), *Introspection and Consciosuness* Oxford: Oxford University Press, 2012, p. 93-127.
Giustina, A. "Introspection Without Judgment", *Erkenntnis* 86, 2021, p. 407-427.
Horgan, T., Tienson, J. L., & Graham, G. "Internal-World Skepticism and Mental Self-Presentation", in *Self-Representational Approaches to Consciousness*, U. Kriegel and K. Williford (eds.), Cambridge, MA: MIT, 2006, p. 191-207.
Horgan, T.&Kriegel, U. "Phenomenal Epistemology: What Is Consciousness that We May Know It So Well?", *Philosophical Issues* 17, 2007, p. 123-144.
Hu, I. "The Epistemology of Immunity to Error Through Misidentification", *The Journal of Philosophy* 114(3), 2017, p. 113-133.
Kripke, S. *Naming and Necessity*, Cambridge MA: Harvard University Press, 1980.

McGlynn, A. “Immunity to Error Through Misidentification and the Epistemology of *De Se* Thought”, in M. García-Carpintero, and S. Torre (eds.) *About Oneself*, Oxford: OUP, 2016, p. 25-55.
Morgan, D. “The Demonstrative Model of first-person thought”, *Philosophical Studies* 172, 2015, p. 1795-1811.
Morgan, D. “Thinking about the Body as Subject”, *Canadian Journal of Philosophy* 49(4), 2019, p. 435-57.
Morgan, D. “Memory and Identity”, *Philosophical Psychology,* 2024.
O’Brien, L. *Self-Knowing Agents*, Oxford: Oxford University Press, 2007.
Palmira, M. “Immunity, Thought Insertion, and the First-Person Concept”, *Philosophical Studies* 177(12), 2020, p. 3833-3860.
Palmira, M. “Questions of Reference and the Reflexivity of First-person Thought”, *The Journal of Philosophy* 119(11), 2022, p. 628-640.
Peacocke, C. *A Study of Concepts*, Cambridge MA: MIT Press, 1992.
Peacocke, C. *Being Known*, Oxford: OUP, 1999.
Peacocke, C. *Truly Understood*, Oxford: OUP, 2008.
Peacocke, C. *The Mirror of the World*, Oxford: OUP, 2014.
Pryor, J. “Immunity to Error Through Misidentification”, *Philosophical Topics*, 26, 1999, p. 271-304.
Recanati, F. *Perspectival Thought*, Oxford: Oxford University Press, 2007.
Recanati, F. “*De re* and *De se*”, *Dialectica* 63, 2009, p. 249-269.
Recanati, F. “Singular Thought: In Defense of Acquaintance”. In: *New Essays on Singular Thought*, ed. by Robin Jeshion, OUP, 2010, p. 141-189.
Recanati, F. *Mental Files*. Oxford: Oxford University Press, 2012a.
Recanati, F. “Immunity to Error through Misidentification: What It Is and Where It Comes From”. In: *Immunity to Error Through Misidentification: New Essays*, ed. by Simon Prosser and François Recanati, Cambridge University Press, 2012b, p. 180-201.
Recanati, F. “First-Person Thought”. In: *Liber Amicorum Pascal Engel*, Julien Dutant, Davide Fassio and Anne Meylan (eds.), Université de Genève, 2014, p. 506-511.
Recanati, F. *Mental Files in Flux*. Oxford: Oxford University Press, 2017.
Shoemaker, S. “Self-Reference and Self-Awareness”, *The Journal of Philosophy*, 65, 1968, p. 555-567.
Shoemaker, S. “Persons and Their Pasts”, *American Philosophical Quarterly* 7, 1970, p. 269-285.
Siewert, C. “On the Phenomenology of introspection”, in D. Smithies and D. Stoljar (eds.), *Introspection and Consciosuness* Oxford: OUP, 2012, p. 129-168.
Verdejo, V. “The Second Person Perspective”, *Erkenntnis* 86(6), 2021a, p. 1693-1711.
Verdejo, V. “Perspectives on *de se* immunity”, *Synthese* 198, 2021b, p. 10089-10107.
Verdejo, V. “On the Rationality of Thought-Insertion Judgements”, *Philosophical Psychology*, 2023.
Wiseman, R. “The Misidentification of Immunity to Error through Misidentification”, *The Journal of Philosophy* 116(12), 2019, p. 663-677.
Wittgenstein, L. *The Blue and Brown Books*, New York: Harper and Row, 1958.
Wright, C. “Self-Knowledge: The Wittgensteinian Legacy”, in *Knowing Our Own Minds*, C. Wright, B. C. Smith, and C. Macdonald (eds.), Oxford: OUP, 1998, p. 15-45.

Acknowledgements

I am grateful to Víctor Verdejo for his tremendous written comments on a previous draft. Thanks also to two anonymous referees for this journal. Work on this article has received

funding from the Spanish Government's Ministerio de Ciencia, Innovacion y Universidades under grant agreements RYC2018-024624-I and PID-2021-123938NB-100.