**Immunity, Thought Insertion, and The First-Person Concept**

Michele Palmira

University of Barcelona, BIAP/LOGOS Research Group

Penultimate Draft

Forthcoming in *Philosophical Studies*

<https://doi.org/10.1007/s11098-019-01411-z>

Please refer to the published version when possible

**Abstract**

In this paper I aim to illuminate the significance of thought insertion for debates about the first-person concept. My starting point is the often-voiced contention that thought insertion might challenge the thesis that introspection-based self-ascriptions of psychological properties are immune to error through misidentification relative to the first-person concept. In the first part of the paper I explain what a thought insertion-based counterexample to this immunity thesis should be like. I then argue that various thought insertion-involving scenarios do not give rise to successful counterexamples to the immunity of the target class of self-ascriptions. In the second part of the paper I turn to defend a *Metasemantic* *Explanation* of why the immunity thesis holds. The Metasemantic Explanation rests on a reference-fixing story about the mental ‘I’ whose key contention is that introspective impressions play an essential role in fixing its reference. It is part of my argument in favour of the proposed reference-fixing story, as well as of the Metasemantic Explanation, that they respect the paradigmatic features of self-ascriptions of inserted thoughts.

**1 Introduction**

In the *Blue Book* (1958), Ludwig Wittgenstein maintains that any introspection-based self-ascription of occurrent mental properties, such as ‘I am thinking that it is raining’, is such that I cannot be mistaken that it is *I* who am instantiating the relevant mental property. To use Sidney Shoemaker’s (1968) influential label, this class of self-ascriptions is *immune to error through misidentification relative to the first-person concept.* Since then, philosophers of mind have widely taken immunity to error through misidentification as one of the most beneficial entry points to the central questions of the nature and representation of the self.

At the dawn of the literature on immunity to error through misidentification, G.E.M. Anscombe (1975) and Shoemaker (1968) joined Wittgenstein in maintaining that any introspection-based self-ascription of occurrent mental properties is immune to error through misidentification*.* (Call this *The Introspective Immunity Thesis*.) Despite such a distinguished roster of early endorsers and its apparent obviousness, however, The Introspective Immunity Thesis is no longer regarded as sacrosanct. Several philosophers (see e.g. Campbell 1999, Cappelen and Dever 2013, García-Carpintero 2018, Hu 2017, Recanati 2007, Salje 2016) have indeed taken thought insertion – a delusion associated with schizophrenia whose hallmark is that subjects report that the thoughts they have introspective access to are not their own – to be a phenomenon of the mind which might provide strong evidence against the universal generality of The Introspective Immunity Thesis.

John Campbell (1999, 2002) and Annalisa Coliva (2002a, 2002b) have debated the significance of thought insertion for the truth of The Introspective Immunity Thesis. Their debate operates under a specific conception of error through misidentification which, adopting James Pryor (1999)’s terminology, we may call “*de re*” misidentification. Yet, following the lead of Pryor (1999), several authors (see García-Carpintero 2018, Hu 2017, McGlynn 2015, Recanati 2012, Salje 2016, Wright 2012a) have acknowledged the existence of another variety of error through misidentification: what we may call, following again Pryor, error through “*which-object*” misidentification (*wh*-misidentification for short).

Despite the broad agreement both on the existence of error through *wh*-misidentification and the significance of thought insertion for debates about immunity to error through misidentification and the self, reading through the extensive literature on immunity surprisingly reveals that comparatively little attention has been so far devoted to the question of how thought insertion might provide a counterexample to The Introspective Immunity Thesis as formulated in its *wh*-misidentification variety.[[1]](#footnote-1) The first aim of this paper is to make progress with this question: I explain what a thought insertion-based counterexample should be like and argue that various thought insertion-involving scenarios do not give rise to successful counterexamples to The Introspective Immunity Thesis.

A second urgency surrounds the present investigation. Irrespective of how one answers the *extensionality question* – “Is such-and-such a class of first-person judgements immune/vulnerable to error through *wh*-misidentification?” – one has to face an *explanatory question* – “Why is such-and-such a class of first-person judgements immune/vulnerable to error through *wh*-misidentification?”. Focusing on introspection-based self-ascriptions of psychological properties, once we have answered the extensionality question by determining whether there are successful thought insertion-based counterexamples to The Introspective Immunity Thesis, we have to answer the explanatory question by providing an account of the immunity/vulnerability of such a class of self-ascriptions.

Plausibly, any account of why The Introspective Immunity Thesis holds/does not hold has to respect the paradigmatic features of self-ascriptions of inserted thoughts. This reveals that thought insertion is also central to the issue of how to explain the immunity/vulnerability of introspection-based self-ascriptions of psychological properties. And yet, little attention to this fact has been paid in the contemporary literature.[[2]](#footnote-2)

 The second aim of this paper is to explain why The Introspective Immunity Thesis holds. I offer a *Metasemantic Explanation* to the effect that introspection-based self-ascriptions of psychological properties are immune to error through *wh*-misidentification because introspective impressions – which act as grounds of such self-ascriptions – play an essential role in fixing the reference of the first-person concept. The Metasemantic Explanation, I maintain, both accommodates and vindicates the paradigmatic features of self-ascriptions of inserted thoughts.

In §2 I introduce the relevant notion of (immunity to) error through misidentification. In §3 I explain what it takes to raise a thought insertion-based counterexample to The Introspective Immunity Thesis. In §4, I consider and rebut three alleged thought insertion-based counterexamples to The Introspective Immunity Thesis. In §5-5.3 I develop the Metasemantic Explanation. In §6 I summarise the main findings of the paper.

**2 (Immunity to) Error Through Misidentification**

This section provides an introduction to the notion to (immunity) to error through misidentification. Given that the paper focuses exclusively on error through *wh*-misidentification (hereafter error through misidentification), I won’t go over the vast literature on how to distinguish between different varieties of misidentification and what these varieties have in common.[[3]](#footnote-3)

Let us focus on a judgement of the form ‘a is F’ and make two preliminary clarifications. First, 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. So, a judgement involves an error through misidentification only if there is an error about the concept in the subject position ‘a’,[[4]](#footnote-4) and not if there is an error about the concept in the predicative position ‘F’. To capture this, I’ll use the expression “error through misidentification relative to the concept ‘a’”. Bear also in mind that an error through misidentification occurs partly in virtue of the *grounds* upon which subjects make their judgements. In agreement with the literature (see Coliva 2006, McGlynn 2015, Pryor 1999, Wright 2012a), I’ll take a judgement’s grounds to be the causal/explanatory basis of one’s judgement: so, if a subject S’s judgement involves an error through misidentification relative to a certain concept ‘a’, then S judges ‘a is F’ on the basis of certain grounds Gs.

This said, consider now the following case (slightly revised from Pryor 1999: 281):

(SKUNK)

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, on the basis of the testimony of my friend Gustav, 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. However, even if there are several skunks in my garden, none of them is the small animal I see.

Since the falsity of the judgement ‘That animal is a skunk’ can be imputed either to an error about the concept in the subject position ‘That animal’ or to an error about concept in the predicate position ‘being a skunk’, and since error through misidentification concerns the former type of error only, we cannot explain what goes wrong when we make an error through misidentification in terms of the falsity of judgement ‘That animal is a skunk’. Pryor (1999) insightfully suggests looking at whether one’s grounds can be defeated qua epistemic warrant (hereafter warrant)[[5]](#footnote-5) for the singular judgement without being defeated qua warrant for a general judgement of the form ‘Something/one is F’.

Now, the grounds for my judgement ‘That is a skunk’ are constituted by my smelling experiences, my testimonial-based belief that not all skunks have the characteristic white stripes and my visual experiences as of a small animal without white stripes in front of me. Let us suppose that these grounds offer me a defeasible warrant for ‘That is a skunk’. Suppose that, later on, Gustav’s best friend, Alice, comes along and tells me that Gustav makes misleading claims about the bodily features, such as shape, colour, and peculiarities of animals. This, surely, sheds a negative light on the grounds upon which I base my singular judgement: the new piece of information discredits my grounds qua warrant for ‘That is a skunk’. However, receiving the information that my friend Gustav makes misleading claims about the bodily features of animals in no way discredits my smelling experiences qua warrant for the judgement ‘There is a skunk nearby’.

Thus, in (SKUNK) my smelling experiences allow me to detect a feature, i.e. the presence of a skunk, in my garden. This is what warrants me to judge ‘There’s a skunk nearby’. However, I have made a mistake in singling out which thing it is that is a skunk. This is so since the distinctive odour I am smelling still leaves it open whether it is the smallest of these animals I see that is a skunk. This is what it means that my judgement ‘That is a sunk’ involves an error through misidentification.

On closer inspection, (SKUNK) exemplifies just *one* way for a judgement to involve an error through misidentification. The error has to do with mistakenly answering the question “which thing is F?” by judging that it is *a* that is F. However, insofar as judgements involve an error through misidentification if they are mistaken qua answers to the question “which thing is F?”, the mistake can also occur if one judges of *a* to be not-F when in fact it is. Consider this example, due to Ivan Hu (2017: 121):

(PARK)

I decide to visit Yellowstone National Park for the first time. Upon arriving, I come to falsely but justifiably believe of this place (Yellowstone) that it is not within the Rocky Mountains. My friend, amused at my geographical ignorance, informs me that we are in the Rocky Mountains.

In (PARK), even if my geographical evidence warrants me to judge ‘Something is within the Rocky Mountains’, it falls short of warranting me, in light of my friend’s testimony, to judge that this place is not within the Rocky Mountains. So, my judgement ‘This place is not within the Rocky Mountains’ amounts to an error through misidentification, in that I misidentify which thing is within the Rocky Mountains − the place I am thinking about through the demonstrative ‘this’ in fact is, but I think that it is not.

To generalise, error through misidentification occurs when:

S has Gs that warrant S to judge ‘Someone is F’;

S is warranted to judge “a is (not) F’ partly on the basis of Gs;

Unbeknownst to S, *a* is not (is) F.[[6]](#footnote-6)

Take (SKUNK): my smelling experiences are the grounds that warrant the existential ‘There’s a skunk nearby’. Such smelling experiences are part of my grounds for ‘That animal is a skunk’; yet, my grounds for that judgement are also constituted by my visual experiences as of a small stripeless animal in front of me and my testimonial-based belief that not all skunks have the characteristic white stripes. Now, these three pieces of evidence provide me with a warrant for ‘That is a skunk’; however, upon learning that Gustav is not a reliable source of information about animals’ features, I acquire a defeater which undercuts my warrant for ‘That animal is a skunk’. My mistake in singling out which thing it is that is a skunk is explained by the fact that my smelling experiences leave it open which thing it is that is a skunk and do not provide me, by themselves, any warrant for the singular judgement.

We can now also define immunity and vulnerability through error misidentification:

**Immunity/vulnerability to error through misidentification:**

S’s judgement ‘a is (not) F’ made on grounds Gs is immune/vulnerable to an error through misidentification relative to the concept ‘a’ just in case it is impossible/possible that Gs warrant S to judge ‘Someone is F’ without warranting S to judge ‘a is (not) F’.

From this, it follows that at the roots of immunity/vulnerability to error through misidentification there is a modally-flavoured epistemic dependence/independence between singular judgements and existential judgements. This dependence/independence has to do with whether or not Gs leave any justificatory gap between ‘Someone is F’ and ‘a is (not) F’. Focusing on vulnerability and independence, I propose unpacking this justificatory gap idea as follows:

Epistemic Independence

It is possible for S to acquire defeating evidence that undercuts S’s Gs qua warrant for ‘a is/is not F’, but leaves them intact *qua* warrant for ‘Something/one is F’.

Epistemic Independence talks of *undercutting* − as opposed to *additive* − defeating evidence.[[7]](#footnote-7) An undercutting defeater defeats the evidential support for *p* without providing evidential support for its falsity. An additive defeater, by contrast, defeats the evidential support for *p* by providing evidence for its falsity. By taking the relevant defeater to be undercutting, we ensure that one’s warrant for the existential judgement is provided by one’s original evidence alone, thereby ruling out the possibility that such warrant is packed into the newly acquired defeating evidence, as happens with additive defeaters.

Having clarified the main features of immunity to error through misidentification, the next section addresses the question of how thought insertion can give rise to a counterexample to The Introspective Immunity Thesis.

**3 How to Raise a Thought Insertion-Based Counterexample to The Introspective Immunity Thesis**

Let me begin with mentioning a thought insertion report made by a deluded subject S:

I look out of the window and I think the garden looks nice and the grass looks cool, but the thoughts of Eamonn Andrews come into my mind. There are no other thoughts there, only his… He treats my mind like a screen and flashes his thoughts on to it like you flash a picture.

(In Mellor 1970: 17)

John Campbell’s influential model of thought insertion (1999, 2002) allows us to unpack the content of the report. Campbell maintains that the best way to account for the fact that S disowns being the thinker of the thought [the garden looks nice] [[8]](#footnote-8) while, at the same time, being aware of this thought on an introspective basis, is to recognise two strands in the notion of the ownership of a thought: being the owner of a thought *qua* its *author* and *qua* its *recipient*. Thus, if we retain the traditional idea that introspective awareness of our own thoughts is a way of latching onto ourselves qua thinkers of such thoughts, and given the fact that schizophrenic patients do have introspective awareness of inserted thoughts, we can take S to be making rationally intelligible reports by spelling them out as follows:

(1) I am recipient-thinking [the garden looks nice] but Eamonn Andrews is author-thinking it.

(2) I am recipient-thinking [the garden looks nice] but I am not author-thinking it.

The basic idea of a thought insertion-based counterexample to The Introspective Immunity Thesis can be put like this: by failing to ascribe authorship of thoughts to themselves on introspective basis, S is making a mistake in answering the question ‘which thing is author-thinking [the garden looks nice]?”. S says that it is not themselves, when in fact it is. However, S is still warranted to judge ‘Somebody is author-thinking [the garden looks nice]’. So, a thought insertion-based counterexample is going to be like (PARK) and unlike (SKUNK).[[9]](#footnote-9) Let us turn this basic idea into something more precise.

To begin with, our focus will be on the negative self-ascriptive part of the deluded subjects’ reports, namely:

(N1) Eamonn Andrews and not me is author-thinking [the garden looks nice].

(N2) I am not author-thinking [the garden looks nice].[[10]](#footnote-10)

The relevant existential judgement will therefore be:

(Ex) Somebody is author-thinking [the garden looks nice].

Secondly, let’s focus on what The Introspective Immunity says. The Thesis reads as follows: Any introspection-based self-ascription of occurrent mental properties is immune to error through misidentification*.* This formulation tells us that The Introspective Immunity Thesis is silent on what happens when S’s self-ascriptions are based both on introspection and some other grounds. So, The Introspective Immunity Thesis is successfully refuted only if introspection – and introspection *alone* – fails to warrant ascribing the relevant mental property to oneself while still warranting the judgement that somebody has that property. Note that the analogous condition is satisfied in cases of error through misidentification such as (SKUNK): my smelling experiences warranting the existential ‘There is a skunk nearby’ are also (part of) the grounds for my judgement ‘That animal is a skunk’. What we have in (SKUNK) is that the smelling experiences alone fail to warrant ascribing the property of being a skunk to the smallest animal I see while nonetheless warranting the existential ‘There is a skunk nearby’. By analogy: S’s warrant for (Ex) must arise from *the same introspective grounds* which are deployed by S while judging (N1) and (N2). To give a label to this *desideratum*, let us call it Preservation of Grounds.

 Thirdly, we should note that The Introspective Immunity Thesis is philosophically ambitious, as it consists in a universally quantified statement that ranges over any self-ascription of psychological properties based on introspection, irrespective of whether it’s a pathological or a non-pathological one. So, all that’s needed to refute The Introspective Immunity Thesis is one case. Now, if (N1) (N2) are precisely meant to be cases which enable us to refute the Thesis, then the grounds on which S bases (N1) and (N2) must be of the very same type of any other self-ascription of mental properties made basis of introspection falling under the scope of The Introspective Immunity Thesis. Call this *desideratum* Type-Sameness of Introspective Grounds.

 To ensure that S’s negative self-ascriptions (N1) and (N2) meet Type-Sameness of Introspective Grounds in a way which might give rise to a counterexample to the Introspective Immunity Thesis, I suggest taking *introspective impressions* to be the relevant grounds of (N1) (N2) and (Ex), in that this ensures that the deliverances of introspection in cases of mistaken and correct self-ascriptions are subjectively indistinguishable.[[11]](#footnote-11) Subjective indistinguishability is needed for two reasons. For one thing, it is (plausibly) a necessary condition for mistaken and correct self-ascriptions of psychological properties to meet Type-Sameness of Introspective Grounds. For another, subjective indistinguishability explains why a deluded S has no reason to distinguish between the deliverances of introspection in cases where S’s self-ascriptions of thoughts are correct, and cases where they are not (e.g. cases of thought insertion). This fact explains why S’s negative self-ascriptions such as (N1) and (N2) are rationally intelligible despite being mistaken, just as Campbell’s model of thought insertion has it.

 Combining Preservation of Grounds and Type-Sameness of Introspective Grounds with the relevant kind of epistemic independence required by the notion of vulnerability to error through misidentification, we are in a position to precisely state what it takes to construct a thought insertion-based counterexample to The Introspective Immunity Thesis:

Constraint

By Preservation of Grounds:

S’s grounds Gs on which S bases (N1) ‘Eadmonn Andrews and not me is author-thinking [the garden looks nice]’ and (N2) ‘I am not author-thinking [the garden looks nice]’ must also be the grounds on which S bases (Ex) ‘Someone is author-thinking [the garden looks nice]’.

By Type-Sameness of Introspective Grounds:

Gs must S’s introspective impressions as of a thought passing through S’s mind.

By Epistemic Independence:

It must be possible for S to acquire defeating evidence that *undercuts* S’s introspective impressions qua warrant for ‘I am not author-thinking [the garden looks nice]’ but leaves them intact qua warrant for ‘Someone is author-thinking [the garden looks nice]’.

Armed with Constraint, we can now turn to assess purported thought insertion-based counterexamples to The Introspective Immunity Thesis.

**4 Rebutting Thought Insertion-Based Counterexamples to The Introspective Immunity Thesis**

In this section I will be mainly focusing on the only, as far as I am aware, well-worked out attempt at constructing a thought insertion-based counterexample to The Introspective Immunity Thesis, due to Hu (2017). I will work out two different readings of Hu’s thought insertion scenario and argue that none of them succeeds qua counterexample to The Introspective Immunity Thesis. I will then briefly turn to adapt a pattern of counterexample proposed by Joel Smith (2006) to the case of thought insertion and show what’s wrong with it.

Hu (2017: 125, fn. 35) asks us to consider the case in which a schizophrenic subject S’s grounds for (N1) and (N2) are undercut via testimony: S learns that their medical condition has hallucinatory effects. Secondly, Hu observes that S is able to report the content of the inserted thought. Thirdly, he maintains that disavowing ownership of the thought by attributing it to somebody else presupposes that the thought belongs to someone. Fourthly, Hu stipulates that S does not countenance the alleged possibility, hypothesised by Lichtenberg, of thinking without author-thinkers.

 I will begin with arguing that Hu’s fourth consideration is necessary to establish that S’s grounds for (N1) and (N2) are undercut while his grounds for (Ex) are not. Crucially, this will lead us to realise that there are two ways of interpreting Hu’s counterexample.

To begin with, it is eminently plausible to maintain that the testimonial evidence received by S, i.e. that their condition makes their introspective impressions hallucinatory, does not *ipso facto* tell S that their introspective impressions are misleading *only when* they are deployed to ground (positive or negative) ascriptions of mental properties to someone in particular. There is nothing in the very idea of one’s introspective (or perceptual- or memory-based) impressions being hallucinatory that guarantees that they are epistemically defective only when they are used to ground singular, as opposed to general, ascriptions of properties. So, Hu’s first consideration, while necessary to ensure that S receives defeating evidence which undercuts S’s grounds for (N1) and (N2), does not suffice to establish that (N1) and (N2) are cases of error through misidentification.

 Secondly, the fact that S reports the content of the introspected thought, together with the testimonial evidence mentioned above, does not guarantee that S has a reason to think that somebody is the author of the thought with content *p*. Indeed, by entertaining the possibility of thinking without author-thinkers, one might report the content of T without presupposing that someone has authored it.

 The third of Hu’s considerations draws on the following feature of (N1): S ascribes authorship of the introspected thought to a particular individual (i.e. Eamonn Andrews). This, Hu observes, presupposes that someone author-thinks [the garden looks nice]. This is, no doubt, a structural feature of the relevant self-ascription, in that any singular mental ascription entails an existentially general one. However, for one thing, it is not clear how such a feature can ensure that the grounds are not undercut qua warrant for the general ascription but are undercut qua warrant for the singular one. The fact that attributing authorship of a thought to a particular individual presupposes that the thought is authored by someone in general does not *ipso facto* show that there’s a justificatory gap between (N1) and (Ex). For another, the delusion of thought insertion does not necessarily involve ascriptions of authorship to an individual in particular. So this third consideration could not explain why a negative self-ascription such as (N2), in which S does not ascribe authorship of the thought to anyone in particular, involves an error through misidentification.

 So the first three considerations offered by Hu, taken together, do not suffice to establish that S’s grounds for not ascribing authorship of the thought to themselves are undercut while being left intact *qua* grounds for judging that somebody is thinking it. However, if we build into the case that S rules out the possibility of there being thinking without author-thinkers, it becomes reasonable for S to believe that the reason why S has an introspective impression as of the thought [the garden looks nice] passing through their mind is that someone, rather than no one, is author-thinking it. The same can be said with respect to (N2): given that S rules out the possibility that the introspected thought has no author, it is reasonable for S to believe that someone is author-thinking it. Thus, S is indeed warranted to judge (Ex) even if their grounds for judging that it is not themselves who is author-thinking are undercut. This ensures that Epistemic Independence is met.

 On closer inspection, however, S’s taking there to be no thinking without author-thinkers – let’s henceforth take this to be a belief of S’s whose content, let’s stipulate, is [thoughts have authors] – can be epistemically relevant in two ways: either that belief is part of S’s grounds for (N1) and (N2), or it is a background presupposition of these self-ascriptions. Within the immunity debate, the distinction between self-ascriptions’ grounds and background presuppositions has been firstly theorised in Coliva (2006), and subsequently endorsed by several authors (e.g. García-Carpintero 2018, Recanati 2012, Wright 2012). Roughly put, something is part of S’s grounds for a judgement only if it appears in the subject’s psychology. For instance, if I see someone reading *Infinite Jest* and I also believe that the person is Alice, I go on to judge ‘Alice is reading *Infinite Jest*’ by consciously inferring it from that identity belief together with my perceptual experience. I can also 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”. In both examples, my identity belief (i.e. that person is Alice) is foregrounded in my psychology: in this sense, that belief is part of the grounds for my judgement ‘Alice is reading *Infinite Jest*’. By contrast, let’s take the memory-based self-ascription ‘I was in Scotland five years ago’. As persuasively argued by Evans (1982), I don’t form that judgement by (partly) consciously inferring it from the belief that I am the person from whose past this memory impression derives; nor would I offer such a belief in an attempt to rationally reconstruct the basis of my self-ascription, for I would most likely only say that I came to that judgement on the basis of my memory experiences. This has led Coliva and many others to conclude that the identity belief [I am the person from whose past this memory impression derives] is not foregrounded in my own psychology, but it is rather a background presupposition of my self-ascription.

Given the distinction between grounds and background presuppositions, S’s belief [thoughts have authors] can either be part of S’s grounds for (N1) and (N2), or be a background presupposition of those self-ascriptions. I will take both options in turn and argue that none of them helps turn Hu’s scenario into a successful counterexample to the Introspective Immunity Thesis.

As has emerged previously, without the belief [thoughts have authors], S couldn’t reasonably maintain that their testimonial evidence does not undercut their grounds for (Ex). If that belief is part of S’s grounds, however, it is not S’s original introspective grounds – and *only* them – that warrant S to judge ‘Someone is author-thinking [the garden looks nice]’ despite their being undercut *qua* grounds for judging that it is not themselves who is author-thinking it. However, Preservation of Grounds maintains that it must be the case that S’s introspective impressions *alone* fail to warrant them to judge that it is not themselves who is author-thinking [the garden looks nice] while, at same time, warranting them to judge that Someone is author-thinking [the garden looks nice]. So, the option of putting the belief [thought have authors] into the subject’s grounds leads to a violation of Preservation of Grounds. For this reason, this reading of Hu’s scenario can’t give rise to a successful counterexample to The Introspective Immunity Thesis.

Let’s now consider the second option to the effect that [thoughts have authors] is a background presupposition of any self-ascription of psychological properties. Surely, as emerges from the contemporary literature on external world scepticism, some background presuppositions are needed in order for any judgement to be warranted by experience. Take for instance my judgement ‘There’s a tree in front of me’ grounded on a visual perceptual experience as of a tree in front of me. It certainly makes sense to say that this experience falls short of warranting my judgement if it is called into doubt that there exists an external material world or that my senses are reliable. So, by analogy, the same should be true of introspective impressions: the background presupposition [thoughts have authors] must stay put in order for such impressions to afford us a warrant for our self-ascriptions of psychological properties.

The foregoing makes the following reading of Hu’s scenario appealing. Since S’s belief [thoughts have authors] is a background presupposition of (N1) and (N2), the scenario now satisfies Preservation of Grounds, for the grounds for (N1) and (N2) are solely introspective. Since the scenario already met Type-Sameness of Introspective Grounds and Epistemic Independence, it follows that Constraint is met: Hu’s scenario does constitute a counterexample to the Introspective Immunity Thesis.

Appealing though this may be, I will argue that, on closer inspection, this revised Hu’s scenario is not theory-neutral with respect to the notion of warrant. For this reason, we should revoke the status of counterexample to it.

 To begin with, let me introduce the well-established distinction between so-called *liberal* and *conservative* approaches to how experiences can provide us with a warrant for empirical beliefs. According to liberalism (see e.g. Pryor 2000, 2004), my perceptual experience as of a hand in front of me provides me with a *prima facie* warrant for the belief that there’s a hand in front of me in light of the background presuppositions that the external world exists, my senses are reliable, and the like, if I have no reason to doubt such background presuppositions. Crucially, according to liberalism, I need not have any prior warrant for the background presuppositions in order for experiences to provide warrants for my ordinary beliefs. According to conservatism (see Wright 2004, 2007, 2012b, 2014), by contrast, perceptual experiences afford me warrants for ordinary empirical beliefs in light of certain background presuppositions only if I have a prior warrant – to be understood in terms of a non-evidential and un-acquired *entitlement* to trust – for such background presuppositions.

Extending the liberal/conservative approaches to introspective warrant, I turn now to argue that Hu’s scenario works as a counterexample in the second way suggested above only if we accept a liberal view of warrant. To see why, notice that the background presupposition [thoughts have authors] is short for a general principle of the form: For all thoughts T, there exists an author of T. We can generate a specific instance of this principle by replacing T with the thought [the garden looks nice]. Hence, the background presupposition [thoughts have authors] ends up entailing, given the thought [the garden looks nice], [someone is author-thinking that the garden looks nice]. This existential presupposition is precisely the content of the existential judgement (Ex) ‘Someone is author-thinking [the garden looks nice]’. [[12]](#footnote-12) This, I submit, leads to an undesired result once we complete the picture with the conservative theory of warrant.

If we assume that introspective experiences do warrant self-ascriptions of psychological properties, we are also assuming that [thoughts have authors] is warranted. Now, bear in mind that [thoughts have authors] entails, assuming the existence of a thought with content [the garden looks nice], [someone is author-thinking that the garden looks nice]. If we take – as the conservative does (see e.g. Wright 2004) – warrant to be closed under entailment, and if [thoughts have authors] is warranted, it follows that [someone is author-thinking that the garden looks nice] is warranted. As a consequence, if we adopt a conservative view of warrant and we grant that the relevant background presupposition is [thoughts have authors], it follows that introspective impressions will never be able to provide a first (or enhance one’s previous) warrant for the existential judgement ‘Someone is author-thinking [the garden looks nice]’, for that judgement must already be epistemically warranted by default in order for introspective impressions to be able to justify ordinary self-ascriptions of psychological properties. So, (N1)-(N2) and (Ex) have different epistemic support.

 Let me articulate this line of reasoning in some detail. In order for a counterexample to The Introspective Immunity Thesis to be amenable to a conservative theory of warrant, it must be possible that our introspective impressions + our warrant by default (entitlement) can – *together* – provide us with a warrant for a background presupposition [T has an author] without thereby providing a warrant for a singular ascription of authorship. Yet, this possibility is ruled out by conservatism. Take the perceptual case first: perceptual experiences do not and cannot contribute to the warrant of the background presuppositions of the epistemic practice of warranting our ordinary empirical beliefs on perceptual grounds because of the so-called failure of transmission of warrant (Wright 2004, 2007, 2012b). [[13]](#footnote-13) To illustrate it, let’s focus on Moorean-style arguments purporting to show that we are perceptually warranted to believe the relevant background presuppositions, such as the following:

1. Here is a hand.
2. If there’s a hand here, then there is an external world.
3. There is an external world.

Wright’s idea is that normal perceptual experiences as of a hand in front of one can warrant (I) only if one is already warranted in accepting (III). So, if one has a perceptual warrant for (I) then, by the principle of closure for warrant across entailment, one also obtains a warrant for (III). And yet, one’s warrant for (III) does not transmit from the premises (I) and (II), for (III) has to be antecedently warranted in order for (I) to be warranted at all. This makes a Moorean-style argument like (I)-(II)-(III) epistemically circular. To avoid epistemic circularity, conservatism is bound to reject the claim that our perceptual experiences + our warrant by default (entitlement) can – *together* – provide us with a warrant for a background presupposition such as that there is an external world. So, the (I)-(II)-(III) argument cannot be used to enhance one’s previous warrant to accept (III).

Let us now draw an analogy between the perceptual case and the introspective case by envisaging a Moorean-like argument purporting to show that we can be introspectively warranted to believe the relevant background presuppositions, such as the following:

1. Here’s a thought T.
2. If here’s a thought T, then T has an author.
3. T has an author.

Just like its (I)-(II)-(III) cousin, the (i)-(ii)-(iii) argument affords us an intelligible explanation of how introspective impressions could warrant (iii). Yet, the conservative is bound to regard the (i)-(ii)-(iii) argument just epistemically circular as its perceptual cousin.[[14]](#footnote-14) For the conservative maintains that the relevant background presuppositions can only be warranted by default, and it is in virtue of the fact that we are warranted by default (or entitled, in Wright’s terminology) to trust such background presuppositions – the conservative maintains – that our introspective impressions can justify our ordinary self-ascriptions of psychological properties.

To recapitulate: Let us grant that certain background presuppositions must be in place in order for our experiences to provide us with warrants for our judgements. Once we couple this way of reading Hu’s scenario with a plausible theory of warrant, i.e. conservatism, the foregoing considerations show that the scenario cannot constitute a counterexample to The Introspective Immunity Thesis, for (N1)-(N2) and (Ex) will end up having different epistemic support, *contra* Epistemic Independence.

On the face of it, a liberal view of warrant does not run up against the worry just raised, for we need not have any warrant for the background presuppositions in order for experiences to provide warrants for our ordinary beliefs. So, one might think that the counterexample can be redeemed by simply adding to the scenario that liberalism about warrant is correct. Note that this would make the counterexample to The Introspective Immunity Thesis highly theory-laden, in that such counterexample would have to carry a commitment to a specific theory of how experiences provide us with warrants for ordinary beliefs in light of certain background presuppositions. As a consequence, Hu’s thought insertion-based counterexample would stand or fall with the falsity of the conservative view of warrant (and the truth of liberalism).[[15]](#footnote-15) This, I submit, should lead us to revoke the status of counterexample to Hu’s revised scenario. Let me explain.[[16]](#footnote-16)

Consider some philosophically familiar scenarios, such as the Gödel and Schmidt’s devised by Kripke (1980), the messy shopper’s devised by Perry (1979), the Smith and Jones’s devised by Gettier (1963). All these scenarios are meant to act as counterexamples to philosophical theses, respectively: the descriptive theory of proper names, the traditional model of propositional attitudes, the definition of knowledge as justified true belief. Now, the (alleged) philosophical significance of those scenarios rests on the fact that they constitute new *explananda* for our philosophical theories. Theoretical responsibility dictates that the description of an *explanandum* be as independent as possible from the set of theses that might eventually constitute its *explanans*. (Note, for instance, that Perry’s messy shopper scenario does not assume Perry’s own revision of the traditional theory of propositional attitudes, as witnessed by the fact that the alternative Lewisean approaches do engage with the messy shopper scenario without any problem.)

The same holds for Hu’s scenario. That scenario is meant to constitute an *explanandum* that challenges the extensional adequacy of a philosophical *explanans*, i.e. The Introspective Immunity Thesis. As has emerged in §2, the articulation of the notion of immunity to error through misidentification features the notion of warrant, but does not commit us to any specific theory thereof. So, a description of a thought insertion scenario should not commit us to a specific theory of warrant, for that would lead to a characterisation of the *explanandum* which is already biased towards a certain *explanans*. Moreover, if we constructed a thought insertion-based scenario that assumes liberalism and we granted the status of counterexample to The Introspective Immunity Thesis to it, we would be begging the question against those theorists – such as Wright (2012a), for instance – who advocate The Introspective Immunity Thesis without also accepting liberalism. Thus, the fact that Hu’s case succeeds as a counterexample only if we take a far from mandated stance on how experiences can warrant our judgements makes the counterexample bereft of the kind of theoretical neutrality we customarily require of counterexamples.[[17]](#footnote-17)

We have been looking closely at two interpretations of Hu’s scenario since they had the potential to satisfy Epistemic Independence. This *desideratum* proves hard to be satisfied, as witnessed by the fact that other counterexamples clearly fail because they do not meet it. To illustrate this point briefly, let me elaborate on a pattern of counterexample proposed by Smith (2006: 279). Imagine that a doctor tells S that S’s introspective impressions as of being the recipient of a thought of which S is not the author are all incorrect, and that S is in fact the author of the introspected thoughts. This warrants S to judge that someone is author-thinking [the garden looks nice]. However, this is evidence that (N1) and (N2) are false, that is, the evidence is additive as opposed to undercutting. This means that Smith’s recipe fails to guarantee the kind of epistemic independence between the singular judgement and the existential judgement that is required in order for there to be vulnerability to error through misidentification. As a consequence, this alternative specification of the scenario fails to satisfy Constraint since it does not satisfy Epistemic Independence.

Summing up: I have considered some ways in which we can devise a thought insertion-based counterexample to The Introspective Immunity Thesis, and I have found all of them wanting. It is now time to develop an account of why The Introspective Immunity Thesis holds.

**5 The Metasemantic Explanation of The Introspective Immunity Thesis**

*5.1 Motivation and agenda*

I will argue for a *Metasemantic Explanation* of The Introspective Immunity Thesis to the effect that introspection-based self-ascriptions are immune to error through misidentification because introspective impressions play an essential role in fixing the reference of the first-person concept. I want to focus here on developing the Metasemantic Explanation for two main reasons. First, the two alternative and prominent *deflationary* and *content-based* accounts of immunity, namely views seeking to explain immunity either in terms of the absence of an identification component in the ascriptions’ grounds or background presuppositions, or in terms of their distinctively *de se* content, have been heavily criticised in the literature.[[18]](#footnote-18) As I see it, the current state of the debate on deflationary and content-based accounts requires of us not so much to rehearse those criticisms or attempt to vindicate versions of one of these two accounts – this would in fact threaten the debate to go stale – but to develop underexplored alternatives. Secondly, it seems to me that the authors who have considered the possibility of offering a Metasemantic Explanation of immunity have not taken proper account of the significance of thought insertion for the question of how the reference of the first-person concept is fixed.[[19]](#footnote-19) So, it is important to set the record straight and work out the precise content of the Metasemantic Explanation in order to put it on the table as a serious option.

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:[[20]](#footnote-20)

**Bare Token-Reflexive Rule:**

‘I’ =ref the thinker of this thought

According to the **Bare** **Token-Reflexive Rule**, any token of ‘I’ refers to the thinker of that token. That is to say, any token first-person thought is about the person whose token it is, namely the person who produces the token.

The **Bare** **Token-Reflexive Rule** is not an explanation of why first-person thoughts have the references that they do. Rather, the rule makes it explicit what seems to be an obvious and partly constitutive fact of what a first-person thought is, that is: first-person thoughts are thoughts that have the semantic profile of referring to the producer of tokens of such thoughts. However, we cannot rest content with the **Bare Token-Reflexive Rule**, for two reasons. First, the right-hand side of the **Bare** **Token-Reflexive Rule** features a demonstrative expression, i.e. ‘this’. Thus, the rule does not really deliver a referent for ‘I’ unless this determinable aspect of the rule gets determined. This calls for an explanation, for we should wonder: how does one latch onto the right thought, namely the occurrent thought with content [*p*]?

To put the same point differently: the **Bare** **Token-Reflexive Rule** can be seen as a function that, taking a token thought as an argument, is meant to return the thinker of that token as a value. However, insofar we do not have an answer to the question of how the token thought is fed into the function as an argument, that is, insofar as we do not have an answer to the question “How does ‘this’ pick out the occurrent thought?”, we have not yet uncovered the pattern of reference of the first-person concept. Call this the *Demonstrative Determination Problem*.

Secondly, as has emerged above, an adequate description of thought insertion reports rests on the distinction between being *a thinker-qua-author* and being a *thinker-qua-recipient* of a thought. Since the **Bare** **Token-Reflexive Rule** deploys the notion of a “thinker” of a thought, we should ask: “Which thinker? The thinker-qua-author or the thinker-qua-recipient of the thought?”. There indeed is a metaphysical possibility that the thinker-qua-author and the thinker-qua recipient come apart – recall that this possibility is what allows us to make reports of inserted thoughts such as (1) and (2) rationally intelligible by spelling them out via (N1) and (N2). Now, in light of both this possibility and the fact that deluded subjects deploy the first-person concept in their reports, there is an ambiguity within the notion of a “thinker” which prevents the **Bare** **Token-Reflexive Rule** from revealing the semantic profile of the first-person concept (Bear in mind that since the rule articulates what a first-person thought constitutively is, “I” should have the same pattern of reference across all metaphysically and epistemically possible worlds). Call this the *Thinker Determination Problem*.

In my view, the foregoing observations set the agenda for the theorist who wants to defend a Metasemantic Explanation of The Introspective Immunity Thesis appealing to the token-reflexive model of first-person thought: this explanation can be developed only if the Demonstrative Determination Problem and the Thinker Determination Problem get solved. In the next subsection I take up this twofold task.

*5.2 The Full Token-Reflexive Rule*

Let me tackle the Demonstrative Determination Problem first, and begin with the idea that occurrent thoughts are phenomenally conscious: the passing of a thought through one’s mind appears in one’s stream of consciousness, and there is something it is like for one to have a thought *as one has it*. To forestall misunderstandings, claiming that occurrent thoughts have a phenomenal character does not amount to claiming that the phenomenal character of thoughts is different from the phenomenal characters that are present in non-cognitive mental states. Whether or not thoughts come with a *sui generis* phenomenology is object of a substantive debate between endorsers and deniers of so-called cognitive phenomenology. However, as pointed out in Chudnoff’s (2015) and Smithies’s (2013) survey works on cognitive phenomenology, detractors of cognitive phenomenology do not deny that cognitive states also have a phenomenal character: rather, they deny that such phenomenal character exhibits a mark of *sui generisity* which sets it apart from sensory phenomenology. Thus, I take it to be harmless to assume that occurrent thoughts have a phenomenal character. Importantly, this extends to thought insertion: reports of inserted thoughts are reports of occurrent thoughts.

This clarified, the hypothesis I want to defend is the following: the phenomenal conscious nature of the occurrent thought, together with the introspective impression of it, are jointly sufficient conditions for there to be a mental demonstration to the thought S is presently thinking.[[21]](#footnote-21) Thus, whenever S uses ‘I’ to report that there’s a thought passing through S’s mind on introspective grounds, that token of ‘I’ is guaranteed to pick out the right referent of ‘this’, i.e. the occurrent thought. My argument in favour of this hypothesis goes as follows:

1. The thought that S is presently thinking has content [p] and phenomenal character Φ. (Assumption)
2. S self-ascribes a thought introspective grounds. (Assumption)
3. S’s introspective impression as of a thought passing through their mind is partly constituted by the thought’s phenomenal character. (Assumption)
4. If S has an introspective impression as of a thought passing through their mind, S has an introspective impression as of a thought with Φ. (From 1 and 3)
5. If S self-ascribes a thought on introspective grounds, S self-ascribes a thought with Φ. (From 4)
6. S self-ascribes a thought with Φ. (From 2, 5, and MP)
7. If S self-ascribes a thought with Φ, S self-ascribes the thought they are presently thinking. (From 1)
8. S self-ascribes the thought they are presently thinking. (From 6, 7, and MP)

Assumptions (1) and (2) are harmless in the present context, as the cases we are focusing on are introspection-based self-ascriptions of occurrent thoughts. Let me then dwell on the crucial assumption (3) for a moment.

Offering a cast-iron defence of assumption (3) comes down to arguing against accounts of self-knowledge which deny the existence of a (partly) constitutive relation between the introspective state and the introspected state. (See Gertler 2011 for an overview and further references). This task cannot be responsibly undertaken in the space of this paper. However, I shall sketch out an account of introspection that affords the means to securing the partly constitutive link between introspective states and introspected states needed in order for assumption (3) to hold while, at the same time, avoiding some of the (allegedly) implausible implications that are usually associated with accounts of self-knowledge that posit relations of constitution between the introspective state and the introspected state.

On carefully worked-out new *acquaintance* views of introspection,[[22]](#footnote-22) the relationship between *certain* phenomenally conscious states and the introspective impression as of there to be such a state in one’s mind is constitutive: the introspective state is partly constituted by the introspected state.[[23]](#footnote-23) On such views, then, *for certain phenomenal states*, introspective impressions cannot give false information as to what phenomenal states get introspected, as the latter partly constitute the former. Now, as my guarded existential formulations suggest, the acquaintance view of introspection I am endorsing does not claim that introspection is *tout court* infallible. Let me explain.

As Gertler (2012) points out, the acquaintance view of introspection need neither assume that every experience is introspectively accessible, nor is it committed to the thesis that we can always get phenomenal states right. The first point is compatible with claiming that S can access the phenomenally conscious thought they are presently thinking. In fact, this latter claim is needed in order for us to pose the question of whether or not any introspection-based self-ascriptions of mental properties is immune to error through misidentification. The second point is compatible with the idea that even if we might not be introspectively able to discriminate some complex phenomenal characters – think of the phenomenal difference between our visual experience as of a hen with 48 visible speckles and our visual experience as of a hen with 47 visible speckles – this does not *ipso facto* show that we are unable to introspectively discriminate less complex phenomenal characters. As I see it, the what-is-likeness of thinking a thought is considerably less complex than the what-is-likeness of having a visual experience as of a hen with 47 speckles. A third important point is that since we are here focusing on introspection of occurrent thoughts, the question of whether or not introspection systematically alters the introspected state is immaterial, as the relevant introspective impressions give S information about S’sphenomenally conscious thought at the time S has that thought.

These three remarks point towards the following: while the acquaintance view of introspection does not maintain that any phenomenal state can be infallibly introspected, thereby failing to guarantee that *any* introspective impression is partly constituted by the introspected phenomenal state, it is plausible to think that phenomenally conscious occurrent thoughts do fall within the remit of the acquaintance view. That is to say, it is plausible to think that the phenomenal character with which occurrent thoughts come is amongst those phenomenal states of which we are introspectively aware through acquaintance. This enables us to claim that by attending to the phenomenal mark of the thought via their introspective impressions, S can’t but be introspecting the thought with phenomenal character Φ, for there could not be any introspective impression as of a thought passing through S’s mind without such an impression be (partly) constituted by the phenomenally conscious thought S is presently thinking.

The foregoing allows us to solve the Demonstrative Determination Problem as follows. Consider S’s introspection-based self-ascriptions of occurrent thoughts: If such introspective grounds are partly constituted by the phenomenal character Φ of the thought S is presently thinking, then S’s introspective awareness of Φ allows S to attend to the thought they are presently thinking. Thus, whenever S tokens ‘I’ on introspective grounds, S’s introspective impressions work as a mental demonstration which is guaranteed to pick out the right referent of ‘this’, i.e. S’s occurrent thought, in virtue of the fact that the phenomenal character of such thought partly constitutes S’s introspective impressions.[[24]](#footnote-24) Thus, whenever S deploys the first-person concept on introspective basis, the problem of how the referent of ‘this’ gets determined is solved.

Importantly, this solution of the Demonstrative Determination Problem carries over to cases in which schizophrenic patients token the first-person concept; so, the solution respects the widely endorsed contention that thought insertion does not show that thought insertion-deluded subjects lack the first-person concept.

Let us now tackle the Thinker Determination Problem. Let’s assume, somewhat standardly (see e.g. Peacocke 2008: 111), that if S possesses and deploys the first-person concept competently, S deploys ‘I’ in accordance with the token-reflexive rule. This is so since the rule underwrites the canonical patterns of a use that S must be disposed to make of ‘I’ in order to count as a competent user. Now, if we solved the Thinker Determination Problem by taking the thinker of T to be the author of T, it follows that the token-reflexive rule should read as follows:

‘I’ =ref the author-thinker of thisintrospectively accessed thought

 This token-reflexive rule underwrites the following introduction and elimination rules for ‘I’:

*x* is the author of the introspectively accessed T

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Iauthor-intro

 I am *x*

 I am *x*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Iauthor-elim

 *x* is the author of the introspectively accessed T

However, S’s self-ascriptions of inserted thoughts do not follow such rules. For one thing, S denies that they are the author of the thought they are introspectively aware of, so they cannot eliminate ‘I’ according to Iauthor-elim. For another, (N1) shows that if S ascribes authorship of T to anyone at all, they ascribe it to someone else than themselves; so, S is not disposed to introduce ‘I’ via Iauthor-intro. So, either S’s own use of ‘I’ challenges the token-reflexive rule of ‘I’ as specified above, or S doesn’t use ‘I’ competently. Yet, the last option is ruled out by the widely held description of thought insertion proposed by Campbell, as there’s no reason to think that self-ascriptions of inserted thoughts are not genuine self-ascriptions. Therefore, a token-reflexive rule in which the thinker is the author-thinker of T fails to explain S’s competent deployment of ‘I’.

Let’s now solve the Thinker Determination Problem by taking the thinker of T to be the recipient of T. This solution gives rise to:

‘I’ =ref the recipient-thinker of thisintrospectively accessed thought

This token-reflexive rule underwrites the following introduction and elimination rules for ‘I’:

*x* is the recipient of the introspectively accessed T

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Irecipient-intro

 I am *x*

 I am *x*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Irecipient-intro

 *x* is the recipient of the introspectively accessed T

Clearly, S’s self-ascriptions (N1) and (N2) do not challenge these inferential patterns. So, the fact that S uses the first-person concept competently and therefore obeys to the rule determining its pattern of reference is best explained by taking the token-reflexive rule to be specified in terms of the notion of recipient-thinker, as opposed to author-thinker. This gives us a neat solution to the Thinker Determination Problem: the producer of the thought is the thinker-qua-recipient of the thought.

We are now in a position to state the full content of the token-reflexive rule for ‘I’:

**Full Token-Reflexive Rule:**

‘I’ =ref the thinkerrecipient of thisintrospectively accessed thought

The **Full Token-Reflexive Rule** departs from the **Bare Token-Reflexive Rule** in that it offers a substantive explanation of what fixes the reference of ‘I’. Surely one might ask what general principle of reference determination explains why introspective impressions essentially contribute to fixing the reference of ‘I’. One such general principle might be that singular mental reference is governed by an epistemic constraint, to be perhaps understood along the lines of a liberal interpretation of the notion of acquaintance. Adopting this principle of reference determination, we can claim that introspective impressions essentially contribute to fixing the reference of ‘I’ since they enable us to be acquainted with ourselves. For the purposes of this paper, however, we can remain neutral as to what principle of reference determination operates in the background of the **Full Token-Reflexive Rule**.[[25]](#footnote-25)

Let us take stock. I have argued that the **Full Token-Reflexive Rule** accounts for the pattern of reference of the first-person concept. The **Full Token-Reflexive Rule**’s commitment to the idea that our introspective experiences as of thoughts passing through our minds are partly constituted by the phenomenal character of such thoughts makes the proposed account controversial. However, I have harnessed new acquaintance views of introspection to assuage some of the usual worries raised against the constitution claim. Moreover, those philosophers who have some affinity to the token-reflexive explanation of first-person thought could reason as follows: given that the acquaintance view of introspection features in the best (perhaps unique) explanation of how the reference of ‘I’ gets fixed, this provides further support – via inference to the best explanation – in favour of such an account of introspection. So, the **Full Token-Reflexive Rule**’s commitment to the constitution claim makes the account on offer both interesting and defensible.

To conclude this section, I should like to emphasise how thought insertion has been relevant to the question of how the reference of the first-person concept is fixed, thereby leading to the **Full Token-Reflexive Rule**. Note that the Thinker Determination Problem for the token-reflexive rule emerges precisely from Campbell’s description of thought insertion. Crucially, my argument for the Full Token-Reflexive Rule rests on the fact that self-ascriptions of inserted thoughts are genuine *self*-ascriptions; that is, patients do possess and use competently the first-person concept. This demonstrates the significance of the features of thought insertion for working out the referential pattern of the first-person concept.

*5.3 The Introspective Immunity Thesis Explained, Metasemantic-Style*

The **Full Token-Reflexive Rule** is integral to a proper development of the Metasemantic Explanation of The Introspective Immunity Thesis. Here is how the explanation goes: introspection-based self-ascriptions are immune to error through misidentification because introspective impressions play an essential role in fixing the reference of the first-person concept. Taking introspective impressions to constitute the grounds of such self-ascriptions, this provides us with an account of The Introspective Immunity Thesis which works at the right, i.e. epistemic, explanatory level, for the account singles out a property of the grounds of one’s self-ascription. The account is *metasemantic* since the relevant property of the grounds is that they contribute essentially to fixing the reference of the first-person concept.

Accordingly, The Introspective Immunity Thesis is explained as follows: insofar as S’s introspective impressions as of a thought passing through S’s mind warrant S to judge ‘Someone is F’, such introspective impressions can’t but warrant S to judge ‘I am F’. This epistemic dependence is guaranteed to hold in virtue of the fact that S’s introspective impressions essentially contribute to fixing the reference of the first-person concept as stated by the **Full Token-Reflexive Rule**: whenever S tokens the concept ‘I’, the concept picks out the individual having introspective impressions as of an occurrent thought passing through their mind.

Armed with this, let us now repeat (N1), (N2) and (Ex) from above:

(N1) ‘Eamonn Andrews and not me is author-thinking [the garden looks nice]’

(N2) ‘I am not author-thinking [the garden looks nice]’

(Ex) ‘Someone is author-thinking [the garden looks nice]’

Bear in mind that in order for there to be vulnerability to error through misidentification, (Ex) has to be warranted by S’s introspective impressions. I’ve argued earlier that attempts at showing that subjects making judgements such as (N1) and (N2) can nonetheless have a warrant for (Ex) fail. That is to say, (N1) and (N2) cannot be used to show that The Introspective Immunity Thesis fails. We can now deploy the metasemantic considerations offered above to explain why this is the case. The **Full Token-Reflexive Rule** has it that S’s introspective impressions as of T passing through their mind allows S to latch onto themselves qua recipient-thinker of T. So, if S uses competently the first-person concept, S can’t but refer to themselves qua recipient-thinker of T. From this competent use of ‘I’, S can infer a warrant for the logically weaker existential: ‘Someone is recipient-thinking [p]’. This gives us an explanation of how S could have a warrant for ‘Someone is recipient-thinking [p]’ on the basis of how introspective impressions contribute to determining the reference of ‘I’ in introspection-based self-ascriptions. However, we cannot offer the same explanation for how S’s introspective impressions could afford S a warrant for ‘Someone is author-thinking [p]’, since S’s competent deployment of the first-person concept does not guarantee that S latches onto themselves qua author-thinkers of the relevant thought.

This leaves us with the following: if introspective impressions could afford S a warrant for (Ex), the explanation of this fact would have to be totally disjoint from the explanation of why those very introspective impressions could afford S a warrant for ‘Someone is recipient-thinking [p]’. Note that this would turn introspection into a rather gerrymandered warrant-conferring cognitive faculty, for we would have to subscribe to the following picture: on the one hand, introspection can warrant judgements such as ‘I am recipient-thinking [p]’ and ‘Someone is recipient-thinking [p]’ because introspection allows subjects to think of themselves qua recipient-thinkers of thoughts. On the other hand, introspection can warrant the existential ‘Someone is author-thinking [p]’: the reason why is still unclear, but surely has nothing to do with how the first-person concept refers. This picture of introspection qua warrant-conferring faculty is untenable.

To generalise the point: there is an important explanatory connection between introspection and the referential pattern of the first-person concept. We can appeal to that connection to explain how we can use introspection to warrant recipient-thought ascriptions; however, we cannot appeal to it in order to sustain the possibility of using introspection to warrant author-thought ascriptions. This possibility requires accepting the rather mysterious fact that introspective impressions can warrant S to judge both ‘Someone is author-thinking [p]’ and ‘Someone is recipient-thinking [p]’, but for totally different reasons. This would make introspection an epistemic faculty which confers warrants in an implausibly gerrymandered way. To avoid such an unpalatable result, we’d better take introspective impressions not to provide S with any warrant for (Ex).

Summing up: we cannot explain the idea that existential author-thought ascriptions are warranted by introspective impressions given the fact that those introspective impressions allow S to use ‘I’ to refer to themselves qua recipient-thinker of the relevant thought. By contrast, the metasemantic role played by introspection does enable us to explain existential recipient-thought ascriptions. This supports the following two points: First, since the question of whether introspection-based self-ascriptions of psychological properties are immune or vulnerable to error through misidentification can be sensibly raised only if we grant that the relevant existential generalisation is warranted, and since we can make sense of this possibility only relative to existential *recipient*-thought – as opposed to *author*-thought – ascriptions, we should only focus on whether there is any justificatory gap between the existential ‘Someone is recipient-thinking [p]’ and the singular ‘I am recipient-thinking [p]’. This explains why thought insertion-based scenarios cannot give rise to counterexamples to The Introspective Immunity Thesis. Secondly, the Metasemantic Explanation tells us why the target justificatory gap does not exist: insofar as my introspective impressions warrant me to judge that someone is recipient-thinking [p], such impressions also warrant them to judge that it is I who am recipient-thinking [p], for the concept ‘I’ refers to myself qua recipient-thinker of [p] in virtue of my introspective awareness as of a thought with content [p] passing through my mind.

I have deployed metasemantic considerations in order to explain why existential judgements of the form ‘Someone is author-thinking [p]’ cannot be warranted, thereby explaining why introspection-based self-ascriptions are not vulnerable to error through misidentification. Surely this explanation is not theory-neutral, as it requires taking a stance on how the first-person concept refers. Yet, it’s unclear how much neutrality must be required of an explanation of a given phenomenon. While theoretical responsibility dictates that we offer a neutral description of a phenomenon – bear in mind our previous discussion of the second reading of Hu’s scenario – it’s possible to have different competing explanations of that phenomenon that differ precisely on their commitments to potentially controversial theses. Now that the Metasemantic Explanation has been given a fair chance, further works will have to compare and contrast that explanation with competing ones.

To conclude this section, let us get back to (N1) and (N2). Despite being rationally intelligible, (N1) and (N2) are false – thoughts cannot really be inserted! The falsity of these judgements, however, cannot be explained by in terms of their being instances of error through misidentification: S’s uses of ‘me’ and ‘I’ in (N1) and (N2) have to pick out the recipient-thinker of [the garden looks nice], and they correctly do so in virtue S’s introspective awareness as of the thought [the garden looks nice] passing through their mind. So, the falsity of (N1) and (N2) has to be explained in terms of S’s making an error concerning the concept in the predicate position: S in fact are the author of the thought [the garden looks nice], for thoughts cannot really be inserted.

**6 Conclusion**

In this paper I have investigated the significance of thought insertion for the questions of the extension and explanation of the thesis that introspection-based self-ascriptions of psychological properties are immune to error through misidentification. To address both questions, I have made two preliminary points. First, I have offered a principle, Constraint, specifying the conditions that must be met in order for there to be a counterexample to the Introspective Immunity Thesis. Secondly, I have maintained that self-knowledge, understood here as one’s introspective awareness of the phenomenally conscious nature of one’s own occurrent thought, plays an essential role in fixing the reference of ‘I’.

On these grounds I have undermined three thought insertion-based counterexamples to The Introspective Immunity Thesis. I have then offered a Metasemantic Explanation of The Introspective Immunity Thesis whose key contention is that, if S’s introspective impressions warrant S to judge ‘Someone is thinking [p]’ they warrant S to judge ‘I am thinking [p]’ in virtue of the fact that introspective impressions play an essential role in fixing the reference of ‘I’. The existence of this metasemantic connection also explains why self-ascriptions of inserted thoughts are not vulnerable to error through misidentification.**[[26]](#footnote-26)**

**References**

Anscombe, G. E. M. (1975) “The First Person”, in Guttenplan, S. (ed.) *Mind and Language* (pp. 45-65): Oxford: Clarendon.

Campbell, J. (1994) *Past, Space, and Self*, Oxford: Clarendon.

Campbell, J. (1999) “Schizophrenia, the Space of Reasons, and Thinking as a Motor Process”, *The Monist* 82(4): 609-625.

Campbell, J. (2002) “The Ownership of Thoughts”, *Philosophy, Psychiatry and Psychology*, 9(1): 35-39.

Cappelen, H. and Dever, J. (2013) *The Inessential Indexical*, Oxford: OUP.

Chalmers, D. (2003) “The Content and Epistemology of Phenomenal Belief”, in Q. Smith and A. Jokic (eds.) *Consciousness: New Philosophical Perspectives*, Oxford: OUP.

Chudnoff, E. (2015) *Cognitive Phenomenology*, London: Routledge.

Coliva, A. (2002a) “Thought Insertion and Immunity to Error through Misidentification”, *Philosophy, Psychology and Psychiatry* 9(1): 27-34.

Coliva, A. (2002b) “On What There Really Is to Our Notion of Ownership of a Thought”, *Philosophy, Psychology and Psychiatry* 9(1): 41-46.

Coliva, A. (2006) “Error Through Misidentification: Some Varieties”, *Journal of Philosophy* 103(8): 403-425.

Coliva, A. (2015) *Extended Rationality: A Hinge Epistemology*, London: Palgrave.

Coliva, A. (2018) “Review of ‘About Oneself: De Se Thought and Communication”, *Analysis* 78(4): 780-784.

de Vignemont, F. (2012) “Bodily Immunity to Error”, in Prosser, S. and Recanati, F. (eds.), pp. 224-246.

Evans, G. (1982) *The Varieties of Reference*, Oxford: Clarendon.

García-Carpintero, M. (2017) “The Philosophical Significance of the De Se”, *Inquiry* 60(3): 253-276.

García-Carpintero, M. (2018) “*De Se* Thoughts and Immunity to Error Through Misidentification”, *Synthese* 195: 3311-3333.

Gertler, B. (2011) *Self-Knowledge*, London: Routledge.

Gertler, B. (2012) “Renewed Acquaintance”, in D. Smithies and D. Stoljar (eds.), *Introspection and Consciosuness* (pp. 93-127), Oxford: Oxford University Press.

Gettier, E. (1963) “Is Justified True Belief Knowledge?”, *Analysis* 23(6): 121-123.

Horgan, T. and Kriegel, U. (2007) “Phenomenal Epistemology: What Is Consciousness that We May Know It So Well?”, *Philosophical Issues* 17: 123-144.

Hu, I. (2017) “The Epistemology of Immunity to Error Through Misidentification”, *The Journal of Philosophy* 114(3): 113-133.

Kriegel, U. (2009) *Subjective Consciousness: A Self-Representational Theory*, Oxford: Oxford University Press.

Kripke, S. (1980) *Naming and Necessity*, Cambridge MA: Harvard University Press.

Langland-Hassan, P. (2015) “Introspective Misidentification”, *Philosophical Studies* 172: 1737-1758.

McGlynn, A. (2015) “Immunity to Error Through Misidentification and the Epistemology of *De Se* Thought”, in García-Carpintero, M. and Torre, S. (eds.) *About Oneself* (pp. 25-55) Oxford: OUP.

Mellor C.S. (1970) “First Rank Symptoms of Schizophrenia,” *The British Journal of Psychiatry* 117(536): 15-23.

Morgan, D. (2015) “The Demonstrative Model of first-person thought”, *Philosophical Studies* 172: 1795-1811.

O’Brien, L. (2007) *Self-Knowing Agents*, Oxford: OUP.

Palmira, M. (2019) “Arithmetic Judgements, First-Person Judgements, and Immunity to Error Through Misidentification”, *Review of Philosophy and Psychology* 10(1): 155-172.

Peacocke, C. (2008) *Truly Understood*, Oxford: OUP.

Peacocke, C. (2014) *The Mirror of the World*, Oxford: OUP.

Perry, J. (1979) “The Problem of the Essential Indexical”, *Noûs* 13: 3-21.

Prosser, S. and Recanati, F. (eds., 2012) *Immunity to Error Through Misidentification. New Essays*, Cambridge: CUP.

Pryor, J. (1999) “Immunity to Error Through Misidentification”, *Philosophical Topics*, 26: 271-304.

Pryor, J. (2000) “The Skeptic and the Dogmatist”, *Noûs* 34: 517-549.

Pryor, J. (2004) “What’s Wrong with Moore’s Argument?”, *Philosophical Issues* 14(1): 349-378.

Recanati, F. (2007) *Perspectival Thought. A Plea for Moderate Relativism*, Oxford: OUP.

Recanati, F. (2012) “Immunity to Error Through Misidentification: What It Is and Where It Comes From”, in Prosser, S. and Recanati, F. (eds.), pp. 180-201.

Salje, L. (2016) “The Subjective Perspective in Introspection”, *Journal of Consciousness Studies* 23 (3-4): 128-145.

Shoemaker, S. (1968) “Self-Reference and Self-Awareness”, *The Journal of Philosophy*, 65: 555-567.

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: 249-315.

Smith, J. (2006) “Which Immunity to Error?”, *Philosophical Studies* 130(2), 273-283.

Smithies, D. (2013) “The Nature of Cognitive Phenomenology”, *Philosophy Compass* 8(8): 744-754.

Wittgenstein, L. (1958) *The Blue and Brown Books*, New York: Harper and Row.

Wright, C. (1987) “On Making Up One’s Mind: Wittgenstein on intention”, reprinted in Wright, C. (2001), *Rails to Infinity*, pp. 116-142, Cambridge MA: Harvard University Press.

Wright, C. (2004) “On Epistemic Entitlement: Warrant for Nothing (And Foundations for Free?)”, *Aristotelian Society Suppl. Vol.* 78: 167-212.

Wright, C. (2007) “Perils of Dogmatism”, In Nuccetelli, S. (ed.), *Themes from G. E. Moore: New Essays in Epistemology and Ethics*, pp. 25-48 Oxford: Oxford University Press.

Wright, C. (2012a) “Reflections on François Recanati’s ‘Immunity to Error Through Misidentification: What It Is and Where It Comes From’”, in Prosser, S. and Recanati, F. (eds.), pp. 247-280.

Wright, C. (2012b) “Replies Part IV: Warrant Transmission and Entitlement’, in Coliva, A. (ed.), *Mind, Meaning, and Knowledge: Themes from the Philosophy of Crispin Wright*, pp. 451-86, Oxford: Oxford University Press.

Wright, C. (2014) “On Epistemic Entitlement (II): Welfare State Epistemology”, in Dodd, D. and Zardini, E. (eds.) *Scepticism and Perceptual Justification* (pp. 213-247) Oxford: OUP.

1. A notable – yet partial – exception is Hu (2017), which will be used as a foil at various stages of my discussion. The exception is “partial” since Hu devotes only one page to the construction of a thought insertion-based counterexample to The Introspective Immunity Thesis and does not explicitly engage with the question of what features a thought insertion scenario should have in order to appropriately challenge The Introspective Immunity Thesis. Arguably, Hu’s main focus is to carefully discuss various cases of error through misidentification in order to come up with a more precise definition of the phenomenon. [↑](#footnote-ref-1)
2. To forestall misunderstandings: I am not claiming that no attempt has been made to situate thought insertion within a more general account of introspection and psychological explanation; in fact that is the main project pursued in Campbell (1999). I am instead alerting the reader to the fact that those philosophers putting forward explanations of The Introspective Immunity Thesis have not paid enough attention to the question of whether or not certain features of thought insertion place constraints on an explanation of The Introspective Immunity Thesis. I will offer concrete examples below. [↑](#footnote-ref-2)
3. I refer the reader to Coliva (2006), (2018), García-Carpintero (2018), Hu (2017), McGlynn (2015), Palmira (2019) and Wright (2012a). [↑](#footnote-ref-3)
4. In agreement with the literature (see e.g. Coliva 2006, García-Carpintero 2018, Wright 2012a), I take these to be *singular* – as opposed to *descriptive* – concepts. [↑](#footnote-ref-4)
5. I use the expression “warrant” in the standard epistemological sense: epistemically warranted judgements have the distinctive value or quality of bearing a strong connection to the truth of the judged content. I will discuss different theories of warrant below. [↑](#footnote-ref-5)
6. This formulation captures, I believe, the core of the phenomenon without going through the rather complex characterisation offered by Pryor (1999). See García-Carpintero (2018), Hu (2017) and McGlynn (2015) for critical discussions of Pryor’s own formulation, and for other attempts at simplifying it. [↑](#footnote-ref-6)
7. For further discussion, see McGlynn (2015: 41), García-Carpintero (2018: 3320) and Pryor (1999: 284). [↑](#footnote-ref-7)
8. I shall henceforth use [] to refer to propositional contents of thoughts. [↑](#footnote-ref-8)
9. Some authors (e.g. de Vignemont 2012, Langland-Hassan 2015) maintain that thought insertion presents no counterexample to The Introspective Immunity Thesis since it merely shows the possibility of a false negative, as opposed to a false positive error. Since thought insertion-based counterexamples should be like (PARK), and since (PARK) is a genuine case of error through misidentification, this way of downplaying the significance of thought insertion is off target. [↑](#footnote-ref-9)
10. We’ll need (N2) since not all cases of thought insertion are such that the deluded subjects do ascribe authorship of thought to anyone in particular. [↑](#footnote-ref-10)
11. I use “impression” as synonymous with “experience” and “appearance”. [↑](#footnote-ref-11)
12. Note that both conservatism and liberalism enable Hu’s revised scenario to meet Preservation of Grounds. If we asked S to reconstruct their grounds for (N1) and (N2), it’d be really surprising to hear them citing their belief [thoughts have authors], for it’d be much more likely that they cite their introspective impressions only. This gives us reason to think that S’s grounds for (N1) and (N2) are introspective. However, if conservatism is true, S’s introspective impressions can afford S a warrant for (N1) and (N2) only if S is entitled to trust the background presupposition [thoughts have authors]. [↑](#footnote-ref-12)
13. Wright (2014) tentatively recants his early take on the transmission failure; yet it’s unclear whether the view he ends up with really coheres with conservatism as is presented in Wright (2004, 2007, 2012b). So, I will stick to the original version of conservatism. [↑](#footnote-ref-13)
14. Two clarifications are needed. First, one might observe that (ii) is more controversial than (II). However, I grant (ii) for the sake of argument. Secondly, one might observe that in order for S to have a warrant for (III), S must be introspectively aware of T. From this fact, one might derive that introspective impressions do contribute, together with S’s entitlement, to warranting (III). Surely without accessing T, S couldn’t have any warrant for (iii), but this doesn’t show that S’s introspective access to T also warrants (iii). Rather, the conservative theorist will maintain that S’s introspective access to T is a merely *enabling*, as opposed to *warranting* (or justifying), condition for S to be warranted – by default and by default only! – to trust (iii). I thank an anonymous referee for this journal for urging me to clarify both points. [↑](#footnote-ref-14)
15. In fact, the situation is worse than that, for a transmission-failure problem arises even if we adopt Coliva’s (2015) *moderate* view of perceptual warrant, which is a genuine alternative to both liberalism and conservatism. On Coliva’s view, in order to be perceptually warranted to believe that *p*, it’s enough to have a certain course of experience with content *p* together with the *assumption* of a background presupposition such as there is an external world, absent defeaters. On such a view, the warrant for [here is a hand] is also bound to fail to transmit to the background presupposition [there is an external world]. The same holds, *mutatis mutandis*, for a moderate account of introspective warrant. This spells further trouble for the theoretical neutrality of Hu’s scenario. [↑](#footnote-ref-15)
16. In what follows I expand on some remarks made by García-Carpintero (2017: 254-5). [↑](#footnote-ref-16)
17. An anonymous referee pointed out that the foregoing discussion has been running two distinct questions together. The questions are: “Do self-ascriptions of inserted thoughts involve an error through misidentification?”; “Do self-ascriptions of inserted thoughts falsify The Introspective Immunity Thesis?”. While I agree that these are two distinct questions, I believe that this is a distinction without a difference for the purposes of this paper. If the arguments of this section are sound, (N1) and (N2) involve an error through misidentification only if liberalism is true, and this should lead us to revoke the status of counterexample to Hu’s scenario. So, one could answer the former question in the affirmative while, at the same time, answering the latter question in the negative. However, given that one of the two aims of this paper is precisely to establish whether The Introspective Immunity Thesis is falsified by self-ascriptions of inserted thoughts, claiming that (N1) and (N2) involve an error through misidentification provided that liberalism is true wouldn’t make any difference as to the question of whether The Introspective Immunity Thesis is falsified; the only claim that would be falsified is The Introspective Immunity Thesis + liberalism. But, again, that’s not the target of this paper, nor – arguably – is the target of the whole immunity debate. [↑](#footnote-ref-17)
18. See Wright (2012a) for a deflationary account and Recanati (2007) (2012) for a content-based account of immunity. See García-Carpintero (2018) and McGlynn (2015) for criticism of the former, and García-Carpintero (2018), McGlynn (2015) and Wright (2012a) for criticism of the latter. [↑](#footnote-ref-18)
19. Some examples: García-Carpintero (2018) and Hu (2017) maintain that thought insertion presents a counterexample to The Introspective Immunity Thesis, but in the first part of the paper I have argued against this. Peacocke (2014) seems to think that the standard formulation of the token-reflexive rule for ‘I’ will be enough to produce a Metasemantic Explanation of The Introspective Immunity Thesis; yet, as I will argue below, the standard formulation of the token-reflexive rule suffers from two neglected indeterminacy problems which make it unsuitable to account for The Introspective Immunity Thesis. Since, as I will demonstrate below, one of the indeterminacy problems stems from the possibility of distinguishing between thinkers-qua-authors and thinkers-qua-recipients deployed by Campbell to describe the phenomenon of thought insertion, this gives us further reason to think that thought insertion is relevant the project of capturing the pattern of reference of the first-person concept. [↑](#footnote-ref-19)
20. See, among others, Campbell (1994), García-Carpintero (2018), O’Brien (2007), Peacocke (2014). [↑](#footnote-ref-20)
21. García-Carpintero (2018) also claims that introspection is crucial for fixing the reference of ‘I’. [↑](#footnote-ref-21)
22. I shall follow Gertler (2012)’s insightful presentation of acquaintance views. However, other philosophers, e.g. Chalmers (2003), Horgan and Kriegel (2007), have argued for similar points. I should also note that some of the authors discussing The Introspective Immunity Thesis favourably look at the constitution thesis under scrutiny: García-Carpintero (2018) endorses the acquaintance-based development of it offered by Gertler, whereas Peacocke (2014), Shoemaker (1994) and Wright (1987) subscribe to the constitution claim in the context of different theories of self-knowledge. This, to my mind, shows that assumption (3) does not beg the question against many participants to the immunity debate. [↑](#footnote-ref-22)
23. For present purposes it does not make a difference whether it is a phenomenal *property* or a phenomenal *state* which gets introspected. Henceforth I will ignore this distinction. [↑](#footnote-ref-23)
24. Even if we accept the thesis that having a phenomenally conscious state entails being aware of it, this might still not suffice for mentally demonstrating the mental state in question. Since one may be pre-reflectively or peripherally aware of one’s conscious mental state without thereby attending to it (see Kriegel 2009: Ch. 5) and since, plausibly, attention plays a key role in fixing the reference of demonstratives (see Campbell 2002), simple awareness of a thought may not suffice for latching onto it. This is why introspective impressions play a crucial role in fixing the reference of the demonstrative ‘this’ featuring in the token-reflexive rule for the first-person concept, for they are needed in order for there to be a mental demonstration to the occurrent thought S is presently thinking. I’m very much indebted to Anna Giustina for discussion of this point. [↑](#footnote-ref-24)
25. See Morgan (2015) for an insightful discussion of some of these issues. [↑](#footnote-ref-25)
26. **Acknowledgements.** I wish to thank audiences at the 2018 LOGOS Seminar and the 2018 LOGOS *DeSe* Reading Group in Barcelona, the Diaphora *Self-Knowledge* workshop in Paris, the SLMFCE IX conference in Madrid, as well as an anonymous referee for this journal for valuable feedback on various parts of this material. I am particularly grateful to Manuel García-Carpintero, Anna Giustina, Daniel Morgan, Hichem Naar, Léa Salje, and Carlota Serrahima for providing detailed comments on previous drafts of this paper. Extra special thanks are due to the messy shoppers, Carlota and Daniel. [↑](#footnote-ref-26)