



Immunity to error through misidentification: some trends

Annalisa Coliva & Michele Palmira

To cite this article: Annalisa Coliva & Michele Palmira (03 Jul 2024): Immunity to error through misidentification: some trends, *Philosophical Psychology*, DOI: [10.1080/09515089.2024.2371423](https://doi.org/10.1080/09515089.2024.2371423)

To link to this article: <https://doi.org/10.1080/09515089.2024.2371423>



Published online: 03 Jul 2024.



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



View related articles [↗](#)



View Crossmark data [↗](#)



Immunity to error through misidentification: some trends

Annalisa Coliva^a and Michele Palmira^b

^aPhilosophy, University of California at Irvine, Irvine, CA, USA; ^bDepartment of Logic and Theoretical Philosophy, Complutense University of Madrid, Madrid, Spain

ABSTRACT

According to a prominent strand of thought in analytic philosophy of mind, certain judgments of the form “*a* is *F*” are such that, although one can be mistaken about what property it is that *a* has, one cannot be mistaken that it is *a* that has the relevant property. Judgments of this kind are said to be immune to error through misidentification (IEM). This article has two main aims. On the one hand, it responds to a need for a systematization of the debate about immunity to error through misidentification, which consists of multiple contributions in epistemology, philosophy of language, philosophy of mind, and metaphysics that have accumulated throughout the last 80 years – since the publication of Ludwig Wittgenstein’s *Blue Book* – and which no one – to the best of our knowledge – has drawn together in an exhaustive and accessible way. On the other hand, the article will take a stand on matters concerning the nature, scope, explanation, and significance of immunity to error through misidentification.

ARTICLE HISTORY

Received 22 January 2024

Accepted 5 June 2024


KEYWORDS

Immunity to error through misidentification; first-person thought; self; self-knowledge; introspection; token-reflexivity

1. Introduction

Suppose that, upon looking in a mirror, I judge: “My legs are crossed”. Unbeknownst to me, the legs reflected in the mirror are not my own but someone else’s, say Fred’s. Suppose now that I formulate the same judgment by attending to my proprioceptive experiences: while it might be the case that my legs are not really crossed – say, my left calf is merely brushing against my right knee – I cannot be mistaken that it is *I* who am instantiating the relevant bodily property. The former judgment involves, to use Sidney Shoemaker’s (1968) nowadays-popular label, an *error through misidentification*, whereas the latter appears to be *immune* to such kind of error.

Immunity to error through misidentification has been suggested to have wide-ranging implications for longstanding debates about the self and the first person, such as debates about Cartesian *vs.* non-Cartesian conceptions

CONTACT Michele Palmira  michelepalmira@gmail.com  Facultad de Filosofía Edificio B, Philosophy Building B, Ciudad Universitaria, Madrid 28040, Spain

The authors are listed alphabetically and they contributed equally to this work.

© 2024 Informa UK Limited, trading as Taylor & Francis Group

of the self (see Evans, 1982), the semantic status of the first-person pronoun “I” (see Anscombe, 1975; Bar-On, 2004; Wittgenstein, 1958), the special epistemology of first-person thought (see Coliva, 2003, 2006; Pryor, 1999; Wright, 2012), the temptation to postulate various exotic entities, such as *res cogitantes*, transcendental egos and the like, as the objects of reference of first-person thought (see Coliva, 2012; Peacocke, 1999), and the need of revising the traditional theory of propositional attitudes (see Cappelen & Dever, 2013).

This article has two main aims. On the one hand, it responds to a need for a systematization of the debate about immunity to error through misidentification, which consists of multiple contributions in epistemology, philosophy of language, philosophy of mind and metaphysics that have accumulated throughout the last 80 years – since the publication of Ludwig Wittgenstein’s *Blue Book* – and which no one – to the best of our knowledge – has drawn together in an exhaustive and accessible way. On the other hand, the article will take a stand on matters concerning the nature, scope, explanation, and significance of immunity to error through misidentification.

We begin by going over the history of the notion of immunity to error through misidentification (henceforth “IEM”),¹ originated in Wittgenstein’s *Blue Book* (1958), further developed by Elizabeth Anscombe in “The first person” (1975), and belabored by Sydney Shoemaker (1968, 1970) and Gareth Evans (1982). We take this excursus to be both of interest in its own right and helpful to set the stage for the main questions addressed in the contemporary debate.

2. Recent history

2.1. Wittgenstein

In the *Blue Book* Wittgenstein draws attention to two different uses of the word “I” (Wittgenstein, 1958, pp. 66–67); namely, the use of “I” “as subject” and “as object”. Examples of the first kind are judgments in which a psychological property is attributed to oneself such as in “I have a toothache”, “I think it will rain”. Examples of the second kind of use are “I have a bump on my forehead”, “I have grown six inches, where a subject self-ascribes physical properties to oneself.”²

According to Wittgenstein, it’s with the second kind of use that the following kind of mistake is possible: one’s judgment may be false not only because the predication component is mistaken but also because the subject of the judgment has been mistakenly identified. By contrast, in the former case one’s judgment may be false only because of an error about what property it is that oneself has. Thus, it has become customary in the

literature on the topic, since Shoemaker (1968), to take uses of “I” as object vulnerable to error through misidentification (EM, hereafter). Notice that, for Wittgenstein, a mistaken identification of a subject consists in a mistake in *recognizing* it – that is, in taking the object presented to one to be identical to oneself. For instance, I see a reflection of a person in a shop window with a prominent bump on their forehead, I take that person to be myself and I judge that *I* have a bump on my forehead. However, unbeknownst to me, that person is someone else.

By contrast, according to Wittgenstein, this kind of mistake is not possible in the case of the use of “I” as subject. The examples Wittgenstein offers of such kind of use of “I” Wittgenstein offers cases when I am non-inferentially attributing to myself a psychological property, like being in pain or thinking that *p*. For, whenever I am in a position to judge, on the basis of introspection,³ “Someone has a toothache”, I am *ipso facto* in a position to judge that *I* have a toothache. According to Wittgenstein, this is so because no identification – viz. no recognition – of a person is involved in these cases. As he puts it, “it is impossible that in making the statement ‘I have a toothache’ I should have mistaken another person for myself, as it is to moan with pain by mistake, having mistaken someone else for me” (1958, p. 67). Hence, these uses of “I” have been taken to be IEM. Importantly, IEM isn’t the same as infallibility: when I judge that “I have a toothache”, it is possible that I’m mistaken about the fact that it is toothache that I’m feeling (suppose that what I’m really feeling is gum pain and not toothache), but I cannot be mistaken about the fact that it is I who am undergoing that feeling. Wittgenstein also appeared to be sensitive to Lichtenberg’s anti-Cartesian claim,⁴ of Humean descent, that in being aware of thinking a thought, one is not aware of a self – conceived of as a mental entity – who is doing the thinking. In such a case, not only would there be no recognition of a self as oneself, but there would be no self presented to one in the first place.

From the fact that no identification – i.e., no recognition – of a person is involved, Wittgenstein draws the conclusion that those judgments in which “I” is used as a subject are not *about* a person at all. This, in turn, entails that these uses of “I” are not about a subject and are not genuinely referring to it. These conclusions are known in the literature as the no-subject and the no-reference view, respectively (Anscombe, 1975; Parfit, 1984).

2.2. Anscombe

In “The first person” (1975), Elizabeth Anscombe radicalizes Wittgenstein’s claim and argues that “I” is *never* used as a referring expression. First, *if* “I” were a referring expression, it would have a sense (a Fregean *Sinn*, cf. Frege, 1918-19). Accordingly, there would be a way of thinking of its referent

which should be such that by means of it (i) one would necessarily *refer* to oneself, and (ii) one would necessarily *know* that one is referring to oneself. For Anscombe, (i) is not enough to characterize the sense of “I” because one’s proper name could fulfill that condition, whence the need to supplement it with (ii).⁵ Thus, in the explanation of the sense of “I” – “I” is the expression that everyone uses to refer to oneself” –, “oneself” is in fact an occurrence of the *indirect reflexive* pronoun, which is nothing but the very occurrence of “I” in indirect speech. This account of the sense of “I” would thus be circular and could not clarify what it means to have I-thoughts, according to Anscombe.⁶

Second, if “I” were to refer to one’s self, that self – conceived as an object – should be necessarily present and recognized as such in each use of “I”. This, according to Anscombe, excludes the body as a candidate for playing the role of the self, for one could use “I” in a state of sensory deprivation, or could not recognize one’s body as one’s own. Nor could the self be identical to a bundle of memories, for one could use “I” in a state of amnesia. According to Anscombe, only a Cartesian ego – that is a thinking substance – or even “a stretch of one” (1975, p. 31) would do. For it would be present in each act of thinking of oneself and – by endorsing the idea that it would be transparently given to one – it would secure the fact that one would know that one is indeed presented with one’s own self. Thus, Anscombe concludes, if “I” were a referring expression, “Descartes was right about what [its] referent was” (*ibid.*). According to Anscombe, to avoid embracing the Cartesian conception of the self, we should deny the main premise – viz., that “I” is a referring expression at all.

According to Anscombe, the role of “I” is similar to the role of “it” in “it’s raining”, which is not a referring one, but rather an indication that rain is occurring at the place and time in which “it” was uttered or thought. Similarly, “I” indicates one’s awareness of the fact that certain states and actions are occurring where and when “I” is being uttered or thought. Such states and actions are “subjectless”, according to Anscombe (1975, p. 36), as they do not contain a representation of a subject. Likewise, their self-ascriptions “do not involve the connection of what is understood by a predicate with a distinctly conceived subject” (*ibid.*).⁷ Realizing that “I” is not a referring expression should thus free us from “the (deeply rooted) grammatical illusion of a subject” (*ibid.*) with all the problems that that illusion generates (See [Section 5.2](#)).

Anscombe and Wittgenstein’s views on IEM seem to leave us with a dilemma: either we are bound to adopt a no-reference view about the first person or the attempt to make sense of IEM should lead us to posit metaphysically extravagant referents for “I”. Anscombe and Wittgenstein embraced the first horn of the dilemma, but the claim that the mental or linguistic “I” does not refer is highly revisionary and we do regard it as

a last resort move. Our dissatisfaction with the no-reference view of “I” is common ground in the dispute between Sidney Shoemaker and Gareth Evans. They both agree that we need not accept a no-reference view of the first-person concept to make sense of IEM. Yet, Shoemaker and Evans agree on this much. But they disagree over whether IEM is a phenomenon that does in the end speak in favor of a Cartesian conception of the self.

2.3. *The Evans-Shoemaker dispute*

In “Self-reference and self-awareness” (1968), Sydney Shoemaker distances himself from the no-reference view propounded by Wittgenstein and Anscombe. Yet, he endorses Wittgenstein’s claim that all and only (non-inferential) psychological self-ascriptions are “immune to error due to a misrecognition of a person, or, as I shall put it, they are immune to error through misidentification relative to the first-person pronoun” (Shoemaker, 1968, p. 556).⁸ He then defines EM as follows:

[T]o 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” (1968: 557).⁹

Non-inferential psychological self-ascriptions, in contrast, are such that if one knows some particular thing to be φ one cannot make the mistake of asserting “I am φ ” because one mistakenly thinks that the thing one knows to be φ is identical to oneself (ibid.).

For Shoemaker, even though no identification – that is, no recognition – of oneself is taking place in the use of “I” as subject, “I” refers to oneself. In this respect, “I” is similar to a perceptual demonstrative – like “this” or “that” – which may refer to a particular object, while not involving any recognition of it. In fact, the absence of recognition and therefore of any identification explains why, according to Shoemaker, these uses of “I” are immune to error through misidentification.

For Shoemaker, however, the referent of the demonstrative is determined by the speaker’s intentions, whereas the referent of “I” is determined by the rule that “I” is the expression one uses to refer to oneself.¹⁰ Furthermore, whereas a demonstrative may fail of reference, if one is hallucinating an object, “I” cannot be subject to this kind of failure. Yet, realizing that “I” does not behave exactly like a demonstrative should not lead to the conclusion that it is not a referring expression at all.

According to Shoemaker, self-reference without self-identification must be possible because any self-identification will involve the identity statement “*a* = I” and, to entertain such an identity statement, one needs a prior grasp

of what “I” refers to. Yet, he does not think that self-awareness is awareness of an object that is presented to one in thought.

Hence, for Shoemaker, “I” is puzzling. For it partly behaves as a demonstrative in that it refers without identification (understood as recognition) and yet it does not involve singling out or being presented with oneself as an object in thought. Yet, he thinks that this puzzlement depends on the endorsement of a perceptual model of self-awareness (or self-knowledge). In contrast to the perceptual model, according to him, “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” (1968, pp. 563–564). Thus, according to Shoemaker, by conceiving of self-knowledge as not involving any substantive cognitive achievement concerning the singling out of a subject in thought to which psychological properties would inhere, one can fully appreciate how in the use of “I” as subject, “I” does refer, even if no recognition or even singling out of oneself as an object of thought occurs.

As has emerged previously, Shoemaker initially sided with Wittgenstein in claiming that all and only (non-inferential) psychological self-ascriptions are IEM. In “Persons and their past” (Shoemaker, 1970), Shoemaker partly corrects his earlier views and holds that also some memory-based self-ascriptions which are not based on information acquired through any identification component are IEM. IEM is thus “*preserved* in memory” (1970, p. 270). Furthermore, Shoemaker concedes that “remembering” entails retaining knowledge of one’s own past. Still, he thinks we may envisage cases of “quasi-memory” (particularly in fictional cases of fission or of partial brain transplant).¹¹ That is, cases in which, at t_1 , (i) one has an apparent memory of oneself* being F at t_0 . Yet, (ii) the apparent memory embodies information deriving from the perception of that event by a person who is not necessarily oneself.

According to Shoemaker, this shows that also in the normal run of cases, memory-based self-ascriptions in which one remembers a past event “from the inside” (1970, p. 273) are based on the identity “I = the person whose past is responsible for the memory impressions I am having”. Such an identity is true and hence “in our world all quasi-remembering is remembering” (1970, p. 270). Still, in a different possible world, one’s memory impressions “from the inside” could in fact derive from someone else’s past thus making that identification component false. Memory self-ascriptions are therefore only *de facto* immune to error through misidentification: whereas in the real world the person whose past is responsible for one’s memory impressions is oneself, in a different possible world where fission and partial brain transplant occurred, it could be someone else.

In *The Varieties of Reference* (1982), Gareth Evans takes issue with Shoemaker’s account of memory-based self-ascriptions. Evans proposes

a theory of I-thoughts, rather than statements, which conceives of them as similar to demonstrative and here-thoughts. Thus, he is not primarily interested in the linguistic use of “I”, but rather in thoughts containing the first-person concept. His discussion of IEM takes place in the context of specifying the special ways in which one has knowledge of oneself *qua* oneself – a kind of knowledge of oneself that is available only to the subject. By so doing, Evans intends to build on Frege’s observation, in “The thought” (1918–19/1956: 298), that “everyone is presented to himself in a particular and primitive way, in which he is presented to no-one else”. Before presenting Evans’s views on IEM, however, it’s important to give a brief sketch of his ideas about propositional thinking in general.

Evans’s starting point is the observation that, insofar as thoughts have concepts as constituents, they must be systematically connected in such a way that the thought “John is happy” has something in common both with the thought “Harry is happy” and the thought “John is sad”. This gives rise to a condition on propositional thinking that Evans “Generality Constraint”, which he states as follows (Evans, 1982, p. 104): “If a subject can be credited with the thought that *a* is *F*, then he must have the conceptual resources for entertaining the thought that *a* is *G*, for every property of being *G* of which he has a conception.” In the case of I-thoughts, this means that one must be able to entertain thoughts which are not available to one simply on the grounds of these special ways of thinking of oneself. For instance, one must be able to make self-ascriptions of non-psychological properties which depend on other people’s testimony, like “I was born in 1973”. Only in this way would one’s first-person concept be a concept of an objective entity. For these thoughts would involve an identification of the form “I = the person whose documents say that she was born in 1973”, where the second term of the identity would be available, at least in principle, to anyone, and refers to a person rather than to a mental entity like a Cartesian ego. Moreover, if one’s self-ascription of a psychological property obeys the generality constraint, this means that it must be possible for one to ascribe that property to some other individual *a*. Thus, for Evans, to have I-thoughts we must conceive of ourselves as persons, potentially locatable in a “spatio-temporal map of the world” (1982, p. 211).

The adoption of the generality constraint helps us see how Evans’s take on I-thoughts departs from both the Cartesian and the no-reference views of the first person. If the Cartesian view were correct, our self-ascriptions of psychological properties wouldn’t bear any systematic connection to our self-ascriptions of non-psychological properties, for we wouldn’t be in a position to predicate non-psychological properties of a purely mental substance. If the no-reference view were correct, ascribing a psychological property *F* to other individuals by thinking “*a* is *F*” wouldn’t bear any systematic connection to ascribing the same property to ourselves, for in

such a case there would be no subject *a* to which *F* is ascribed. Both views, then, violate the generality constraint.

Evans further objects to the conclusion that “I” is not a referring expression reached by Anscombe based on the observation that we can have I-thoughts in states of sensory deprivation or amnesia. For him, in contrast, it is enough for “I” to refer, to “be *disposed* to have such thinking controlled by information which may become available to [one] in each of the relevant ways” (1982, p. 216, emphasis added). Dispositions, in turn, are had even when they are not deployed, like in cases of sensory deprivation or amnesia.

Evans then delves into a characterization of “each of the relevant ways”. IEM, for Evans, is key in this respect and, contrary to Wittgenstein and Shoemaker, he notices that, besides non-inferential mental self-ascriptions,¹² also some physical and memory-based self-ascriptions (even when originally grounded on an identification) are IEM. Consider again “My legs are crossed” based on proprioception. In such an event, writes Evans, “it seems . . . not to make sense for a subject to utter ‘Someone’s legs are crossed, but is it I whose legs are crossed?’” (ibid.). He then adds that in this case there is an identification of a person without any identity judgment. He perceptively notices that “the word ‘identify’ can do us a disservice here” (1982, p. 218), because of its ambiguity between “individuation” (or “singling out” an object) and “recognition”. Somatic proprioception and memory, for Evans, are ways of knowing about our physical and past properties which do not involve the recognition of a subject and yet produce or retain individuating knowledge about it. Similarly, perception allows us to know our position, orientation, and relation to other objects in the world without any identification. By focusing on the way one knows of instantiating a property, rather than by focusing on the kind of property in question (mental or physical), Evans is able to dislodge the idea that only (non-inferential) mental self-ascriptions are IEM, by showing that also physical self-ascriptions are IEM when based on proprioception.

More specifically, for Evans, there is no gap “between the subject’s having information (or appearing to have information) in the appropriate way, that the property of being *F* is [or was] instantiated, and his having information (or appearing to have information) that *he* is [or was] *F*” (1982, p. 221). That is, “for him to have, or appear to have, the information that the property is [or was] instantiated *just is* [emphasis added] for it to appear to him that *he* is [or was] *F*” (ibid.).

Evans discards the idea that some articulation should be added in these cases due to the possibility of deviant causal chains, through which we would be receiving or storing information about someone else’s body, whereabouts, or past. The possibility of deviant causal chains would merely show that these self-ascriptions are open to error, but it would not show that they are based on an identification component of the form “I = the person

whose body I am receiving information from/whose whereabouts I am perceiving/whose past is responsible for the memory impressions I am having”.

Such collateral information, for Evans, would be involved “if one *knew* one was in the abnormal situation described; but it is surely too sophisticated to be discerned as an element in the normal case of judgements of the kind we are considering” (ibid., emphasis added). Furthermore, he objects to the possibility that these faculties could be “identity-neutral” (1982, p. 244) ways of having knowledge of the relevant states of the person, in the sense that faculties such as memory, introspection, proprioception offer knowledge of an object which doesn’t leave its identity with the subject as an open question.¹³ These special ways of gaining or retaining knowledge about oneself have been forged through natural selection, for Evans, and give us knowledge of ourselves as physical and temporally enduring entities. Thus, our first-person concept (or *Idea*, as he calls it) is firmly anti-Cartesian.

Summing up. Shoemaker maintains that the possibility of there being q-memory cases show that, in such cases, one’s memory-based self-ascriptions rely on an identity component and are therefore vulnerable to EM. Evans’s reply is that this possibility doesn’t affect the point that memory, just like introspection, doesn’t offer recognitional knowledge of the subject *qua* bearer of such-and-such properties. So, memory-based self-ascriptions of physical properties are identification-free and, for this reason, always IEM. In contemporary times, some authors have sided with Shoemaker, others with Evans, and yet others have tried to offer a conciliatory resolution of their debate. What’s common to all such strategies is the disentanglement of different notions of EM and IEM, to which we now turn.

3. The varieties of (immunity to) error through misidentification

The kind of error involved in the notion of error through misidentification does not concern *what* property it is that *a* has but whether it is *a* that instantiates the relevant property. Bear also in mind that an error through misidentification occurs partly in virtue of the *grounds* upon which subjects make their judgments. The same judgment, “My legs are crossed” is vulnerable to EM when made based on the perception of a reflection, but IEM when made based on proprioception. Thus, we need to be attentive to a judgment’s grounds and not just to its content. In agreement with the literature (see Coliva, 2006; McGlynn, 2016; Palmira, 2020; Pryor, 1999; Wright, 2012), let us take one’s judgment’s grounds to be the explanatory basis of one’s judgment. So, let us think of the judgment’s grounds as the psychological causes of the formation of one’s judgment, as opposed to the

thinking of them as premises in the conscious reasoning one performs to arrive at one's judgment. For instance, I can judge "Alice is reading *Infinite Jest*" without making any conscious inference. Yet, if I were asked to explain why I did judge that Alice is reading *Infinite Jest*, I would say something like: "I've seen a person reading *Infinite Jest*, and that person is Alice".

This clarified, as the survey of the early history of EM has made apparent, the participants to the debate understood that notion as involving a mistaken identification component of the form "I = a" as part of one's grounds for the final self-ascription "I am F". Nowadays, this kind of EM is known as *de re* misidentification. Let us say that *de re* EM occurs when:

- (i) S judges "a is F"
- (ii) S's warrant for "a is F" rests on their warrant for judging, of some b, and of a, that b is F and that b = a.
- (iii) However, unbeknownst to S, $b \neq a$.

The corresponding IEM would then be due to the absence of such an identification component as part of one's grounds for the judgment "a is F". The Shoemaker-Evans dispute can thus be seen as a disagreement concerning which grounds for the relevant self-ascription contain such an identification component.

More recently, Jim Pryor (1999) has proposed another variety of EM, labeled "which-misidentification" (wh-EM, hereafter) and has used it partly to arbitrate the Evans-Shoemaker dispute. Consider the following case, due to Pryor (1999, p. 281):

I smell a skunky odor, and see several animals rummaging around in my garden. None of them has the characteristic white stripes of a skunk, but I believe that some skunks lack these stripes. Approaching closer and sniffing, I form the belief, of the smallest of these animals, that it is a skunk in my garden. This belief is mistaken. There are several skunks in my garden, but none of them is the small animal I see.

The key difference between *de re* and wh-EM is that while, in the former case, the grounds one has for one's judgment justify one in believing of some particular object *b* that it is *F*, and one mistakenly takes *b* to be identical with *a*, in the latter case one's grounds do not justify one in believing of a particular object *b* that it is *F*, but only that *something* is *F*. As Pryor (1999, p. 283) puts it: "In cases of *which*-misidentification, I go wrong not in *re-identifying* the thing I know to be *F* as some other thing; rather I go wrong in figuring out *which thing* is *F* in the first place". On Pryor's view, *de re* and wh-EM are different instances of the same phenomenon since they specify two distinct ways in which one's grounds warrant the existential judgment "Something is F" without thereby warranting one to judge

“a is F”. As a consequence, at the core of IEM there is an epistemic dependence between singular and existential judgments which stems from the fact that one’s grounds do not leave any justificatory gap between “Something is F” and “a is F”.¹⁴

There are different proposals on how to unpack this epistemic dependence condition in the case of wh-IEM. On Pryor’s view (1999, p. 284), if “a is F” is wh-IEM then it’s not possible for one to acquire defeating evidence that *undercuts* one’s grounds qua warrant for “a is F” but leaves them intact qua warrant for “Something is F”.¹⁵ Equipped with this notion, Pryor envisages the case in which one is told that some of one’s memories are quasi-memories (“q-memories” for short), and that none of one’s memories as of being F derive from one’s own past (1999, p. 285). According to Pryor, this information undercuts one’s memory-based grounds for judging “I was F” without *ipso facto* undercutting one’s grounds for “Someone was F”. Thus, while memory-based self-ascriptions are *de re* IEM, they are open to wh-EM. By so arguing, Pryor sides with Shoemaker in the Shoemaker-Evans dispute.

As observed by Daniel Morgan (2019, pp. 444–5), however, it’s unclear that the example really satisfies Pryor’s undercutting defeat condition on wh-EM, for the information that one is victim of quasi-memories appears to be *additional* information that warrants the existential judgment.¹⁶ So, it’s unclear that Pryor’s own definition returns the verdict that memory-based self-ascriptions are open to wh-EM.

Morgan (2019, pp. 446–7) suggests an alternative interpretation of wh-IEM. The core idea is that the grounds warranting “a is F” do not offer independent knowledge of “Something is F”, where the notion of independence knowledge is spelled out as follows: the existential judgment cannot be known on those grounds without the singular judgment being true.¹⁷ On Morgan’s notion of wh-IEM, memory-based self-ascriptions turn out to be wh-IEM: one’s judgment “Someone was F” in a quasi-memory scenario is, to put it in Evansian terms (1982, p. 221), a “wild shot in the dark” that can at best be luckily true, thereby falling short of knowledge. Thus, *contra* Pryor, even if we construe the Shoemaker-Evans dispute as a dispute about wh-IEM we can side with Evans in this dispute.

Coliva (2006) criticizes wh-EM. First, Coliva claims that wh-EM does not respect the fact, appreciated by both Shoemaker and Evans, that q-memories are not identity neutral. That is, they do give one information that one was or seems to have been *F*, and not just that something was or seems to have been *F*. Secondly, and more generally, Coliva contends that there is no genuine phenomenon of wh-EM. Focusing on Pryor’s example, she maintains that the rational grounds for the judgment “This animal is a skunk” do involve an identification component. For I smell a skunky odor, and this gives me grounds for:

- (1) The animal (in my garden) which is actually responsible for this odor I smell is a skunk.

I then go close to one of the animals and sniff, while still smelling the skunky odor. I then judge (3) “This animal is a skunk” *because* I believe:

- (2) This animal (I can now see) = the animal (in my garden) which is actually responsible for this odor I smell.

According to Coliva (2006, p. 413): “The source of Pryor’s mistake lies . . . in too narrow a conception of the range of *concepts* which may feature in an identification component. An identification component may involve either *de re* concepts . . . [; or] else, it may involve non-*de re* but still singular concepts”.¹⁸ To illustrate the difference: in the case where, upon looking in a mirror, I judge: “My legs are crossed”, the false identity claim “that person = myself” features only *de re* concepts, that is, concepts that are grounded in the subject’s identifying knowledge of the objects in question. By contrast, Coliva maintains, in Pryor’s example the false identity claim is “This animal = the animal (in my garden) which is actually responsible for this odor I smell”, which involves a descriptive concept that uniquely singles out the object in question.

Wright (2012) has presented two further examples to argue in favor of wh-EM. The key feature of them is that the grounds at one’s disposal would seem to support only an existential generalization. For instance, I see footprints in the sand going around in circles and form the belief “Someone is going around in circles”; I then conclude that I am going around in circles. Or else, I go to a palmist, who tells me that this week I will be incredibly lucky, and I form the belief that I will win the lottery, based on the existential generalization that someone who wins the lottery is incredibly lucky.

However, it merits note that unless one takes oneself to be identical to the person who has made the footprints in the sand, it is hard to see how one could rationally arrive at the self-ascription “I am going around in circles”.¹⁹ The palmist case too, albeit based on a fallacious inference, will have to be grounded on a further identification of oneself with the person who is going to win the lottery this week. Thus, it remains unclear whether wh-EM is a genuine alternative to *de re* EM.²⁰ If so, it remains unclear whether wh-EM can be appealed to in order to resolve the Evans-Shoemaker dispute we have examined in [Section 1.3](#).

Coliva has then proposed a different resolution of the Evans-Shoemaker dispute, based on the distinction between a subject’s available grounds for their judgment and a judgment’s background presuppositions. Let us take memory-based self-ascription “I was in Scotland five years ago”. As persuasively argued by Evans (1982), one neither arrives at such a self-ascription by

going through a piece of conscious reasoning that takes identity belief “I am the person from whose past this memory impression derives” as a premise, nor would one offer such a belief as an explanation of why one is making that self-ascription. For, if one were asked how one arrived at self-ascription, one would most likely only cite one’s memory experiences. This has led Coliva and many others (see e.g., García-Carpintero, 2018; Wright, 2012) to say that the identity belief “I am the person from whose past this memory impression derives” is not foregrounded in one’s own psychology and, for this reason, is not part of the grounds of one’s self-ascription. However, that identity belief is a background presupposition of one’s self-ascription. If (one knew that) the identity belief were not in place, then one couldn’t rationally hold “I was in Scotland five years ago”.

One can thus 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. One may then argue that while memory-based judgments are IEM relative to one’s grounds for the judgment, they remain vulnerable to error through misidentification relative to background presuppositions. Coliva (2006, p. 420) thus concludes that “Evans would . . . be right on the former understanding of IEM, while Shoemaker would be right on the latter”. Another option would be to maintain that there is EM relative only to one’s grounds but that the presence of an identification in the background presuppositions allows for cases in which, due to the collateral information that one is not in the normal condition of storing information about one’s past, that presupposition could be moved from the background into the grounds, while still respecting Evans’ point that in this case too q-memories would not be identity-neutral.

We have distinguished between three varieties of EM: *de re*-EM occurs when one forms a judgment that is grounded on a false identity belief; *wh*-EM occurs when one forms a mistaken judgment about which thing has a certain property even though one’s grounds still warrant the corresponding existential judgment; background presupposition-EM occurs when one forms a judgment that features in its background presuppositions a false identity belief. We turn now to distinguish two senses in which one’s judgment can be immune to these three types of EM.

The Shoemaker-Evans dispute also suggests that the modal force implicit in the notion of immunity can be understood in two ways: while certain judgments 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, other

judgments are IEM relative to all possible worlds. The former class of judgments will be merely *de facto* IEM, whereas the latter will be logically IEM. Shoemaker's view, for instance, is that memory-based self-ascriptions of physical properties are merely *de facto* IEM, for in a different possible world one's memory experiences could derive from someone else's past. Evans rejects this view though, maintaining that the target self-ascriptions are logically IEM. As we will presently see, this has important consequences regarding the understanding of thought insertion.

Importantly, the distinction between *de facto* and logical IEM crosscuts the distinction between *de re*-IEM, wh-IEM, and background presupposition-IEM: any such variety of IEM is amenable to the *de facto*/logical distinction. So, it is possible that one's judgment be logically *de re*-IEM while, for instance, being only *de facto* background presupposition-IEM. This, on Coliva's view, would happen when a judgment is based on the presence in the background presuppositions of a *contingently true identification component*, which may be false in a different possible world, and on a different state of information of the subject who forms the judgment. By contrast, logical IEM would depend either on the absence of any identification component in the background presuppositions, or else in the presence of an a priori true one, such as, for instance, "I = the thinker of this [introspectively available] thought" or "I = the subject of this [introspectively available] sensation". (for an articulation and defense, see [Section 4.3](#))

McGlynn (2016) proposes to rethink the distinction between *de facto* and logical IEM not as a difference in kind but only in degree of modal safety with respect to the relevant knowledge claims.²¹ In his view, all self-ascriptions are in principle open to EM, but some are modally safer than others. In particular, psychological, non-inferential self-ascriptions would thus be safer than self-ascriptions based on q-memories, or somatic proprioception.

McGlynn's proposal doesn't convince us (see also Coliva 2018). For one, it is important to bear in mind that IEM shouldn't be equated with infallibility, insofar as IEM self-ascriptions allow for the possibility of a mistake in predication; so an account of IEM based on the degree of safety enjoyed by such self-ascriptions looks pointless to account for the phenomenon in its generality, for it is well-known that safety implies truth. For another, McGlynn's account does not explain why some self-ascriptions would be modally insecure regarding the identification of the *subject* who has a certain property, as opposed to being modally insecure concerning the *property* the subject self-ascribes. Were McGlynn to supplement his proposal to account for the difference, it is likely that he would have to involve background identification components similar to the ones presented in Coliva (2006).

4. The scope

4.1. Are non-inferential psychological self-ascriptions logically IEM? – the case of thought insertion

As alluded to in the previous discussion, depending on the varieties of IEM we countenance, we can give different verdicts about which classes of self-ascriptions enjoy a (logical or de facto) IEM status. Nowadays, however, there is wide consensus that the kind of self-ascriptions we have been reviewing, such as bodily self-ascriptions based on somatic proprioception, self-locating ones based on perception, and memory-based ones, are all IEM, even though maybe only de facto so.²² Interestingly, authors such as Anscombe, Evans, Shoemaker and Wittgenstein all agree that introspection-based self-ascriptions of psychological properties are logically IEM. However, the logical IEM of non-inferential psychological self-ascriptions has been challenged by Campbell and Sugden (1999), based on the phenomenon of thought insertion.

Thought insertion is a characteristic symptom of schizophrenia, whereby subjects are introspectively aware of typically emotionally charged thoughts, such as “Kill God”, which they disown and attribute to someone else who allegedly inserted them into their heads (see Frith, 1992). Campbell (2003, p. 39) aims to provide an account of thought insertion that belongs to “a wave of analyses in cognitive neuropsychiatry in which delusions are explained as broadly rational responses to highly unusual experiences”. For this reason, he proposes a distinction between two notions of a thinker of a thought T²³: the thinker *qua* “agent” and the thinker *qua* “owner” of T. In order to count as the author-thinker of T, the thought “must have been generated by me”.²⁴ In order to count as the owner-thinker of a thought, by contrast, what matters is “the possibility of self-ascription of it by me”,²⁵ that is, the fact that I can self-ascribe T in a distinctively direct and non-observational way.²⁶ Campbell uses this distinction to maintain that a deluded S could make introspection-based self-ascriptions of this kind: “I am owner-thinking that *p*, but I’m not author-thinking it”. According to Campbell, the possibility of first-person inserted thoughts would show that one can be introspectively aware of a given thought and yet go wrong in identifying whose thought it is.

To represent a counterexample to the widespread idea that non-inferential psychological self-ascriptions are logically IEM, these cases should be construed as resting on a background assumption of the form “I = the thinker of this thought I am presently aware of”. Furthermore, such a background assumption should be only contingently true.

According to several theorists (Coliva, 2012, 2017; Shoemaker, 1968), however, non-inferential psychological self-ascriptions aren’t based on any identification component – certainly not in a subject’s own grounds for the

judgment – nor, indeed, in the background presuppositions of it. Other theorists, in contrast, allow for the presence of such an identification component, but consider it a priori true (Peacocke, 1999, 2014). Be that as it may, the intuition shared by many is that in being introspectively aware of a given mental state – let it be a phenomenal one like pain, or a propositionally attitudinal one, like a thought or a belief – one is, tautologically, aware of oneself being in pain, or thinking that thought (Shoemaker, 1968, p. 89). As Coliva (2002b, p. 28) puts it: “[I]f X were introspectively aware of Y’s pain, or Y’s belief that it is sunny today, this would mean that X herself would be feeling pain or believing that it is sunny today. Hence that pain or belief would be X’s own”. According to Coliva, this also explains the difference between *de facto* and logical IEM. If one’s self-ascriptions based on somatic proprioception or memory derived from someone else’s body or past, that body or past would not count as one’s own; whereas feeling a given sensation or thinking a given thought suffices for making them one’s own, even if they causally derived from someone else’s mind.²⁷ That is, we don’t have an independent grasp of what counts as one’s own mental state apart from the fact of being introspectively aware of it.

According to Coliva (2002b, b), this is crucial in making us count thought insertion as a symptom of *irrationality* – and thereby of mental *illness* – instead of *erroneous* ascription of a thought to oneself. Consequently, we then try to *cure* the subject who makes such self-ascriptions, rather than merely ask them to *correct* their judgment. That said, Coliva proposes to redescribe thought insertion as a case where while one still owns one’s thoughts, one has lost a sense of authorship and/or agency over them. That is, one is introspectively aware of thoughts that one does not experience as integrated in one’s own cognitive life. As Stephens and Graham (2000: 172–176) point out, when thoughts are phenomenologically experienced as nested in a web of beliefs and desires that one does not recognize as one’s own, this can lead to thinking of them as produced by alien intentionality. Thus, according to Coliva, it is possible to account, causally, for how ascriptions of one’s thoughts to others are possible, without thereby denying that they remain, logically, one’s own.

The Campbell-Coliva dispute on thought insertion operates with the notion of *de re* EM. However, Hu (2017) has argued that the possibility of first-person inserted thoughts shows that introspection-based self-ascriptions of psychological properties are open to wh-EM. Accepting Campbell’s model of thought insertion and deploying Pryor’s definition of wh-IEM, Hu (2017, p. 125, fn. 35) asks us to consider a case in which a schizophrenic subject S who judges “I am owner-thinking *p* but I am not author-thinking it” on the basis of introspection but whose grounds are undercut via testimony: S learns that their medical condition has hallucinatory effects. Hu also stipulates that S does not countenance the alleged possibility, conjectured by Lichtenberg, that there can be thoughts without thinkers. The basic idea is that by failing to

ascribe authorship of thoughts to themselves on an introspective basis, S is making a mistake in answering the question “Which thing is author-thinking p ?”: S says that it is not themselves, when in fact it is. However, S is still warranted in judging “Somebody is author-thinking p ”.

Palmira (2020) observes that S’s taking there to be no thinking without author-thinkers can be epistemically relevant in two ways: either that belief is part of S’s grounds, or it is a background presupposition of the target self-ascriptions. Palmira argues that both options fail to show that the introspection-based self-ascriptions are open to wh-EM. In the former case, it is no longer the case that S’s introspective experience *alone* warrants S to judge “Someone is author-thinking T”. In the latter case, the background presupposition that thoughts have authors ends up entailing, given the thought T, that someone is author-thinking T, which is precisely the content of the existential judgment “Someone is author-thinking T”. Given this, Palmira contends that the only way in which such a background presupposition can be warranted by introspective experiences is by assuming a *dogmatist* conception of warrant to the effect that the target experiences can immediately provide a *prima facie* warrant for the corresponding judgment (see Pryor 2000). On the competing *conservative* and *moderate* accounts of warrant, however, the background assumptions cannot be warranted by our experiences, for we are only either *entitled to trust* (Wright, 2004) or *mandated to assume* them by the very concept of epistemic rationality (Coliva, 2015). This suggests that Hu’s thought insertion-based counterexample is theoretically loaded. According to Palmira, the lack of theoretical neutrality of this variant of the thought insertion scenario is enough to discard its status as a counterexample to the thesis that introspection-based self-ascriptions of psychological properties are logically IEM.

In our view, the foregoing suggests that introspection-based self-ascriptions of psychological properties remain the best candidate class of judgments that exhibit logical IEM. However, we would like to stress that endorsing this epistemological thesis has neither metaphysical implications for the nature of the self as Evans feared, nor does it have the radical semantic repercussions that Wittgenstein and Anscombe pointed out. We will return to this below (Section 4.3, 5.1).

4.2. IEM beyond the first person

Several philosophers (Evans, 1982; Shoemaker, 1968; Wright, 1998, 2012) have variously argued that demonstrative judgments such as “That is black”, spatial and temporal judgments such as “It’s raining here” and “It’s raining now”, second- and third-person judgments such as “You are next to me” and “He’s a long way off”, when made on the appropriate grounds, are IEM. To give just one example, consider the demonstrative judgment “That is

black”. When made in ordinary circumstances on perceptual grounds, the judgment is IEM: it cannot be that something is in fact black but it is not that, i.e., x . Clearly, however, one might find oneself in the circumstance in which x is transparent and set against a black background. In such a scenario, the presupposition that x is the perceived black object is false, something which indicates that the judgment is only *de facto* IEM.

More recently, some authors (Coliva, 2017; Palmira, 2019; Wright, 2012) have ventured the hypothesis that some singular judgments about the natural numbers can be IEM. Consider the following: “4 is a square root of 16”. Suppose that S makes such a judgment on the basis of number theoretical facts, viz. facts about both negative and nonnegative integers. Plausibly, S ’s grounds for “4 is a square root of 16” should be specified as follows:

- (G1) A square root of a number n is a number m such that $m \times m = n$.
- (G2) All integers have a square root.
- (G3) For any two integers m and n : $m \times n = -m \times -n$.
- (G4) $-4 \times -4 = \Pi 4$.
- (G5) $-4 \times -4 = 4 \times 4$.

Now, since (G4) is an instance of b is F, and (G5) is an instance of $a = b$, we have that the pattern of judgment relevant to *de re* EM is met. And yet, since (G5) is necessarily and a priori true, the judgment “4 is a square root of 16” is *de re* IEM. The scope of IEM then extends well beyond the first person.

5. Contemporary accounts of IEM

We now turn to the question of how to explain the IEM status of a certain class of judgments. The foregoing discussion suggests that, for any given class of putative IEM judgments, an account of why that class is/is not IEM has to satisfy four *desiderata*:

- (i) to respect the basis-relative nature of the phenomenon, namely the fact that the same self-ascription can be EM or IEM depending on its grounds or background presuppositions;
- (ii) to the extent that one countenances a distinction between *de re* and *wh*-IEM, an account of IEM should specify how the two kinds of IEM differ;
- (iii) to capture the difference between *logical* and *de facto* IEM;
- (iv) to cover cases of IEM that go beyond the first person.

Our next task is to compare the three main contemporary accounts of IEM against these desiderata.

5.1. The simple account

The first explanation of IEM we focus on traces back to Evans (1982). As we saw earlier Evans's key observation is that a judgment is IEM because its grounds do not feature any identification component. This observation lies at the heart of the *Simple Account*. The account has been dubbed "simple" (García-Carpintero, 2018; Morgan, 2012; Palmira, 2019; Wright, 2012) in virtue of its deflationary character: focusing on the basic case of first-personal or *de se* thoughts, the Simple Account has it that we don't have to tell any special story about the content of experiences that constitute the grounds of these judgments (cf. the Content Accounts, see below), nor do we have to rely on any controversial thesis about self-knowledge (cf. the Metasemantic Account, see below) to explain IEM. Besides Evans, contemporary advocates of the Simple Account are Guillot (2014), Morgan (2012) and Wright (2012).

The Simple Account respects the basis-relative nature of IEM: a judgment such as "My legs are crossed" made by attending to one's proprioceptive experiences is IEM because its grounds do not feature any identity component, whereas the same judgment made upon looking at an image reflected in the mirror is not IEM because its grounds do feature such an identity component. So, the first desideratum is met. This said, however, the account does stand in need of revision to meet the other three desiderata.

As already noted by Pryor (1999, pp. 292–3) and further remarked by Guillot (2014, p. 13), the Simple Account does not explain wh-IEM. Bear in mind that what's (allegedly) distinctive of wh-EM is that, unlike *de re* EM, it does not involve the presence of an identity component in the judgment's grounds. So, we cannot explain wh-IEM by saying that no identity component features in the target judgment's grounds, for the absence of such an identity component is in fact compatible with the judgment's being vulnerable to wh-EM. Guillot (2014) argues that the Simple Account needs to be supplemented with a relativist story about the contents of the experiences that ground one's judgments (more on this in Section 1.2). Wright (2012, pp. 259–60), by contrast, claims that the Simple Account can be extended to explain why a certain class of judgment is wh-IEM. Wh-EM arises when the inferential structure of one's judgment is as follows: one judges "a is F" (e.g., "That is a skunk") on both: (i.) grounds G1 that warrant "Something is F" (e.g., one's olfactory experiences); and (ii.) grounds G2 that interact with other grounds G2 (e.g., one's visual experiences of several small black animals) one exploits to (mistakenly) single out which thing is F (e.g., one singles out a black cat instead). This revision does not affect the deflationary spirit of the proposal, for no explanation of why IEM judgments lack that inferential justificatory structure is given, and it enables the Simple Account to make sense of the IEM status of certain demonstrative, temporal, second-

and third-personal judgments: those judgments are unified by the fact that one's grounds are directly experiential or observational in such a way that they lack the problematic inferential justificatory structure. However, this is not enough to ensure that the Simple Account meets the third desideratum: insofar as the Simple Account offers no explanation of why the target judgments lack the problematic inferential justificatory structure, no IEM-relevant difference in such structure can be appealed to in order to distinguish logical and *de facto* IEM.

Drawing on Coliva (2006), Wright (2012, pp. 270–1) acknowledges that a proprioception-based self-ascription such as “My legs are crossed” is vulnerable to EM although its grounds do not exhibit any inferential structure, as witnessed by the conceivability of abnormal circumstances constructed along the lines of quasi-memories.²⁸ Bear in mind that proprioception-based self-ascriptions do rest on the identification “I = the person whose body is causing such-and-such proprioceptive experiences”: despite not being foregrounded in one's psychological basis for the judgment, such identification is a background presupposition of it. To make sense of this, Wright suggests modifying the Simple Account as follows: the IEM status of a judgment depends on both the non-inferential structure of its grounds and the absence of an identification in the judgment's background presuppositions. This commits the Simple Account to regarding proprioception-based self-ascriptions – as well as all the other self-ascriptions for which abnormal circumstances can be conceived of – as vulnerable to EM tout court. However, as has emerged previously (Section 2), the possibility of abnormal cases can also be taken to show that the target judgments are only *de facto* IEM, for they rely on background presuppositions that are only contingently true (see Coliva, 2006). So, the Simple Account does not have the resources to capture the distinction between *logical* and *de facto* IEM (see García-Carpintero, 2018). Wright (2012, pp. 271–272) explicitly acknowledges that the Simple Account considerably restricts the extent of first-personal IEM. A more promising reply, we believe, is to modify the Simple Account and say that while the logical IEM status of a judgment depends on both the absence of an inference in its grounds and the absence of an identification in the judgment's background presuppositions, the *de facto* IEM status of a judgment depends on the absence of an inference in its grounds only.

5.2. Content accounts

On the standard theory of propositional attitudes, different subjects can think the same thought (call this *shareability*), and a thought's truth-value is determined with respect to what the world is like (call this *absolutism*). Inspired by Frege (1918-9/1956) and Castañeda (1966, 1968), John Perry (1979) and David Lewis (1979) have argued that *de se* thought requires a revision of the standard theory of propositional attitudes. On Perry's

absolutist view, to have *de se* attitudes is to have attitudes toward Russellian propositions, namely propositions that ascribe properties to a certain object, i.e., oneself, which is presented in a distinctively first-personal and non-shareable way (see also Evans, 1982; García-Carpintero, 2018), in the sense that different subjects cannot entertain the same mental representations with the same content. What's responsible for the non-shareability of first-person thought? On what we may call the *Reference-fixing* approach, the first-person mode of presentation is what fixes the pattern of reference of first-person thought but is not itself part of its truth-conditional content, which has oneself as one of its constituents (see Campbell & Sugden, 1999; Evans, 1982; García-Carpintero, 2018; O'Brien, 2007; Peacocke, 2014; Perry, 1979). On what we may call the *Reflexivist* approach, the first-person mode of presentation associated with the mental "I" is part of the content of the *de se* attitudes (see Frege, 1918 –9; Higginbotham, 2003).

On Lewis's *relativist* view, in contrast, to have *de se* attitudes is to have attitudes toward relativized propositions, conceived by Lewis as properties that different subjects can self-ascribe, where the truth of such propositions is evaluated with respect to the thinking subject (see Chisholm, 1981; Loar, 1976; Ninan, 2010; Recanati, 2007).

The Reflexivist view and the Relativist view have inspired two accounts of IEM whose unifying idea is that IEM is explained by the distinctive *content* of the *de se* states. Let us take them in turn.

The *Reflexivist Content Account* of IEM, defended by Higginbotham (2003, 2010), starts from the contention that the content of *de se* attitudes is reflexive: whenever I think a first-person thought such as "I am hungry" I have an attitude toward the proposition that the subject of that very thought is hungry. So, the mental "I" contributes a token-reflexive description, i.e. *the subject of this very thought t*, to the truth-conditions of *de se* attitudes, so, having a *de se* attitude just is having an attitude that involves thinking about oneself as the subject of that attitude, something which is meant to guarantee that one cannot misidentify who the subject of the attitude is. As remarked by Morgan (2012), however, this proposal fails to respect the basis-relative nature of IEM, namely the idea that the same *de se* judgment is/is not IEM depending on the grounds one exploits while making that judgment. Suppose that I judge "I hear trumpets" by inferring it from "The person who is listening to music hears trumpets" and "I am the person who is listening to music". If the Reflexivist Account were correct, from the fact that my judgment "I hear trumpets" involves thinking of myself as the subject of that judgment we would have to conclude, implausibly, that the judgment is IEM.

The Relativist Content Account, defended by Recanati (2007, 2012), bypasses these worries by locating itself at the right explanatory level: what explains the fact that certain self-ascriptions are IEM is, ultimately, the selfless content of the

states that constitute one's grounds for the relevant self-ascriptions. This ensures that the Relativist Content Account, unlike the Reflexivist Content Account, has the resources to account for the basis-relativity of IEM and meet the first explanatory desideratum presented above. The Relativist Content Account has two key features: first, it promises to give us an explanation of why self-ascriptions that are *de re*-IEM are not based on any identification component, something that the Simple Account is unable to do due to its deflationary character (see Wright, 2012). The Relativist Content Account is then able to claim that the absence of an identification component in the judgment's grounds is explained by the distinctive content of the states that constitute such grounds. This is what makes the Relativist Content Account *inflationary*. Secondly, as noted by Wright (2012), the Relativist Content Account seems to accord well with Hume's no-subject thesis to the effect that the self typically is not represented in one's experiences: when I see a blue tree, the content of my experience plausibly only involves the blue tree and not also myself doing the seeing.

The Relativist Content Account provides a straightforward explanation of the IEM status of a class of self-ascriptions that Recanati calls "implicit", that is, those *de se* judgments that have a selfless content (see Recanati, 2007, 2012): if I judge "My legs are crossed" on proprioceptive grounds, my proprioceptive experience has a certain property encoding a subjectless bodily condition, e.g., *having legs crossed* or *legs being crossed*, as a content, whereas the proprioceptive *mode* of the experience – where modes are psychological analogues of illocutionary forces – supplies the self as the sole point of evaluation that makes the content of the judgment truth-evaluable. Thus, the proprioceptive mode ensures that my proprioceptive experiences will warrant judgments about one object only, i.e., myself. The same pattern of explanation applies to implicit *de se* judgments based on introspection, memory, and visual perception. This account extends to cases of implicit *de se* judgments that are wh-IEM. Wh-EM takes place when one's grounds suffice to warrant the existential judgment "Something is F" without thereby warranting one's singular judgment "a is F". Insofar as implicit *de se* judgments are such that one does not in fact predicate something of a particular object, i.e., oneself, this *a fortiori* leaves no room for wrongly identifying which thing is F (See Guillot, 2014, who further argues that the Relativist Content Account can be regarded as a development of the Simple Account).

The Relativist Content Account faces three main problems though. First, as has emerged previously, it's controversial to maintain that the content of one's proprioceptive experiences is subjectless. So, the Relativist Content Account does rest on a controversial picture of the content of experiences.

The second problem concerns the alleged IEM status of "explicit" *de se* judgments, namely *de se* judgments whereby one makes the contrast

between oneself and another person relevant. A judgment such as “My legs are crossed (in contrast to my neighbor’s)” plausibly features the first-person concept, and is therefore explicitly *de se* (Recanati, 2012, p. 196). Recanati acknowledges that, when made on proprioceptive grounds, such a judgment is IEM. To explain this, Recanati (2012, p. 192) maintains that we make explicit *de se* judgments as a result of a process called *reflection*, which is a non-inferential way of making explicit at the level of the judgment what was implicitly supplied by the mode of the relevant selfless experience. Even granting the existence of reflection, however, the Relativist Content Account’s general explanatory claim appears to be mistaken (see Wright, 2012, p. 272): the proprioceptive mode of my current experience as of legs being crossed can supply myself as the object that the experience is concerned with only if the background presupposition *that the person whose body is causing such and such proprioceptive experiences is myself* is not challenged. Now, when such a presupposition gets challenged in abnormal cases, this shows that the target class of *de se* judgments are only *de facto* IEM although the judgment’s grounds in both normal and abnormal cases is the same. So, the Relativist Content Account fails to capture the distinction between *logical* and *de facto* IEM.²⁹

Finally, it’s hard to see how the Relativist Content Account can be extended to demonstrative, second- and third-person judgments (not to mention arithmetical judgments): if I judge “You look lovely today” on ordinary perceptual grounds, my judgment is IEM relative to the second-person concept, but my experience clearly has you as part of its content (see García-Carpintero, 2018; Wright, 2012). So, the Relativist Content Account does not seem to generalize to all classes of judgment that can be IEM.³⁰

5.3. *The metasemantic account*

The key contention of the Metasemantic Account is that the facts that establish the pattern of reference of a certain singular concept “a” are also the facts that explain why the judgment “a is F” is IEM. Versions of the Metasemantic Account have been defended by García-Carpintero (2018), Howell (2006), Palmira (2019), Palmira (2020), Peacocke (2014) and Verdejo (2021).

Let us begin with the widely held view that the reference of the first-person concept is fixed according to the following token-reflexive rule (Campbell, 1994; Coliva, 2003, 2012, 2017; García-Carpintero, 2018; O’Brien, 2007; Peacocke, 2014, but see; Evans, 1982 for criticism and Morgan, 2015 for an alternative Evansian model):

Token-Reflexive Rule:

“I” =_{ref} the thinker of this thought

According to the token-reflexive rule, any token first-person thought is about the person whose token it is, namely the person who produces the token. Importantly, the token-reflexive rule should be regarded as type-individuating the first-person concept by defining its sense, and not as contributing an identity to the truth-conditions of one’s thoughts featuring “I” (see Coliva, 2012, 2017; García-Carpintero, 2018; Palmira, 2020). As noted by García-Carpintero (2018), Guillot (2016), and Palmira (2020), the token-reflexive rule features a demonstrative expression, i.e., “this”. Thus, the rule delivers a referent for “I” only if this determinable aspect of the rule is fixed. Moreover, Palmira (2020, 2022) emphasizes that, in light of the metaphysical possibility of disowned first-person thoughts countenanced by Campbell and others, the very notion of a “thinker” of a thought should be disambiguated: does the rule feature the notion of a thinker-qua-author or of a thinker-qua-owner of a thought?

García-Carpintero (2018) and Palmira (2022) maintain that our ability to have introspective knowledge of the phenomenal features of our occurrent thoughts ensures that we can mentally latch onto the thought we are presently thinking. This reveals that the notion of a thinker of a thought featuring the token-reflexive rule must be specified in the ownership sense, for to be a thinker-owner of a thought is to be introspectively aware of its phenomenal character. Importantly, García-Carpintero (2018) and Palmira (2022) accept the new *acquaintance view* of introspection to the effect that an introspective state targeting the phenomenal character of an occurrent thought is partly constituted by the target phenomenal character (see Gertler, 2012; Giustina, 2021 for a defense of the acquaintance view, and Chalmers, 2003; Horgan & Kriegel, 2007 for other defenses of the constitution thesis).³¹ The Metasemantic Account gives an *inflationary* explanation of IEM, as it will have to rely on substantive (and controversial) theses about one’s awareness of one’s thoughts.

With these points in play, the following explanation of the IEM-status of introspection-based self-ascriptions of psychological properties becomes available: insofar as S’s introspective experiences as of a thought T passing through S’s mind warrant S to judge “Someone is thinking a thought T”, such introspective experiences can’t but warrant S to judge “I am thinking T”. This epistemic dependence is guaranteed to hold in virtue of the fact that S’s introspective experiences enable the thinker to latch onto the thought they are presently thinking and think of themselves first-personally – as per the token-reflexive rule – without any possibility of mistake about the concept in the subject position, for those experiences are partly constituted by the thought’s phenomenal character (See García-Carpintero, 2018; Palmira,

2020). Focusing on introspection-based self-ascriptions of inserted thoughts, Palmira (2020) maintains that insofar as we acknowledge that schizophrenic patients are endowed with the first-person concept, those self-ascriptions will be IEM. Their falsity, however, is explained by the fact that the patients make a predication mistake: patients are mistaken in not ascribing the property of being the author of the introspected thoughts to themselves.³²

Peacocke (2014: Chapter 5: Section 1, Chapter 8: Section 2) offers a different version of the Metasemantic Account. The key contention is that the concepts picking out psychological properties are such that their instantiations warrant self-ascriptions made on the relevant grounds in virtue of the fact that the psychological concepts' possession-conditions make reference to the thinker's own possession of the property. Importantly, the Metasemantic Account would explain why, even conceding the possibility of first-person disowned thoughts, introspection-based self-ascriptions of psychological properties are IEM: since the deluded subject has introspective experiences as of a thought with certain phenomenal properties passing through their mind, this ensures that the subject cannot but ascribe the thought to themselves in light of the fact that those experiences fix the reference of "I" (see Palmira, 2020).

The Metasemantic Account is equipped to capture the basis-relativity of IEM, for it is only those facts that contribute to fixing the reference of the first-person concept that give rise to IEM self-ascriptions. Yet, in order to capture the IEM status of proprioception- and memory-based self-ascriptions, the token-reflexive rule should be modified so as to type-individuate the first-person concept via a number of different contextual relations one bears to oneself, namely those relations that involve one's proprioceptive or memory experiences. This, however, raises an immediate challenge (see Verdejo, 2021): why is it that introspective, proprioceptive and memory experiences contribute to fixing the reference of "I" whereas other types of experiences, e.g., the experience one has when one sees oneself in a mirror, do not?

We believe that the supporter of the Metasemantic Account can acknowledge that introspective, proprioceptive, and memory experiences all contribute to fixing the reference of "I" because they indeed are ways of gaining information about oneself that ensure *de facto* IEM. Yet, introspection has the pride of place precisely because it is the only source of information about oneself which ensures logical IEM.

Turning to the *logical/de facto* distinction, García-Carpintero (2018) suggests that the failure of the judgment's background presuppositions in abnormal circumstances leads to the adoption of more guarded presuppositions of identity. To give an example: the self-ascription "My legs are crossed" made in a case of quasi-proprioception is such that the reference of "my" is fixed relative to the presupposition that

my body *appears* to be the body with such and such arrangement of limbs that caused the relevant proprioceptive experiences. In such cases, the proprioceptive experiences would warrant the judgment “Someone’s legs are crossed” without warranting the judgment “My legs are crossed” precisely because the appearance of identity between my body and the body that is the causal source of those experiences can be mistaken. This explains why proprioception-based self-ascriptions are only *de facto* IEM. So, the Metasemantic Account has the resources to respect the logical/*de facto* distinction.

The Metasemantic Account has been extended to various cases of IEM. Peacocke (2014, pp. 86–99, 106–113) gives an account of the IEM status of *de se* judgments such as “I see a hummingbird in front of me” based on one’s visual experiences and of agentive *de se* judgments such as “I am raising my arms”; García-Carpintero (2018) extends the Metasemantic Account to demonstrative judgments and Palmira (2019) to arithmetic judgments. To give just an example, consider the demonstrative judgment “That keyboard is black”. When made in ordinary circumstances on perceptual grounds, such a judgment is IEM: it cannot be that something is in fact black but it is not that, i.e., x . This is so since one’s perceptual experiences as of a black keyboard contribute to fixing the reference of “that” by determining what the demonstrated entity in the context is, i.e., x is the perceptually salient black keyboard. Clearly, however, one might find oneself in the circumstance in which the keyboard is transparent and set against a black background. In such a scenario, the presupposition that x is the perceived black keyboard is called into question and makes us retreat to the presupposition of identity that x *appears* to be the object that is the causal source of one’s experiences as of a black keyboard. In such a scenario, one’s perceptual experiences would warrant the judgment “Something is black” without ipso facto warranting the judgment “That keyboard is black” precisely because the appearance of identity that gets presupposed can be mistaken. This explains why perception-based demonstrative judgments are only *de facto* IEM.

The previous discussion shows that the Metasemantic Account compares favorably with its competitors. Importantly, the Metasemantic Account helps us reconcile the idea – of Cartesian descent – that introspection-based self-ascriptions of psychological properties are epistemologically special while, at the same time, maintaining that the first-person concept genuinely refers to an object, i.e., the self, endowed with both psychological and physical properties. And there is more: the Metasemantic Account will help us settle two important questions about the wider philosophical significance of IEM.

6. Significance

6.1. Illusions of transcendence

As we saw at the beginning of this paper, IEM has given rise to the “illusion of transcendence” regarding the self. Wittgenstein and Anscombe argued that only a Cartesian ego – or “a stretch” of one – could account for the impossibility of a misidentification when (psychological) self-ascriptions are at stake. To avoid postulating such an entity, they felt compelled to deny that “I” is a referring expression at all, and even that there is such thing as a self, picked out by “I”. As we further saw above, however, this is an overreaction, due to the failure to appreciate that IEM could be due to the fact that (at least some) self-ascriptions are not based on any identification component at all. In such cases, much like demonstrative- and here- and now-thoughts, “I” would still refer to oneself, and this would be compatible with the self being identical to a living human being with both physical and psychological properties.

According to Coliva (2012, 2017), the distinction between *de facto* and logical IEM could give rise to further illusions of transcendence. For it would turn out that only non-inferential psychological self-ascriptions are logically IEM. This, in turn, could give rise to the idea that the self is most basically presented to oneself as the thinker of a given thought and this might be taken to show that it is identical to a Cartesian ego (or to a “stretch” of one). Yet, this would be a non-sequitur. For from the fact that one is presented to oneself as a thinker of a thought it does not follow that only a mental entity can have introspectively available thoughts. That is, being presented to oneself as the thinker of a given thought is entirely compatible with the fact that one is a physical entity capable of entertaining thoughts and of being aware of oneself while doing so. According to Coliva (2012) while IEM and logical IEM can account for the illusion of transcendence,³³ they are in fact compatible with an animalist conception of the self once these phenomena are correctly characterized. Contrary to Evans and McDowell, then, it is not necessary to deny such a distinction to maintain the view that the self is identical to a living human being with both physical and psychological properties. Still, the distinction grants Cartesianism a point – that is, that at the most fundamental level, the sense of “I” (its *Sinn*) is given by the token reflexive rule, in line with the Metasemantic Account of IEM.

Léa Salje (2020) agrees with Peacocke (1999) and Coliva (2012) that illusions of transcendence arise from the peculiar epistemology of first-person thought, but disagrees with Coliva about which epistemic features are responsible for them. In her view, such an illusion arises as a way of mitigating the cognitive dissonance created by the combination of two individually compelling features of first-person thought she defines as follows (Salje, 2020, pp. 740, 754):

Special Insight: In virtue of being the thinker of a conscious I-thought, a subject has privileged noninferential epistemic grounds for first-personal knowledge that the referent of their thought exists.

Ordinary Ignorance: A subject does not, in virtue of being the thinker of a conscious I-thought, have privileged noninferential grounds for knowledge about the nontrivial properties of the referent of their thought.

Yet, if one accepts the Metasemantic Account of IEM, *Ordinary Ignorance* appears to be false. Bear in mind that García-Carpintero's (2018) and Palmira's (2020) versions of that account, the first-person indexical gets its reference fixed thanks to the fact that the subject, upon attending to the thought they are thinking, is introspectively aware of the phenomenal mark that such a thought has for them. So, in virtue of being the thinker of a first-person thought, the subject is in a position to know some of the nontrivial properties of the referent of the first-person thought: I am in a position to know what it is like for the object of my first-person thought, i.e., myself, to undergo the first-person thought I am presently thinking. On various accounts of introspection that make room for the constitution thesis introduced above (Section 1.3), namely the idea that one's introspective experiences of the phenomenal character of an occurrent thought are partly constituted by the target phenomenal character, this kind of knowledge is privileged and non-inferential. Thus, the Metasemantic Account of IEM comes into conflict with Ordinary Ignorance and, as a consequence, with Salje's explanation of illusions of transcendence about the self.

Interestingly, Salje (2020) argues that Ordinary Ignorance is a plausible principle since *de se* thought is singular thought, and Ordinary Ignorance does seem to hold for non-*de se* singular thought. For instance, from the fact that you're thinking of a plant in front of you "That is nice", you cannot know whether the object you're thinking about is a fake plant or a real one. However, supporters of the Metasemantic Account will press the point that *de se* thoughts show that Ordinary Ignorance doesn't hold unrestrictedly for all varieties of singular thought. This brings us to the topic of the alleged distinctiveness of the first-person perspective, to which we turn in the next section.

6.2. *The distinctiveness of the first-person perspective*

As has already emerged previously (Section 1.2), it has been customary in contemporary philosophy of mind to maintain that first-person thought requires a revision of the standard theory of propositional attitudes. In recent work, however, Cappelen and Dever (2013) and Magidor (2015) have challenged the orthodoxy, defending a form of skepticism about the distinctiveness of the first-person perspective.

Although *de se* skeptics mainly focus on action explanation, Cappelen and Dever (2013: Chapter 7) also consider IEM as a motivation for the adoption of nonstandard views of first-person thought. Cappelen and Dever (2013: Chapter 7) undermine this motivation on three grounds: first, IEM is philosophically uninteresting since we can challenge the IEM status of any type of self-ascription, including introspection-based self-ascriptions of psychological properties; secondly, even if IEM is philosophically interesting, it's not restricted to the first-personal or, more generally, to the indexical case; thirdly, the correct explanation of IEM does not have to rely on the allegedly special nature of first-person thought.

We have argued earlier (Section 3), that the *logical* IEM status of introspection-based self-ascriptions of psychological properties is not threatened by the possibility of first-person disowned thoughts. But Cappelen and Dever (2013, p. 131, fn. 4) go as far as saying self-ascriptions such as “I seem to be thinking that the grass is green” are only *de facto* IEM: they maintain that it's possible to imagine that we could be introspectively wired to someone else's phenomenal impressions. As remarked by García-Carpintero (2018: fn. 13), however, this possibility seems to presuppose a controversial perceptual model of introspection and is at odds with various accounts of self-knowledge offered by Coliva (2016), Gertler (2012), Peacocke (1999: Chapter 6) and Shoemaker (1994). This said, it still remains to be seen whether *logical* IEM reaches further than the basic case of introspection-based self-ascriptions of psychological properties. Palmira (2019) suggests that it does, for even some arithmetical judgments are *logically* IEM.³⁴

Finally, Cappelen and Dever (2013, pp. 136–9) argue that the Simple Account and the Content Accounts do not require any revision of the standard theory of propositional attitudes. The Simple Account, given its deflationary nature, does indeed afford neutrality on the nature of first-person thought. Cappelen and Dever's (2013, p. 138) rejection of Higginbotham's Reflexivist Content Account also appeals to the basis-insensitivity of the account we have already discussed above (Section 1.2). As for the Lewisian-inspired relativist account offered by Recanati, Cappelen and Dever rely on an argument (2013: Chapter 5) which underplays the role of relativized contents in a general theory of self-ascription.

Cappelen and Dever do not discuss the Metasemantic Account. That account, as is developed in García-Carpintero (2018) and Palmira (2020), rests on the idea that in order for one to possess the first-person concept one has to think of oneself under an introspective mode of presentation that is not available to anyone else. Thus, if – as we argued in Section 4.3 – the Metasemantic Account offers the overall best explanation of IEM, we have good abductive reasons to depart from the standard theory of propositional attitudes.

7. Conclusions

Since modern philosophy, first-person thought has been regarded as somewhat special. While Descartes located its specialness in the alleged infallibility of self-ascriptions of psychological properties, Wittgenstein's remarks in the *Blue Book* open up the possibility of departing from the somewhat radical Cartesian infallibility thesis while, at the same time, retaining the insight that a certain class of first-person thoughts is indeed special, on account of the alleged impossibility of mistakenly identifying whether it is I who has a certain psychological property. Anscombe brings Wittgenstein's distinction to bear on the question of the nature of the first-person pronoun. Anscombe's point is that if the target self-ascriptions were IEM this would be explainable only by positing a Cartesian ego as the referent of "I", something which leads Anscombe to embrace the radical view that all uses of "I" aren't referential.

Acceptance or rejection of a Cartesian-inspired picture of the mind operates in the background of the Shoemaker-Evans dispute too. Shoemaker makes a Cartesian-sounding point when he argues that even if memory-based self-ascriptions of physical properties and introspection-based self-ascriptions of psychological properties are both IEM in normal circumstances, the former are vulnerable to EM in abnormal circumstances whereas the latter aren't. Evans, by contrast, takes both classes of self-ascriptions to involve a justificatory structure that doesn't feature any identification component, thereby contending that they deserve the same IEM status.

The Shoemaker-Evans dispute led to a clarification of the various ways in which a judgment can involve an error through misidentification as well as of the modal force involved in the notion of immunity. All varieties of error through misidentification are such that while one is not warranted to judge "a is F", one is warranted to judge "Someone is F". This justificatory gap can manifest itself in three ways: the judgment's grounds feature a mistaken identification component (de re-EM); the judgment's grounds do not warrant any judgment about which particular thing is F (wh-EM); the judgment rests on a background presupposition featuring a mistaken identification component (background presupposition-EM). The notion of immunity can be interpreted in two ways: a judgment can be IEM relative to the actual world but EM relative to a possible world in which certain abnormal circumstances occur (de facto IEM), or else it can be IEM relative to any possible world (logical IEM). Acknowledgment or discussion of such distinctions gave rise to a number of different views about which classes of self-ascriptions enjoy one, more than one, all or none of the varieties of (*de facto* or logical) IEM.

Once a certain view about the nature and the scope of IEM is reached, the next task on the IEM theorist's agenda is to explain why a certain class of judgments is IEM in the targeted sense. We have reviewed three explanations of IEM: the Simple Account, the Content Account, and the Metasemantic Account. The Simple Account, championed by Wright (2012), takes the IEM status of judgments to depend on both the absence of a certain inferential structure in their grounds and the absence of an identification component in their background presuppositions. The deflationary spirit of the Simple Account contrasts with more explanatorily ambitious proposals which seek to answer the question of why IEM judgments lack the target inferential structure. The Content Account, developed by Recanati (2007, 2012), offers such an explanation in terms alleged selfless nature of the experiences that constitute the grounds of the relevant self-ascriptions. By contrast, the Metasemantic Account, advocated by García-Carpintero (2018), Palmira (2019, 2020) and Peacocke (2014), focuses instead on the role played by the judgment's grounds in fixing the reference of the first-person concept. We have highlighted the respective pros and cons of these views and concluded that the Metasemantic Account scores higher than its competitors on overall assessment theory.

In closing, we focused on the implications of IEM for debates about the metaphysics and the representation of the self. While IEM doesn't force us, *pace* Anscombe and Wittgenstein, to postulate the existence of transcendent objects of reference for "I", IEM can certainly be seen as one of the sources of such illusions of transcendence. Finally, we stressed that even if we acknowledge that IEM is a phenomenon that extends beyond the first person, this is not enough to show, as *de se* skeptics such as Cappelen and Dever would have it, that IEM can be explained without revising the standard theory of propositional attitudes.

Notes

1. We will use "IEM" both as a label for the phenomenon and as an adjective to characterize judgments that are immune to error through misidentification.
2. For a different interpretation of Wittgenstein's position, see Wiseman (2019).
3. "Introspection" here does not presuppose any specific account of self-knowledge. For a discussion of self-knowledge, see Coliva (2016).
4. See e.Lichtenberg (1971: Section 76).
5. What's under discussion here is what the sense of "I" should be like in order for one to know that one is referring to oneself while tokening the first-person pronoun. Coliva (2002a, 2017) has suggested to take Anscombe's proposal regarding the sense of "I" as bearing onto the nature of the first-person concept, rather than on IEM. For *all* uses of "I" would be such as to exhibit what Coliva terms "the real guarantee" – i.e., the guarantee that, by using "I", (i) one would necessarily *refer* to oneself, and (ii) one would necessarily *know* that one is referring to oneself. By contrast, only some uses of

- “I” would be IEM. More on this in fn. 14. For a different interpretation of Anscombe’s position, see Wiseman (2017).
6. See Section 4–4.3 for various ways of cashing out the sense of “I”, which would not fall prey to the circularity objection raised by Anscombe.
 7. This interpretation of Anscombe is widely shared: see e.Evans (1982), Peacocke (2008) and Wiseman (2017). Stainton (2019) offers a different interpretation to the effect that Anscombe’s main claim is that *if* we have a certain Fregean conception of reference, then “I” is not a referring expression.
 8. Shoemaker will deny this claim in his subsequent work. More on this below.
 9. Shoemaker considers EM and IEM as affecting the use of linguistic terms in statements. Nowadays, it is customary to drop such a qualification, and to take EM and IEM primarily as affecting judgments.
 10. The use of the asterisk is customary to indicate that “oneself” is here the indirect reflexive pronoun. That is, the pronoun that replaces “I” in indirect speech.
 11. Shoemaker envisages cases in which fissions of subjects may occur, such that while sharing a single past, the two resulting entities go on living separate lives and build different memories from then on. He also considers cases of partial brain transplant which would make memories deriving from the donor’s past available to the recipient.
 12. Famously, Evans (1982, pp. 224–235) proposed a transparency account of self-ascriptions of beliefs and perceptions. In these cases, the self-ascription would be based neither on the recognition of a subject, nor on its individuation through its mental properties, but would be merely a fall-out of the transparency method employed to make the self-ascription. Furthermore, due to the Generality Constraint, the “I” in it would depend on thinking of oneself as an element in the objective order. Hence, Evans’ views about psychological self-ascriptions aren’t in the least hospitable to a Cartesian conception of the self.
 13. As we saw, Shoemaker also conceived of q-memories as “from the inside” and therefore as not identity-neutral.
 14. In the case of *de re* EM, one is warranted to judge “Something is F” since one’s grounds warrant “b is F” from which the existential judgment can be inferred.
 15. An undercutting defeater is evidence that undermines the support for a belief without giving support to its negation. By contrast, an overriding defeater is evidence that supports the opposite belief (or a belief incompatible with the original one).
 16. McGlynn (2021) offers a variant of Pryor’s definition whose main idea is that one’s grounds for the existential judgment “Something is F” also warrant one in identifying a particular object *a* as being F but, appearances notwithstanding, those grounds are not caused by the fact that *a* is indeed F. This definition dispenses with the notion of undercutting defeaters.
 17. Morgan motivates his proposal by appealing to the thesis that we cannot obtain knowledge from false premises: if one’s grounds warranted “Something is F” via warranting “a is F”, the fact that “a is F” is false *ipso facto* makes “Something is F” fall short of knowledge since the existential judgment can be inferred from the singular judgment. McGlynn (2021) notes that this thesis is controversial in the ongoing epistemological debate.
 18. For the details of such a distinction, see Coliva (2006, p. 413).
 19. For a different take on Wright’s example, see McGlynn (2021, pp. 2302–2303).
 20. Coliva (2017, p. 247) conjectures that the real guarantee (see fn. 2) and wh-IEM might be very close kins. She rejects their identity, however, claiming (*ibid.*) 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. Palmira (2019) claims that this point is compatible with the idea 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 wh-IEM self-ascriptions.

21. One believes that p safely just in case one couldn’t easily have believed p falsely on the same or a relevantly similar basis. See Pritchard (2005) and Williamson (2000) for more on this notion.
22. McDowell (1997) and Cappelen and Dever (2013) are among the few theorists who still follow Evans in denying de facto IEM.
23. Campbell and Sugden (1999, 2003).
24. Campbell (2003, p. 36).
25. *Ibid.*, p. 35.
26. *Ibid.*
27. Notice, again, that this doesn’t mean that logical IEM is the same as infallibility: one can still make a predicative mistake about what thought one is thinking (is it a belief or a supposition? Is the content of the thought p or q ?) or what feeling one is having (is it a pain or a itch?).
28. In a q-memory case, one has an apparent memory of oneself being F at a certain time but that apparent memory contains information deriving from the perception of that event by someone who is not necessarily oneself.
29. Recanati (2007, pp. 149–154) takes logical IEM to be a primitive property of self-ascriptions of conscious states only. Recanati doesn’t think that the same notion of immunity can receive different modal interpretations, but he rather regards logical IEM and de facto IEM as altogether different types of phenomena. See Guillot (2013, pp. 1805–1808) for further discussion.
30. Recanati (2012) attempts to extend his account to demonstrative thoughts, but he still cannot account for the full scope of IEM.
31. Coliva (2002b, 2012) and Shoemaker (1968) make a related phenomenological observation to the effect that the very fact of being aware of thinking a certain thought makes that thought one’s own, something which would also determine the content of the token-reflexive rule in the ownership sense.
32. This also shows that the Metasemantic Account preserves the difference IEM and infallibility.
33. According to Coliva, there may be other sources of the illusion of transcendence, besides IEM. For a discussion, see Coliva (2012, 2017).
34. Coliva (2017) defends the contention that certain basic arithmetical concepts are indexical on the grounds that they exhibit the real guarantee which, as we have seen above, can be regarded as close kin of Pryor’s wh-IEM.

Acknowledgements

We are grateful to two anonymous reviewers and to the Editor for their comments on a previous draft. Palmira’s 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.

Disclosure statement

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

References

- Anscombe, G. E. M. (1975). The first person. In S. Guttenplan (Ed.), *Mind and language* (pp. 45–65). Clarendon Press.
- Bar-On, D. (2004). *Speaking my mind. expression and self-knowledge*. OUP.
- Campbell, J. (1994). *Past, space, and self*. Clarendon.
- Campbell, J. (2003). The ownership of thoughts. *Philosophy, Psychiatry, & Psychology*, 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*, 82(4), 609–625. <https://doi.org/10.5840/monist199982426>
- Cappelen, H., & Dever, J. (2013). *The Inessential Indexical*. OUP.
- Castañeda, H. (1966). ‘He’: A study in the logic of self-consciousness. *Ratio*, 8, 130–157.
- Castañeda, H. (1968). On the logic of attributions of self-knowledge to others. *The Journal of Philosophy*, 65(15), 439–456. <https://doi.org/10.2307/2024296>
- Chalmers, D. (2003). The content and epistemology of phenomenal belief. In Q. Smith & A. Jokic (Eds.), *Consciousness: New philosophical perspectives* (pp. 220–272). OUP.
- Chisholm, R. (1981). *The first person*. University of Minnesota Press.
- Coliva, A. (2002a). On What there really is to our notion of ownership of a thought. *Philosophy, Psychology and Psychiatry*, 9(1), 41–46. <https://doi.org/10.1353/ppp.2003.0005>
- Coliva, A. (2002b). Thought insertion and immunity to error through misidentification. *Philosophy, Psychology and Psychiatry*, 9(1), 27–34. <https://doi.org/10.1353/ppp.2003.0004>
- Coliva, A. (2003). The first person: Error through misidentification, the split between Speaker’s and semantic reference, and the real guarantee. *The Journal of Philosophy*, 100(8), 416–431.
- Coliva, A. (2006). Error through misidentification: Some varieties. *Journal of Philosophy*, 103(8), 403–425. <https://doi.org/10.5840/jphil2006103824>
- Coliva, A. (2012). Which ‘key to all mythologies’ about the self? a note on where the illusions of transcendence come from and how to resist them. In S. Prosser & F. Recanati (Eds.), *Immunity to error through misidentification. New essays* (pp. 22–45). Cambridge University Press.
- Coliva, A. (2015). *Extended rationality: A hinge epistemology*. Palgrave.
- Coliva, A. (2016). *The varieties of self-knowledge*. Palgrave.
- Coliva, A. (2017). Stopping points: ‘I’, immunity, and the real guarantee. *Inquiry*, 60(3), 233–252. <https://doi.org/10.1080/0020174X.2017.1261999>
- Evans, G. (1982). *The varieties of reference*. Clarendon.
- Frege, G. (1918–9) “The thought a logical inquiry. *Mind*, 65(1956), 289–311. <https://doi.org/10.1093/mind/65.1.289>
- Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia*. Psychology Press.
- 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>
- Gertler, B. (2012). Renewed acquaintance. In D. Smithies & D. Stoljar (Eds.), *Introspection and consciousness* (pp. 93–127). Oxford University Press.

- Giustina, A. (2021). Introspection without Judgment. *Erkenntnis*, 80(2), 407–427. <https://doi.org/10.1007/s10670-019-00111-8>
- Guillot, M. (2013). The limits of selflessness: Semantic relativism and the epistemology of *De Se* thoughts. *Synthese*, 190(10), 1793–1816. <https://doi.org/10.1007/s11229-013-0262-8>
- Guillot, M. (2014). Identification-free at last. Semantic relativism, evans's legacy and a unified approach to immunity to error through misidentification. *Teorema*, 33(3), 7–30.
- Guillot, M. (2016). Thinking of oneself as the thinker: The concept of self and the phenomenology of intellection. *Philosophical Explorations*, 19(2), 138–160. <https://doi.org/10.1080/13869795.2016.1176232>
- Higginbotham, J. (2003). Remembering, imagining, and the first person. In A. Barber (Ed.), *Epistemology of language* (pp. 496–533). Oxford University Press.
- Higginbotham, J. (2010). On words and thoughts about oneself. In I. Stojanovic, F. Recanati, & N. Villanueva (Eds.), *Context-dependence, perspective, and relativity* (pp. 253–282). Mouton de Gruyter.
- Horgan, T., & Kriegel, U. (2007). Phenomenal epistemology: what is consciousness that we may know it so well? *Philosophical Issues*, 17(1), 123–144. <https://doi.org/10.1111/j.1533-6077.2007.00126.x>
- Howell, R. (2006). Self-knowledge and self-reference. *Philosophy and Phenomenological Research*, 72(1), 44–69. <https://doi.org/10.1111/j.1933-1592.2006.tb00490.x>
- 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>
- Lewis, D. (1979). Attitudes *De Dicto* and *De Se*. *Philosophical Review*, 88: 513–43. also in D. Lewis, *Philosophical Papers*, Vol. 1, Oxford: OUP.
- Lichtenberg, G. (1971). In W. Promises (Ed.), *Schriften und briefe* (Vol. II). Munich: Carl Hanser.
- Loar, B. (1976). The semantics of singular terms. *Philosophical Studies*, 30(6), 353–377. <https://doi.org/10.1007/BF00372537>
- Magidor, O. (2015). 'The myth of the *de se*'. *Philosophical Perspectives*, 29, 249–283.
- McDowell, J. (1997). Reductionism and the First Person. In Jonathan Dancy (Ed.), *Reading parfit* (pp. 230–250). Malden, MA: Blackwell.
- McGlynn, A. (2016). Immunity to error through misidentification and the epistemology of *De Se* thought. In M. García-Carpintero & S. Torre (Eds.), *About oneself* (pp. 25–55). Oxford University Press.
- McGlynn, A. (2021). Immunity to wh-misidentification. *Synthese*, 199(1–2), 2293–2313. <https://doi.org/10.1007/s11229-020-02885-9>
- Morgan, D. (2012). Immunity to error through misidentification: What does it tell us about the *De Se*? In S. Prosser & F. Recanati (Eds.), *Immunity to Error Through Misidentification. New essays* (pp. 104–123). Cambridge: Cambridge University Press.
- Morgan, D. (2015). The demonstrative model of first-person thought. *Philosophical Studies*, 172(7), 1795–1811. <https://doi.org/10.1007/s11098-014-0391-z>
- 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>
- Ninan, D. (2010). *De Se* Attitudes: Ascription and communication. *Philosophy Compass*, 5(7), 551–567. <https://doi.org/10.1111/j.1747-9991.2010.00290.x>
- O'Brien, L. (2007). *Self-knowing agents*. OUP.
- Palmira, M. (2019). Arithmetic judgements, first-person judgements, and immunity to error through misidentification. *Review of Philosophy and Psychology*, 10(1), 155–172. <https://doi.org/10.1007/s13164-018-0395-2>
- 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>
- Parfit, D. (1984). *Reasons and persons*. Oxford University Press.
- Peacocke, C. (1999). *Being known*. Oxford University Press.
- Peacocke, C. (2008). *Truly understood*. Oxford University Press.
- Peacocke, C. (2014). *The mirror of the world*. Oxford University Press.
- Perry, J. (1979). The problem of the essential indexical. *Noûs*, 13(1), 3–21. <https://doi.org/10.2307/2214792>
- Pritchard, D. (2005). *Epistemic luck*. Oxford University Press.
- Pryor, J. (1999). Immunity to error through misidentification. *Philosophical Topics*, 26(1), 271–304. <https://doi.org/10.5840/philtopics1999261/246>
- Pryor, J. (2000). The Skeptic and the Dogmatist. *Noûs*, 34(4), 517–549.
- Recanati, F. (2007). *Perspectival thought. A plea for moderate relativism*. OUP.
- 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: Cambridge University Press.
- Salje, L. (2020). Lit from within: First-person thought and illusions of transcendence. *Canadian Journal of Philosophy*, 50(6), 735–749. <https://doi.org/10.1017/can.2020.22>
- Shoemaker, S. (1968). Self-reference and self-awareness. *The Journal of Philosophy*, 65(19), 555–567. <https://doi.org/10.2307/2024121>
- Shoemaker, S. (1970). Persons and their pasts. *American Philosophical Quarterly*, 7, 269–285.
- Shoemaker, S. (1994). Self-knowledge and ‘inner sense’. *Philosophy and Phenomenological Research*, 54(2), 249–315. <https://doi.org/10.2307/2108488>
- Stainton, R. (2019). Re-reading anscombe on ‘I’. *Canadian Journal of Philosophy*, 49(1), 70–93. <https://doi.org/10.1080/00455091.2018.1521695>
- Stephens Lynn, G., & Graham, G. (2000). *When self-consciousness breaks*. MIT Press.
- Verdejo, V. (2021). Perspectives on *De Se* immunity. *Synthese*, 198(10), 10089–10107. <https://doi.org/10.1007/s11229-021-03122-7>
- Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.
- Wiseman, R. (2017). What Am I and what Am I doing? *The Journal of Philosophy*, 114(10), 536–550. <https://doi.org/10.5840/jphil20171141037>
- 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>
- Wittgenstein, L. (1958). *The blue and brown books*. Harper and Row.
- Wright, C. (1998). Self-knowledge: The Wittgensteinian Legacy. In B. S. Wright & C. McDonald (Eds.), *Knowing our own minds* (pp. 13–45). New York: Oxford.
- Wright, C. (2004). I—Crispin Wright warrant for nothing (and foundations for free)? *Aristotelian Society Supplementary Volume*, 78(1), 167–212. <https://doi.org/10.1111/j.0309-7013.2004.00121.x>
- 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: Cambridge University Press.