Immunity to error through misidentification, introspection and thought insertion*

Léa Salje (University of Leeds)

Abstract

As an empirical example of introspective conditions in which the normal sense of self is disrupted, the delusion of thought insertion is of special interest to philosophers investigating the epistemic and phenomenological structures of introspection. A common strategy is to use immunity to error through misidentification as a tool with which to pick apart the implications of thought insertion for our understanding of the faculty of introspection. In this paper I turn that strategy on its head: I draw on our understanding of introspection and of thought insertion to make two correctives to the literature on immunity to error through misidentification. The first is the identification of a formal distinction between two phenomena sometimes conflated under the rubric of misidentification errors. The second is a weakening of the presumed significance of claims to immunity to error through misidentification. With these tightenings to the notion of immunity to error through misidentification in hand, the idea is, we will be in a better position to turn again to questions about the epistemic and phenomenological nature of introspection.

Self-ascriptive judgments that have been formed by introspecting the current state of one's own mental activities are surely immune to errors through misidentification, if any are. I will say that a judgment (say, the judgment that *a is F*), made on the basis of certain grounds (say, grounds *G*), is immune to error through misidentification relative to the use of a singular concept contained in that judgment (*a*) just in case it is not possible that the subject could know by *G* that the property of *F*-ness is instantiated, but make a mistake solely in virtue of misidentifying the referent of *a* as the thing that instantiates it — this is what I will call the *modal formulation* of immunity to error through

*Final draft, forthcoming in *Journal of Consciousness Studies*

misidentification. Put in somewhat friendlier terms, the point is that it seems to be out of the question that through introspection I could come to a correct opinion about *what* I am thinking, but go wrong only on the matter of whether *I* am the one thinking it.

When philosophers claim that a given form of self-knowledge issues in first person judgments with immunity to error through misidentification, it is normally in the service of showing that that there is something special about that way of coming to know about ourselves. What is special is not merely that it keeps us safe from making this particular kind of mistake — *that* doesn't seem to be of any very great interest in itself. What is special about the forms of self-knowledge marked out by this epistemic feature, rather, is that it reveals them to be a way of knowing about oneself that is unmediated through any kind of identification; an identification, after all, would have brought with it the possibility of a *mis*identification. But if it is a way of knowing about oneself that does not involve identifying the source of one's knowledge as oneself, then it must be a form of self-knowledge that already and directly presents its object as oneself. It is a form of self-knowledge that unmediately delivers up the subject to herself *as* herself.[1]

This way of understanding the significance of claims to immunity to error through misidentification has firm roots in the origins of the notion in Wittgenstein's blue and brown books, where he distinguished between two uses of the word 'I': the *use as object*, uses which 'involve the recognition of a particular person' and so in which '[t]he possibility of an error has been provided for', and the *use as subject* in which there is no such recognition, and so no such room for error. (Wittgenstein 1958, p. 67) A natural extension of this way of putting things, and one that continues to hold grounds between many commentators on immunity to error through misidentification, is to say that the impossibility of such an error marks out a peculiarly subjective perspective on oneself — it characterises forms of self-knowledge through which a subject is presented to herself *as* herself, not one through which she discovers herself as just another object encountered in the world.[2]

It would perhaps be something of an understatement to say that questions about the correct formulation of immunity to error through misidentification, and about the range of forms of self-knowledge that are marked by it, remain largely unsettled in literature. One commitment that survives these debates more or less untouched, however, is that if there are any forms of self-knowledge that give rise to judgments that are immune to

[1] I should say that the significance is *at least* that the judgment did not involve an identification in its formative grounds. Some commentators, such as Analisa Coliva and Crispin Wright, have argued that it also signifies a lack of identificatory beliefs in the judgment's backgrounds presuppositions; see (Coliva 2006), (Wright 2012). Nothing in this paper requires that stronger view, so I put it to one side.

[2] See, for instance, (Morgan 2015), (Chen 2011), (Cassam 1997), (Hamilton 2009), (Recanati 2012)

error through misidentification relative to uses of the first person, then introspection is one of them — however the notion of immunity to error through misidentification is to be best untangled, introspective self-ascriptive judgments are a paradigm of a judgment-kind that has it. Indeed, for some this serves as a constraint on the adequacy of formulations of immunity to error through misidentification itself; we must reject out of hand any account on which such judgments turn out to lack the immunity.[3] Whether or not we go that far, it certainly doesn't seem like something to be given up too easily: our introspective awareness of our own mental lives surely gives us a uniquely subjective perspective on ourselves if anything does.

It is sometimes suggested that this assumed point of solid ground — the claim that introspective self-ascriptions are immune to error through misidentification relative to uses of the first person — begins to look rather less solid in light of some of the delusional symptoms associated with schizophrenia. Schizophrenic patients often report a disintegrating sense of the boundaries of the self, sometimes manifested in delusions of control expanding outwardly beyond their own mental and bodily activities, and sometimes in delusions characterised by the converse belief, that an external locus of agency or control is causally intervening in states and events taking place within the limits of their own bodies and minds. In particular, it is sometimes suggested that there is a tension between the above epistemic claim about introspection and *thought insertion*, a delusion currently counted among the key diagnostic symptoms of schizophrenia that some of the thoughts occurring in the subject's introspectively accessible stream of consciousness are not the subject's own. *Prima facie*, the delusion of thought insertion calls into question the idea that introspection is a form of self-knowledge that delivers up a subject to herself as herself, that introspection is a form of self-knowledge that grounds judgments with immunity to misidentification errors. It is precisely a process of person-recognition that seems to be disrupted in these cases, the very kind of recognition whose presence is denied by the claim that introspective judgments are immune to errors of misidentification.

I think that there is both something straightforwardly wrong and something interestingly right in this *prima facie* challenge, both of which are tied up with our understanding of the notion of immunity to error through misidentification. The aim of this paper is to bring them both out, and in doing so to find ways of making that notion more precise. In §2 I present two versions of the challenge, corresponding to two ways of formulating the key notion of immunity to error through misidentification. Under the standard modal formulation given above, I show that the sorts of judgments formed by sufferers

[3] See (Smith 2006) and (Langland-Hassan 2015).

of thought insertion are of the wrong shape to be counterexamples to the claim that introspection grounds judgments with immunity to error through misidentification relative to uses of the first person concept. This resolution to the apparent challenge is blocked off under an alternative, so-called *constitutive*, formulation of the immunity, but I argue that the significance of this fact differs importantly from the significance normally attached to claims of vulnerability or immunity to error through misidentification.

Two findings are contributed to the literature on immunity to error through misidentification and introspection in §3. The first is a corrective about keeping apart these importantly different strands in careful formulations of immunity to error through misidentification. The second is the idea that there is a more important lesson to learn from cases of thought insertion than their attempted use as counterexamples: the fact that introspection is marked by the immunity to error through misidentification of its self-directed judgments does not rule it out as a form of self-knowledge that can also fail in the sorts of errors associated with the constitutive formulation. This means that the significance of claims to immunity to error through misidentification must be qualified. Where a form of self-knowledge is marked with this epistemic feature it is one through which a subject is given to herself *as herself*, but — and this is the qualification — that does not mean that such a mode of being given to oneself is immune to additive interference of the kind dramatised by the delusion of thought insertion.

First, though, I will introduce the delusion itself in more detail.

1 Thought insertion

As one of Schneider's first-rank symptoms, thought insertion is symptomatically sufficient under most diagnostic systems for a diagnosis of schizophrenia. It is normally classified as a delusion (rather than a hallucination), or a firmly held belief that is resistant to rational influence and that cannot be accounted for by appeal to the subject's cultural or religious context. While there is broad variation in the positive confabulatory component of thought insertion reports — how the patients explain the way in which the thoughts came to be there — what unifies them is the common conviction that some of the thoughts occurring in their minds are in some sense not their own, that they are someone (or something) else's thoughts that are merely happening in their minds. A pair of examples cited in a 2001 paper by Christophe Hoerl draw out the reported phenomenology of dissociation from the thought underlying this negative element of the delusion:

> [H]e said, " ... it's like a thought as it comes in ... a thought is very light really,

> inspirational ... it's a light feeling where you feel as though I'm actually thinking it ... or you're receiving it rather ... it's just a thought but it feels logical say ... it feels pretty normal or fits with what I suspect, [I] wonder if that's me ... it felt like a piece of information."
>
> Later he went on: "... you find it strange when some different little thought filters through ... why did I think that at this time of day?" He said you judge it and say, "I don't think that was mine...you can differentiate".
>
> (Allison-Bolger 1999, case 68, cited in Hoerl 2001)

> [S]he said that sometimes it seemed to be her own thought "... but I don't get the feeling that it is." She said her "own thoughts might say the same thing ... but the feeling isn't the same ... the feeling is that it is somebody else's ..."
>
> She was asked if she had other people's thoughts put inside her head. She said "... possibly they are but I don't think of them in that way... they were being put into me into my mind ... very similar to what I would be like normally".
>
> (Allison-Bolger 1999, case 89, cited in Hoerl 2001)

The delusion characteristically also involves a positive confabulatory narrative about the thoughts' source:

> I look out of the window and I think the garden looks nice and the grass looks cool, but the thoughts of [radio personality] 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. (Mellor 1970, p. 17)

> Thoughts come into my head like "Kill God." It's just like my mind working, but it isn't. They come from this chap, Chris. They're his thoughts. (Frith 1992, p. 66)

> [T]he words just came into my head — they were ideas I was having. Yet I instinctively knew they were not my ideas. They belonged to the houses, and the houses had put them in my head. (Saks 2007, p. 29)

The content of the central belief that threads its way throughout these reports is crucially not just that a foreign body of control is influencing the content of the subject's thoughts, or making her think thoughts that she wouldn't have had on her own. The content of the delusional belief is considerably stronger than that; even while they are occurring in the subject's mind, there is a sense in which the subject *does not take these thoughts to be her own*.

According to the dominant view, due in its original version to Frith but taken up in influential form by John Campbell, thought insertion is a disorder of agency. The best way to make sense of these reports, the idea is, is that the subject continues to take the thoughts to be occurring within the limits of her own mind, but does not believe them to be thoughts that she herself *is thinking*. The content of the delusion, then, is that these

thoughts are the subject's own in a posessive or locative sense, but not in an authorship or agentive sense. Competitors to this leading account include what we might call *ill-fittedness* accounts, on which the inserted thoughts fail to properly integrate into the subject's standing mental economy; *endorsement* accounts, on which what matters is that the subject does not endorse the contents of these thoughts as she does the contents of her ordinary thoughts; and Jordi Fernández's recent *self-knowledge* account, on which the deficiency is in the first instance an epistemic one — the subjects are unable to find out about the contents of these beliefs in the normal way by looking outwardly at the state of the affairs with which the belief is concerned because of a hyper-reflexive tendency to focus on the experiences themselves rather than their objects. On Fernández's view, it is the resulting feeling of abnormality that underlies the delusion of thought insertion.[4]

Over the next three sections I put aside the question of how best to make sense of the phenomenological-psychological factors that generate these reports. I focus instead on the reports themselves and what they mean for the claim that introspection is a form of self-knowledge that gives rise to judgments with immunity to error through misidentification. More specifically, I focus for clarity on an idealised abstraction from these reports:

(*S*) Someone is thinking thought *p*, but it isn't me.

2 The challenges from thought insertion

Let's say that *S* is a judgment formed on the basis of introspection by a schizophrenic patient suffering from thought insertion. There might seem to be a challenge to the epistemic status of introspection here of the following kind. *S* is, of course, false. This means that its judger must have made a mistake. What's more, the mistake she has made seems to be one of misidentification; she has misidentified the person undergoing the mental occurrence (that is, herself) with someone else (the thought's attributed source), and as a result is wrong about *who* is thinking the thought, even if she is right about what the thought is. This mistake is important because it shows that there was an identification involved in the judgment's formative grounds — after all, the subject can't come to misidentify the thought's thinker if her judgment on the matter didn't go via an identification in the first place. But given that the judgment was arrived at on the basis of introspection, what this means is that introspection is not an identification-free

[4]For an example of the *ill-fittedness account* see (Graham and Stephens 2000), for the *endorsement account* see (Bortolotti and Broome 2009), for the last account see (Fernández 2010). For Campbell on thought insertion see (Campbell 1999) and (Campbell 2002)

method of judgment-formation. If that's right, then introspection cannot be a form of self-knowledge that issues in judgments that are immune to errors of misidentification after all. Or so the challenge from thought insertion might seem to go.

This challenge is not very deep, but seeing why this is so will be helpful in drawing out a rather more interesting implication of thought insertion cases for claims about the immunity to error through misidentification of introspective judgments. Consider the distinction made by Fréderique de Vignemont between what she calls *false positive* and *false negative* errors:

> There is a *false negative* if one does not self-ascribe properties that are instantiated by [oneself]. False-negative errors have to be contrasted from false-positive errors. There is a *false-positive* if one self-ascribes properties that are instantiated by another. (de Vignemont 2012, p. 229) [5]

Claims about immunity to error through misidentification are concerned with the impossibility of grounds-relative false positive errors, at least on its modal formulation given in the introduction. More specifically, to say that a particular class of self-ascriptive judgments, made on certain grounds, is immune to error through misidentification is to say the following: it is impossible to self-ascribe a property on those grounds when the property instantiation thereby known about is in fact by someone other than oneself. This is to deny the possibility under those conditions of a false positive error of a particular kind.

The claim with which this paper started, that introspection is a form of self-knowledge that gives rise to judgments with immunity to error through misidentification relative to uses of the first person, is just such a claim about the impossibilty of false positive errors under introspection-involving conditions. It says that it is not possible that a subject could form a self-ascriptive judgment on the basis of introspection, and in so doing self-ascribe a property that is in fact instantiated — and known on the basis of introspection to be instantiated — by someone else. It denies the possibility of an introspection-based false positive error.

From here it is no great stretch to see why cases of thought insertion do not constitute a counterexample to the claim that introspection is a form of self-knowledge giving rise

[5] As I am using de Vignemont's notion of a *false positive error* here there is no ambition to provide a general purpose definition of a kind of error that might be extended to discussions in other areas. A more minimal, and perhaps more common, understanding of a false positive error is one in which one self-ascribes a property that one does not in fact instantiate (rather than one that is in fact instantiated by another). This would not be strong enough for present purposes, since that would only be an error of misascription, not one of misidentification. I take this quote to be definitional of the particular kind of error we are interested in for the present discussion, and not the only way to understand the notion of a false positive error. Thanks to an anonymous reviewer for pressing me on this point.

to judgments with immunity to error through misidentification. What cases of thought insertion demonstrate is the possibility of introspection-based false *negative* errors; on the basis of introspection, these subjects fail to self-ascribe a property that they are in fact instantiating. Showing introspection to be a form of self-knowledge that can give rise to judgments that are defective in this negative way, however, does nothing to threaten the claim that one couldn't make a false positive error of the relevant kind on its basis. Even granting that the mistake being made in judgments like *S* is one of a misidentification, then, there is no counterexample here to the claim that our introspection-based self-ascriptions of psychological properties are immune to error through misidentification.[6]

Perhaps, though, there is another way to press the challenge. Even if the identification-errors involved in cases of thought insertion are strictly *compatible* with the claimed immunity to error through misidentification of introspection-based self-ascriptions, don't the former nonetheless give us reason to discredit the latter thesis? After all, the delusion of thought insertion still shows us that, at least for these schizophrenic subjects, the faculty of introspection is identification-involving. Doesn't that give us reason to think that *all* of our introspection-based self-ascriptions have an identification in their formative structure, albeit one that never normally goes wrong in healthy subjects?

This objection is quickly dealt with. To see why it will not take us very far, notice that it rests on an unargued assumption of structural homogeneity between ordinary introspective judgments and introspection-involving judgments made by sufferers of thought insertion. But we have no reason to accept this assumption. Indeed, if it is right that normal self-ascriptive introspective judgments are identification free, then we should only *expect* that other-ascriptive introspection-involving judgments will incorporate an identification into their formative grounds — an identification, that is, between the object about which the introspective grounds provide immediate warrant to make an ascription of and the pronounced object of the subject's judgment. That the formative grounds of a judgment like *S* is identification-involving, then, gives us no reason to project the same structure onto the ordinary workings of introspection.

The resolution just offered to the *prima facie* tension between the delusion of thought insertion on the one hand, and the claimed immunity to error through misidentification relative to uses of the first person concept of introspection-based judgments on the other, rests on a standard modal characterisation of immunity to error through misidentification as given in the introduction: there is no threat so long as we understand immunity

[6] See (Coliva 2002) and (Langland-Hassan 2015) for similar responses to this first version of the thought insertion challenge.

to error through misidentification as the impossibility that the subject could know by certain grounds that a certain property is instantiated, but make a mistake solely in virtue of misidentifying the referent of the relevant concept as the thing that instantiates it. The literature on immunity to error through misidentification, however, is characterised by something of wealth of formulations of its central notion, and the acceptability of this resolution depends crucially on which formulation one takes on. In particular there is one such formulation that I want to consider under which this compatabilist resolution to the challenge seems to be blocked off. This is what we might call the *constitutive* formulation of immunity to error through misidentification. Under it, a first personal judgment is immune to error through misidentification relative to its first personal component just in case it is formed on the basis of knowing a property to be instantatiated on certain grounds, where those grounds ensure that general knowledge of property-instantiation is simultaneously constitutive of singular knowledge that it is instantiated *in oneself*; there is no space between knowing that the property is instantiated and knowing it to be self-instantiated.

This formulation has an impressive pedigree. Sydney Shoemaker, for instance, held what might be called a tautological model of introspection on which awareness of a mental event, like a pain, *just is* awareness of pain in oneself, and it is because of this that self-ascriptions of mental properties display immunity to error through misidentification.[7] Likewise, but more broadly, Gareth Evans asserts that we have various information channels to ourselves (including introspection) for which there is 'no gap' between knowing a property to be instantiated and knowing it to be instantiated in oneself. The 'gapless' nature of these information channels explains the identification-freedom, and so immunity to error through misidentification, of first person judgments arising from those channels.[8] For both writers, the judgments are immune to error through misidentification because the properties are known about in such a way that existential knowledge of property instantiation amounts to singular knowledge of its instantiation in the relevant object — oneself. A more recent example of this way of putting things comes from Beatrice Longuenesse, who writes, '...given the kind of information these judgments are based on, knowing, on the basis of that information, the predicate to be true of anyone at all *just is knowing it to be true of oneself*' (Longuenesse 2012, p.83, emphasis added). On the constitutive characterisation, a judgment has immunity to error through misidentification relative to a use of the first person concept just in case existential knowledge of property instantiation, via the given grounds, amounts to knowledge

[7](Shoemaker 1968, pp. 563-4).
[8]See (Evans 1982, p. 180)

of the property's instantiation in oneself.[9]

The constitutive characterisation of immunity to error through misidentification resurrects the threat from thought insertion to the claim that introspection-based judgments are immune to error through misidentification relative to uses of the first person concept. That is because under the constitutive characterisation that claim becomes the following: that when one has existential knowledge of the instantiation of mental properties on introspective grounds, that is constitutive of knowledge that those properties are instantiated in oneself. But this, as we have seen, is precisely the step that is *not* made in the case of thought insertion. In those cases the subject has existential knowledge of a mental property instantiation on the basis of introspection, but she fails to recognise that is is *she herself* in whom it is instantiated. So long as we understand the notion of immunity to error through misidentification on its constitutive reading, then, cases of thought insertion once again present themselves as counterexamples to the claim that introspection-based judgments are immune to error through misidentification relative to uses of the first person concept. They show that introspection is a way of knowing about mental properties in which — *pace* Shoemaker and Evans — there *is* a gap between knowing them to be instantiated and knowing them to be instantiated in oneself.

This revived challenge should give us reason to be suspicious of the constitutive formulation of immunity to error through misidentification. Even if we don't want to go so far as to treat it as a constraint on adequate accounts of immunity to error through misidentification, we have seen that the immunity of introspective self-ascriptive judgments is a core commitment for many writers in this area. At the very least, then, we should take care not to simply define it away.

In response to the original version of the challenge from thought insertion I drew on de Vignemont's distinction between false positive and false negative errors, and pressed for an understanding of immunity to error through misidentification as a notion concerned with the impossibility of false positive errors. To say that a given judgment, made on certain grounds, is immune to error through misidentification relative to the use of a given concept is to say that it is not possible to form a judgment on those grounds, and in so doing ascribe to that object a property that is in fact instantiated — and known on the same grounds to be instantiated — by something (or someone) other than that object. In the case of a first personal judgment, it is to say that it is impossi-

[9]This constitutive formulation also underlies the widely used test for immunity to error through misidentification that asks whether it makes sense to ask, on the relevant grounds, 'someone is F, but is it me?'. A negative response to this question implies that it would not be possible to know that someone is F on those grounds but not to know that it is me, a characterisation of the immunity that slots into the constitutive formulation just given.

ble to form a self-ascriptive judgment on those grounds while being mistaken through a misidentification in having self- rather than other-ascribed the property. The significance of making such a claim is that it shows that there can have been no identification involved in the formation of the judgment. The subject had immediate, non-identification involving grounds for a self-ascription.

It is obviously a further step to say not only that the subject had immediate, non-identification involving grounds for a self-ascription, but moreover, that she was in a position such that she could not fail to exploit those grounds and form a first personal judgment. That would be a much stronger claim, and one concerned with the impossibility of a false *negative* rather than a false positive error; it says that where the subject is in possession of those grounds, it is impossible that she could fail to self-ascribe a property that she in fact instantiates. Recall, however, that under the constitutive formulation of immunity to error through misidentification, the claim that introspective self-ascriptions are immune to error through misidentification relative to their first personal components becomes the claim that it is impossible, once in possession of existential knowledge of a property instantiation through introspection, to fail to know that the property is *self*-instantiated — that's to say that the stronger claim about the impossibility of introspection-based false negative errors is precisely what is at issue under the constitutive formulation of immunity to error through misidentification. Under the constitutive formulation, then, the notion of immunity to error through misidentification is concerned with false negative, rather than false positive errors.

This is more than just a labelling problem. Of course, 'immunity to error through misidentification' is a philosopher's term of art; equipped with the right definitions, we can use it however we like. Perhaps fans of the constitutive formulation will insist that theirs is the more philosophically important definition, and that the recognition that there have been two independent epistemic phenomena run together under the same label should prompt a turn away from the standard modal notion and towards the constitutive one.

The problem with this move, however, is that it seems to change the subject; it changes the significance of claims about the immunity to error through misidentification of given forms of self-knowledge. To see this, notice that with the modal notion in hand, the bare fact (if it was one) that introspection is a form of self-knowledge that issues in judgments that are subject to errors of misidentification relative to uses of the first person concept would be enough by itself to force a radical revision of our understanding of the faculty of introspection. It would show that introspection is a form of self-knowledge that involves identifying or recognising something found in the world

as oneself. The fact that introspection is immune to these errors, then, tells us something important — it tells us that this is not the case, that introspection is a form of self-knowledge in which we are presented to ourselves under a uniquely subjective perspective. Under its modal formulation, then, the immunity to error through misidentification of introspective self-ascriptive judgments is crucial to the preservation of our best current theories of introspection; it allows us to uphold the orthodoxy that introspection is a form of self-knowledge in which we come to know about ourselves *as* ourselves.

Compare now the constitutive formulation. The bare fact that — as the revived challenge from thought insertion shows — introspection is a form of self-knowledge that gives rise to first person judgments that are vulnerable to misidentification errors under the constitutive reading does not tell us anything like what it told us under the modal formulation. There are, after all, all sorts of reasons why a subject might fail to exploit the non-identification involving grounds for a self-ascription that are available to her, even if she uses those grounds to form an existential judgment. She might be in the grip of an especially tenacious form of self-deception of a kind that prevents her from forming second order judgments about her first order mental states, or perhaps there are repressive psychological mechanisms associated with post traumatic stress disorder at work. She could be under the influence of powerful psychoactive drugs that inhibit the formation of self-ascriptive judgments while leaving open introspective access to the properties that are being instantiated. Likewise, nothing in the current state of the literature in the neurosciences rules out the possibility that brain damage of certain kinds might lead to comparable effects, and so on. Given the diverse menu of possible sources of a false negative error, the fact (if it was one) that such a mistake was precluded would carry very little information about the faculty of introspection. It would tell us only that no item on a long list of assorted forms of possible interference could have taken place. This is not to tell us nothing. But it is to tell us something about the conditions in which the episode of introspection took place, not about the intrinsic structure of introspection itself. So long as we take claims about the immunity or vulnerability to error through misidentification of introspective self-ascriptive judgments to bear on our structural understanding of the faculty of introspection, then, we had better reject the constitutive formulation of immunity to error through misidentification.

3 Two findings

The suggestion is not that ruling in or out introspection-based false negative errors of certain kinds is not of philosophical interest in its own right, only that the significance

of such claims is not the same as the significance traditionally attached to claims about immunity to error through misidentification. There are two findings to this paper. The first is the identification of two structurally contrastive kinds of error that have sometimes been run together under the rubric of introspective errors of misidentification. At its most modest, the directive that falls out of this finding is to take care in keeping these two things apart in discussions of immunity to error through misidentification — though at the end of the last section I also urged, a bit more ambitiously, that the sort of significance normally attached to claims about immunity to error through misidentification makes it natural to opt for what I have been calling the modal notion of immunity to error through misidentification over the constitutive formulation. Even accepting this division, however, there might be interest in pursuing questions about immunity or vulnerability to the kinds of false negative errors we have been discussing. One such source of interest might come from an attraction to what Johannes Roessler has called *transparency*, the claim that '[t]o be introspectively aware of a current episode of thinking that p is to be aware of oneself thinking that p.' (Roessler 2013, p. 1). *Transparency*, or something like it, is certainly an initially compelling thesis in its own right, and — along with the kinds of scenarios raised at the end of the last section — cases of thought insertion will be relevant in its final assessment. But, and this is the point of the first finding, that will not be a question about immunity to error through misidentification.[10]

Another point of interest might be the ways in which the possibilities for false negative errors of this kind interact with the possibility or impossibility of false positive errors of misidentification. The discussion of the last section showed introspection to be a form of self-knowledge that combines immunity to the relevant false positive errors with vulnerability to the kinds of false negative errors associated with the constitutive formulation. What this shows is that the fact that introspective judgments are immune to error through misidentification relative to uses of the first person concept does not guarantee that it is a form of self-knowledge that is also marked by Roessler's transparency; these two epistemic features can, and in the case of introspection seemingly do, come apart.

Through these facts about the epistemic profile of introspection we discover something somewhat surprising about the significance of claims to immunity to error through misidentification. This brings us to the second finding of the paper. The traditional significance of such claims, at least as I have been characterising it, is that it tells

[10]Though it is not made explicit, Roessler himself seems to assume that transparency lines up with immunity to error through misidentification; he writes of transparency, '[t]he idea is familiar from discussions of 'immunity to error through misidentification' (Roessler 2013, p. 5).

us something special about the form of self-knowledge about which the claim is made. It tells us that it is an identification-free form of self-knowledge, and so one in which a subject is immediately given to herself *as* herself. In the introduction I gave some reason for thinking that this — or, at least, something very nearby — is a dominant way of understanding the significance of immunity to error through misidentification. The fact that introspection combines this form of immunity with vulnerability to the kind of introspection-based false negative error brought out by the delusion of thought insertion now drives something of a qualification to this presumed significance. The significance of saying that introspection is a source of judgments that are immune to error through misidentification is that it shows introspection to be a form of self-knowledge in which one is given to oneself as oneself *providing that all goes well* — providing, that is, that the introspective episode is undergone in the absence of additional factors that could interfere with the transition from introspection to introspective self-ascription. In this respect, there is something interestingly right about the challenge from thought insertion with which this paper started. Even if the delusion does nothing to show the identification-dependence of introspection, it shows that introspection is not a faculty in which one is always laid bare to oneself as oneself. The second finding of this paper is that this cannot then be the unqualified significance of claims to immunity to error through misidentification.

As an empirical example of introspective conditions in which the normal sense of self is disrupted, the delusion of thought insertion is of special interest to philosophers investigating the epistemic and phenomenological structures of introspection. A common strategy is to use the notion of immunity to error through misidentification as a tool with which to pick apart the implications of thought insertion for our understanding of the faculty of introspection. In this paper I have turned that strategy on its head: I have drawn on the delusion of thought insertion and our understanding of introspection to say something about the epistemologist's device of immunity to error through misidentification. There have been two central results. The first is the identification of a formal distinction between two epistemic phenomena that are sometimes conflated under the guise of misidentification errors. Both are of interest in their own right, but we must guard against running them together if our discussions about immunity to error through misidentification are to run along straight tracks. The second is something of a weakening of the presumed significance of claims to immunity to error through misidentification. We need not throw out altogether the idea that a form of self-knowledge marked with immunity to error through misidentification is one through which a subject is given to herself as herself. But we must add the qualification that even this non-identification

involving way of being given to oneself is vulnerable to additive interference; even if I am given to myself *as me*, there are all sorts of ways this might be distorted or screened off from myself. With these tightenings to the notion of immunity to error through misidentification, the idea is, we will be in a better position to turn again to questions about the epistemic and phenomenological nature of introspection.

References

Allison-Bolger, V. Y. (1999). "Collection of case histories".

Bortolotti, Lisa and Matthew Broome (2009). "A role for ownership and authorship in the analysis of thought insertion". In: *Phenomenology and the Cognitive Sciences* 8.2, pp. 205–224.

Campbell, John (1999). "Schizophrenia, the space of reasons and thinking as a motor process". In: *The Monist*.

— (2002). "The ownership of thoughts". In: *Philosophy, Psychiatry, and Psychology* 9.1, pp. 35–39.

Cassam, Quassim (1997). *Self and World*. Oxford University Press.

Chen, Cheryl (2011). "Bodily awareness and immunity to error through misidentification". In: *European Journal of Philosophy*.

Coliva, Annalisa (2002). "Thought insertion and immunity to error through misidentification". In: *Philosophy, Psychiatry, and Psychology* 9.1, pp. 27–34.

— (2006). "Error through Misidentification: Some Varieties". In: *The Journal of Philosophy* 103.8, pp. 403–425.

De Vignemont, Frédérique (2012). "Bodily Immunity to Error". In: *Immunity to Error through Misidentification*. Ed. by F. Recanati and S. Prosser. Cambridge University Press.

Evans, Gareth (1982). *The Varieties of Reference*. Ed. by J. McDowell. Oxford University Press.

Fernández, Jordi (2010). "Thought insertion and self-knowledge". In: *Mind and Language* 25.1, pp. 66–88.

Frith, C.D. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Psychology Press.

Graham, G. and G.L. Stephens (2000). *When Self-Consciousness breaks*. MIT Press.

Hamilton, Andy (2009). "Memory and self-consciousness: Immunity to error through misidentification". In: *Synthese* 171.3, pp. 409–417.

Hoerl, Christoph (2001). "On Thought Insertion". In: *Philosophy, Psychiatry, and Psychology* 8.2-3, pp. 189–200.

Langland-Hassan, Peter (2015). "Introspective Misidentification". In: *Philosophical Studies* 172.7, pp. 1737–58.

Longuenesse, Beatrice (2012). "Immunity to error through misidentification: New essays". In: ed. by Simon Prosser and Francois Recanati. Cambridge University Press. Chap. Two uses of 'I' as subject?, pp. 81–103.

Mellor, C.S. (1970). "First Rank Symptoms of Schizophrenia". In: *The British Journal of Psychiatry* 117.536, pp. 15–23.

Morgan, Daniel (2015). "Thinking About the Body as Subject".

Recanati, François (2012). "Immunity to error through misidentification: what it is and where it comes from". In: *Immunity to Error through Misidentification: New Essays*. Ed. by Simon Prosser and Francois Recanati. Cambridge University Press, pp. 180–201.

Roessler, Johannes (2013). *Thought Insertion, Self-Awareness, and Rationality*. Ed. by R. Gipps B. Fulford M. Davies. Vol. The Oxford Handbook of Philosophy and Psychiatry. Oxford University Press.

Saks, Elyn R. (2007). *The Centre Cannot Hold: my journey through madness*. Hyperion Press.

Shoemaker, Sydney (1968). "Self-Reference and Self-Awareness". In: *Journal of Philosophy* 65.19, pp. 555–567.

Smith, Joel (2006). "Which Immunity to Error?" In: *Philosophical studies*.

Wittgenstein, Ludwig (1958). *The Blue and Brown Books: Preliminary studies for the 'Philosophical Investigations'*. Blackwell Publishing.

Wright, Crispin (2012). "Reflections on Francois Recanati's 'Immunity to Error through Misidentification: what it is and where it comes from'". In: *Immunity to Error through Misidentification: New essays*. Ed. by Simon Prosser and Francois Recanati. Cambridge University Press, pp. 247–280.