Immunity and Self-Awareness

Max Seeger
Heinrich-Heine University Düsseldorf

© 2015 Max Seeger
This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 License.
<www.philosophersimprint.org/015017/>

1. Introduction

In thought insertion, a pathological experience found in schizophrenia, patients experience thoughts of which they claim not to be the thinker. In an anarchic hand syndrome, a condition in which a patient’s hand seems to acquire a will of its own, the anarchic hand performs goal-directed movements which the patient is aware of, but claims not to be doing. In a particular case of somatoparaphrenia, a delusion in which subjects disown parts of their bodies, one subject is aware of touches delivered to her disowned hand yet purportedly claims that these sensations are not experienced by her, but by her niece. These cases have been alleged to refute the philosophical dictum that one cannot doubt who is the thinker of the thoughts one is aware of, who is the agent of the actions one is aware of, and who is the subject of experiences one is aware of. In particular, these cases have been claimed to refute the thesis that self-ascriptions of mental states based on introspection are immune to error through misidentification (cf. Campbell 1999a, Marcel 2003, Lane & Liang 2011).

What does it mean to be in error through misidentification? We can intuitively distinguish two kinds of mistakes that could afflict a judgment of the form ‘a is F’. One error pertains to the object, a; the other error pertains to the property, F. Suppose, for instance, that upon seeing a man running, whom I take to be Forrest, I judge that Forrest is in a hurry. If the man I’ve seen is not Forrest, my judgment is in error.
through misidentification. If the man is not in a hurry, my judgment is in error through mispredication.

The classical Immunity Thesis is the claim that error through misidentification is not possible in introspection-based self-ascriptions of mental states. Very roughly, this means that if one self-ascrives a mental state of which one is introspectively aware, then one cannot go wrong in ascribing the state to the wrong subject. This contrasts with exteroception-based self-ascriptions. Suppose I judge that my coat is dirty upon seeing a reflection in a shop window. This judgment is vulnerable to error through misidentification, as it may happen that I mistake somebody else’s reflection for mine. In that case, I would attribute the property which I correctly perceived (wearing a dirty coat) to the wrong subject. Note that I am the wrong subject of attribution not in the sense that my coat isn’t dirty (for even if my coat happened to also be dirty, the judgment would still involve a misidentification), but in the sense that the reflection I’ve actually seen is not mine, but somebody else’s. When self-ascriptions are based on introspection, the Immunity Thesis states, this kind of error is not possible.²

Various forms of critique have been raised against the Immunity Thesis (henceforth: Immunity), both theoretically and empirically motivated. In this paper, I discuss whether the three pathological cases of alienation introduced above, thought insertion, anarchic hand syndrome, and somatoparaphrenia, refute Immunity. I argue that, given a proper understanding of the thesis, the putative counterexamples either fall outside the scope of the thesis or do not involve a misidentification and therefore do not refute Immunity.

2. The Immunity Thesis

To properly assess whether an alleged counterexample actually undermines the Immunity Thesis, we have to get clear on what exactly the thesis states. Regrettably, this crucial step has been neglected by those who claim to have refuted the thesis. In particular, two questions have to be answered.

1. Definition of misidentification: What is an error through misidentification?

2. Scope of the thesis: Which judgments are claimed to be immune to such error?

Let us take them up in turn. First, what is an error through misidentification? Defining the notion of misidentification potentially raises a host of controversial issues. Luckily, we can ignore most of these, since the arguments for rejecting the putative counterexamples do not depend on any particular approach or definition. A rough answer to the question what it means to be in error through misidentification has already been given above by the intuitive distinction between misidentification and mispredication. Disregarding, for the moment, the possibility that a judgment can exhibit both errors at the same time, the distinction can be put this way: In cases of misidentification, one attributes the right property to the wrong object; in cases of mispredication, one attributes the wrong property to the right object. A more precise definition, the one I appeal to in what follows, is this: A judgment of the form ‘a is F’ is in error through misidentification if and only if the source object (i.e. the object from which the predication information derives) is different from the target object (i.e. the object to which the predicate is applied).³

This definition of error through misidentification differs substantially from the most common way of defining it, which is in epistemic terms. The most straightforward definition holds that a judgment is in error through misidentification if and only if it is (epistemically) based on a false identity proposition (cf. Evans 1982, pp. 180–182). James

². Many authors hold that immunity to error through misidentification is exhibited not only in first-personal thought, but also in demonstrative ways of thinking about objects, time or space. In this paper, I am concerned solely with the immunity of first-personal thought.

³. This conception of error through misidentification is inspired by Simon Prosser’s observation that “a judgment is immune to error through misidentification in just those cases where the source object and the target object cannot differ” (2012, pp. 161ff.).
Pryor (1999) argues that this definition leaves out cases of which-misidentification. Roughly, these are cases in which a subject knows a property to be instantiated and goes wrong in singling out who or what is the bearer of that property. While the details are debated, most writers share an epistemic approach in the sense that they define error through misidentification in terms of the rational or epistemic grounds of the judgment. The epistemic nature of this approach is also reflected in the most common definition of immunity to that error, viz. the definition in terms of retreat to existential generalization: a judgment of the form ‘a is F’ is immune to error through misidentification (IEM) if and only if any ground for doubt that a is F is ipso facto ground for doubt that anything is F. My own definition bypasses certain epistemic issues (e.g. regarding the assessment of cross-wiring scenarios) by construing error through misidentification in rather metaphysical terms as the coming apart of source and target.

But, as I said, my argument against the putative counterexamples does not depend on any particular definition. My definition construes error through misidentification in the widest possible sense and is therefore the one most favorable to the critics of Immunity. To see this, consider how the epistemic approach defines error through misidentification in terms of a judgment’s being based on a false “identifying” belief — that is, in terms of being based either on a false identity belief (‘this man is Forrest’) or on a false singling-out belief (‘this is the animal I am smelling’). The identifying belief in question can be false only if source of predication information and target object are distinct. The satisfaction of the epistemic criterion therefore requires, but is not exhausted by, the satisfaction of the metaphysical criterion. Hence, my construal of error through misidentification allows for more judgments to be classified as involving such error than any other construal. This in turn renders the Immunity Thesis, i.e. the thesis that one cannot make this error in certain kinds of judgments, more vulnerable than it would be given any other construal of error through misidentification.

Let us then turn to the second question, regarding the scope of the thesis: Which judgments are typically claimed to be immune to the type of error just specified? The classical Immunity Thesis goes back to Ludwig Wittgenstein (1958) and Sydney Shoemaker (1968) and holds that introspection-based self-ascriptions of mental states are immune to error through misidentification. Both Wittgenstein and Shoemaker illustrate the thesis with self-ascriptions of phenomenal states, beliefs, actions or intentions to act, and perceptions. Crucially, the self-ascriptions in these examples are to be understood as based on first-personal awareness of these states. Gareth Evans (1982) adds that self-ascriptions of bodily states are IEM when based on proprioception.

First, I shall clarify the notion of self-ascription. Wittgenstein and Shoemaker discuss the phenomenon on the linguistic level, i.e. they are concerned with sentences in which a speaker self-ascribes a state. Today, Immunity is typically discussed on the mental level; it is construed as a thesis about thoughts, especially about judgments, about oneself. Below, I briefly explore the possibility of broadening the scope of the thesis to mental states in general, in particular to feelings, as long as they carry representative content.

Now, which judgments are claimed to be IEM? I suggest that the essence of the Wittgenstein-Shoemaker-Evans idea can be summarized by two criteria which delineate the scope of the thesis. Immunity applies to all and only those judgments that satisfy what I dub the target and the source criterion. Roughly, a judgment satisfies the target criterion if and only if it is a self-ascription; a judgment satisfies the source


5. To prevent misunderstanding: I do not mean to say that my suggested definition is entirely non-epistemic; the notion of a source certainly is an epistemic notion.

6. Some authors may hold that a judgment fails to be immune to error through misidentification simply in virtue of being based on an identity assumption, even if that assumption is necessarily true (i.e. if source and target cannot come apart). Thanks to an anonymous reviewer of this journal for pointing this out. Unfortunately I do not have the space to address this issue here.
criterion if and only if it is based on a first-personal way of gaining information. Let me elaborate.

The target criterion restricts the scope of Immunity to de se self-ascriptions, a point that seems quite undisputable. Consider how Wittgenstein distinguishes two different uses of the word ‘I’ (1958, pp. 66f.) and how Shoemaker’s claim is about “error through misidentification relative to the first-person pronouns” (1968, p. 556; my emphasis). The target criterion hence specifies that a judgment falls within the scope of the Immunity Thesis only if the subject self-ascribes a property. The Immunity Thesis regarding the first person is silent about judgments in which a property is attributed to another person (or to oneself, but in a mode that is not de se). Obvious though this point may seem, it is well worth pointing out, since some of the putative counterexamples do not satisfy this criterion.

The source criterion restricts Immunity to apply solely to judgments that are based on first-personal ways of gaining information. Notably, one and the same judgment (e.g. ‘my legs are crossed’) can be immune when based on one type of information channel (proprioception), but vulnerable to error through misidentification when based on another type of information channel (vision). Roughly, the divide is between “inner” and “outer” information channels. That is to say, Immunity applies to judgments based on introspection or proprioception, broadly construed, but not to judgments based on external perception.

I hasten to add several qualifications. First, external perception may deliver some information in a first-personal way. Most importantly, due to its egocentric frame of reference, external perception arguably delivers information about one’s relation to other objects in a first-personal way (Evans 1982, p. 222). Further, according to the transparency thesis regarding self-knowledge, my knowledge that I do not believe there to be a third world war is not gained through introspection but by looking out into the world (cf. Evans 1982, p. 225). Second, some judgments may be IEM without being based on first-personal ways of gaining information. For instance, Frédérique de Vignemont (2012) argues that certain visual perceptions (roughly: looking down one’s body) deliver IEM judgments. Also, judgments such as ‘I exist’ and ‘I am thinking’ may be IEM when based solely on rational deliberation. I mention these possible exceptions just to put them aside, since they are not directly relevant to the cases under discussion.

Let me add a clarification regarding the terms ‘proprioception’ and ‘introspection’. The term ‘proprioception’ is sometimes used in a narrow sense, denoting a specific inner sense modality that can be distinguished from other inner senses such as interoception and kinesthesia. Here, I want ‘proprioception’ to be understood in a broad sense as denoting all first-personal, inner ways of being aware of bodily states. The term ‘introspection’ is notoriously hard to define. For the purposes of this paper, I intend it to denote the first-personal ways of being aware of mental states.

I assume that, the above exceptions aside, target and source criteria are singly necessary and jointly sufficient for delineating the scope of Immunity. However, it may be suggested that a third criterion is necessary to fully capture the classical idea, namely a criterion specifying the type of predicate involved in the judgment. Wittgenstein and Shoemaker are specifically interested in the self-ascriptions of mental states; Evans is interested in the self-ascription of certain bodily states. A third criterion suggests itself accordingly: A judgment falls within the scope of Immunity only if it involves the ascription of a mental or a certain bodily property. Let’s call this the property criterion.

We can ignore the property criterion for two reasons. First, ignoring it is favorable to the critics of Immunity and therefore dialectically harmless. If at all, the property criterion would provide yet another

---

7. I should note that this way of putting the idea does not sit well with a different view, expressivism, sometimes attributed to Wittgenstein. The idea of expressivism is that sentences involving the word ‘I’ as subject are not self-ascriptions in any good sense, but are rather expressions of the mental states in questions. Taking this idea to the extreme, Anscombe (1975) famously argues that the first-person pronoun, when used as subject, is not a referring term at all. Thanks to Jérémie Lafraire for pointing this out.

8. Thanks to an anonymous reviewer of this journal for pressing me on this.
anchor point to exclude putative counterexamples. Second, and more importantly, a restriction to mental (and certain bodily) states is already in place in virtue of the source criterion. That is to say, the kinds of judgments that this additional criterion would exclude from the scope are already excluded by the source criterion: we do not have introspective or proprioceptive awareness of, say, our date of birth, our driving skills, or the color of our hair.

3. The Putative Counterexamples

Given the above picture of error through misidentification and the Immunity Thesis, a judgment constitutes a counterexample to the thesis if and only if the judgment (a) involves an error through misidentification (i.e. source and target are not identical) and (b) falls within the scope of the thesis (i.e. satisfies source and target criteria). As I will now show for each of the putative empirical counterexamples in turn, they all fail on either (a) or (b) and hence do not refute Immunity.

3.1 Thought Insertion

The first empirical counterexample that has been raised against the Immunity Thesis is the phenomenon of thought insertion in schizophrenia (Campbell 1999a). In thought insertion, subjects claim that certain thoughts they are experiencing are not their own thoughts. Typically they also claim that the thought in question is in fact somebody else’s thought. Here is one widely cited report: “Thoughts are put into my mind, 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). John Campbell takes this to be an (at least) prima facie threat to Immunity.

A patient who supposes that thoughts have been inserted into his mind by someone else is right about which thoughts they are, but wrong about whose thoughts they are. So thought insertion seems to be a counterexample to the thesis that present-tense introspectively based reports of psychological state cannot involve errors of identification. (1999a, pp. 609f.)

The discussion of thought insertion requires a few conceptual clarifications. First, I distinguish two judgments typically found in thought insertion: the disowning judgment (‘this is not my thought’) and the external attribution (‘this is Chris’s thought’) (cf. Stephens & Graham 2000, p. 152). Secondly, I appeal to an established distinction of what it means for a thought to be one’s own. On the one hand, a thought can be my own in the sense that I am the one experiencing the thought, that the thought is introspectively available to me or that the thought appears in my stream of consciousness. Judging a thought to be one’s own in this sense is what I call the ownership attribution. On the other hand, a thought can be my own in the sense that I am the producer, the active thinker, or the causal origin of the thought. Judging a thought to be one’s own in this sense is what I call the authorship attribution.9

According to the standard interpretation of thought insertion, subjects deny authorship but not ownership for the thought. To disambiguate, the judgments ‘this is not my thought’ and ‘this is Chris’s thought’ can thus be rephrased as the judgments ‘I am not the author of this thought’ and ‘Chris is the author of this thought’, respectively. A common response to the alleged counterexample is to concede that authorship ascriptions are not IEM but to insist that ownership ascriptions are (e.g. Gallagher 2000, Vosgerau & Voss 2014). As we will see shortly, this concession is in fact too quick.

To begin, we need to ask which of the two judgments is supposed to be the counterexample to the Immunity Thesis. The external attribution (‘Chris is the author of this thought’) can easily be shown to fall outside the scope of the thesis. It clearly fails the target criterion since it is not a self-ascription.10 It arguably also fails the source criterion

9. There is no universally accepted terminology, but roughly the same distinction can be found in Campbell (1999a), Gallagher (2000), Stephens & Graham (2000), Sousa & Swiney (2013).

10. Coliva (2002, p. 30) was the first to note this point.
since it is not based on introspective awareness of a thought experience. Rather, it is a confabulated answer to the question who made that thought appear within the subject’s mind. Hence, the external attribution doesn’t undermine the Immunity Thesis.

The disowning claim (‘I am not the author of this thought’) fares much more promising with respect to the scope criteria. I begin with the target criterion. Is the claim ‘I am not the author of this thought’ a self-ascription? To be sure, subjects deny having a certain property rather than ascribe a property to themselves. This seems to conflict with paradigmatic cases of immunity to error through misidentification, in which subjects ascribe properties to themselves rather than deny them. It is tempting, therefore, to construe Immunity as a thesis solely about positive self-ascriptions and to reject thought insertion as a counterexample simply on the grounds that it is not a positive self-ascription.\textsuperscript{11}

However, such a construal of the thesis would require a criterion to distinguish positive claims (i.e. claims to the effect that one instantiates a property) from negative claims (i.e. claims to the effect that one does not instantiate a property). The problem is that the grammatical structure of a thought’s linguistic expression is not a good guide to whether a thought involves a positive self-ascription or not. Consider a thought that could be expressed equally well by the sentences ‘I don’t want to go to sleep yet’ and ‘I want to stay up longer’. Grammatical structure suggests that the first sentence denies a property whereas the second sentence ascribes a property to the speaker. But this superficial difference should not matter for questions regarding Immunity.

More importantly, it is not at all clear why there should be an epistemic difference between positive and negative claims about the instantiation of properties when these claims are based on the same kind of information. What matters regarding the scope of Immunity is not whether I ascribe or deny a property; what matters is whether I make a judgment about myself or about somebody else. The fact that the paradigmatic cases are always positive self-ascriptions is misleading. Consider Wittgenstein’s examples for the use of ‘I’ as subject: “I try to lift my arm, ‘I think it will rain’, ‘I have toothache’” (1958, p. 66). The fact that properties are ascribed rather than denied is not central to the case. The point could be made equally well with the judgments ‘I do not try to lift my arm’, ‘I do not think it will rain’ and ‘I do not have a toothache’. When these negative claims are based on introspection, there is no way that I could mistake somebody else’s not trying to lift their arm for my own not trying, or somebody else’s not being in pain for my own not being in pain. Hence, Immunity is restricted not to positive self-ascriptions, but to judgments which are about the subject (positively or negatively). In other words, the target criterion excludes from the scope only those judgments that are about somebody else. In particular, the target criterion is satisfied by claims such as ‘I am not the author of this thought’ or ‘I am not the agent of this movement’.

We now turn to the source criterion. The question here is: Are disowning claims in thought insertion based on introspection? The answer depends on what the primary thought experience at the core of thought insertion is like. There are two views on the matter.\textsuperscript{12} According to the endorsement approach, the thought is already represented as not one’s own in the primary thought experience. The judgment ‘I am not the author of this thought’ simply expresses this experience. If this is correct, the judgment is based on introspective awareness of the primary thought experience and hence satisfies the source criterion. In contrast, according to the explanationist approach, the primary thought experience does not represent the thought as not one’s own (although it may represent the thought as strange in other respects). The explanationist approach holds that the disowning is an attempt to explain or rationalize the occurrence of the thought (e.g. because it has an unwelcome or unfamiliar content); it is not based directly on

\textsuperscript{11} See e.g. de Vignemont’s (2012) distinction between false positives and false negatives. To be fair, her main aim is to exclude external attributions from the scope of Immunity. But the way she puts the distinction may indeed suggest that only positive self-ascriptions fall within its scope.

\textsuperscript{12} The following distinction has been introduced by Pacherie et al. (2006). Sollberger (2014) applies it to thought insertion.
introspection of the thought in question, but rather on confabulation or some kind of inference. If this is correct, it is debatable whether the disowning claim satisfies the source criterion. I want to stay neutral with respect to the two approaches and will simply grant to the critic that the judgment satisfies the source criterion.

The disowning claim (‘I am not the author of this thought’) hence might actually satisfy the two scope criteria. It fails as a counterexample for a different reason: It does not involve a misidentification. To put it simply, it is not the case that the subject has information of somebody else’s having a property F and goes wrong in ascribing that property to himself. Since we are dealing with a negative self-ascription, perhaps it should be put this way: it is not the case that the subject has information of somebody else’s not having a property F and goes wrong in denying that property of himself. The mistake that is involved in the disowning claim is a misprediction, not a misidentification.

The same point can be made in terms of source and target. Remember: for there to be a misidentification, the object from which the information derives has to be different from the object to which a property is ascribed. In construing the disowning claim as a self-ascription, I assumed that the target object is the subject himself. This assumption is necessary for the judgment to satisfy the target criterion. Now, for target and source to come apart, the source would have to be something or somebody other than the subject. That is, the information on which the judgment is based would have to derive from somebody else. But that is clearly not the case. However the subject may have come up with the judgment that he is not the author, it certainly isn’t the case that he mistook somebody else’s not being the author for his own.

If this result seems surprising, it may be because disowning and external attribution have not been kept properly apart. In some sense, it is true that thought insertion involves both a misidentification and a self-ascription: The external attribution involves a misidentification and the disowning involves a self-ascription, but there is not one single judgment that both is a self-ascription and involves a misidentification.

It may be objected that one should look not at the doxastic component of the delusion, i.e. the judgment ‘I am not the author’, but rather at the phenomenal component, i.e. the experience of a lacking sense or feeling of authorship. Although Immunity is typically construed as a thesis about judgments, I will grant for the sake of argument that it can be extended to feelings, as long as these are construed as states that carry a representational content. The idea of extending the thesis would roughly be this: If I am experiencing a feeling with the content, say, ‘I am hungry’, then it is not possible for the feeling to veridically represent that someone is hungry, but to misrepresent who is hungry. Now, if there is any representational content to the feeling of a lack of authorship, it must be the same as the content of the disowning claim discussed above, namely ‘I am not the author’. Hence both can be treated on a par. For the same reason that the judgment of non-authorship does not involve a misidentification, the feeling of non-authorship does not involve a misidentification either. To wit, the feeling does not ascribe the right property to the wrong subject, but rather ascribes the wrong property to the right subject. Hence, none of the judgments or feelings involved in thought insertion refutes the Immunity Thesis.

3.2 Anarchic Hand Syndrome
Anthony Marcel (2003, pp. 80f.) claims that anarchic hand syndrome refutes the Immunity Thesis regarding bodily agency. In anarchic hand syndrome, one hand of the patient performs “unintended but complex, well-executed, goal-directed actions” which often compete with what the patient actually intends to do (op. cit., p. 77). The hand itself is not (necessarily) disowned by the patient; i.e. the patient (perhaps) acknowledges that it is his or her own hand. The crucial aspect is that the actions which this hand performs are not intended by the subject.

13. These two very different properties, being the author of a thought vs. having the sense of being the author of a thought, are often conflated in the debate. For the distinction between judgment of agency and feeling of agency see Synofzik et al. (2008).
and appear to the subject as if not done by him. Marcel suggests that these subjects make an identification mistake in denying agency.

The anarchic hand phenomenon, just like thought insertion, may involve both a judgment of non-agency and a feeling of non-agency. It is not clear which of the two Marcel has in mind as counterexample, but since the representational content is the same in both cases (‘I am not the agent of this action’), it suffices to discuss only one of them; I will discuss the judgment.

I will present two objections against Marcel’s case. Before I do so, let me mention, just to put aside, a critique raised by Christopher Peacocke (2003, p. 109). Peacocke holds that the movements of the anarchic hand are not actions at all and that therefore there is no mistake involved in the subject’s denial of agency. Peacocke’s guiding assumption seems to be that for the movement to be a φ-ing it would have to essentially involve a trying to φ, which it does not. Since the following arguments can do without that assumption, I will grant to Marcel for the sake of argument that the hand’s movement is an action.

The first objection is that the judgment does not satisfy the source criterion. Let us assume that there is something we may call first-personal action awareness. Plausibly, this kind of action awareness would involve introspective awareness of the action intention. However, as Marcel himself stresses, in anarchic hand syndrome “awareness from the inside of relevant intention, effort, and will are lacking” (2003, p. 79; Marcel’s emphasis). In what sense, then, can it be said that the subjects are first-personally aware of the anarchic action? Subjects may have proprioceptive (i.e. first-personal) awareness of the hand’s movements (op. cit., p. 81). But that does not suffice for first-personal action awareness. For it is the goal-directedness of the hand’s behavior that makes it an action. But clearly, subjects are not aware of this goal-directness first-personally, via awareness of the intention, but are aware of the goal-directedness in a third-personal kind of way. For instance, when they see (and feel) the anarchic hand unbutton the shirt which their other hand is trying to button up, they perceive the hand’s movement as an unbuttoning. But they do so in a third-personal way, just as a bystander would perceive the movement as an unbuttoning. Crucially, they are not first-personally aware of an intention to unbutton the shirt. The fact that, next to seeing the hand, they also perceive the hand’s movement internally does not make a relevant difference in terms of action awareness. For the case to challenge Immunity, it would have to involve first-personal action awareness. For subjects do not deny that it is their hand that is moving; they deny that they are moving it.¹⁴

My second and more fundamental objection is, again, that the denial of agency does not involve a misidentification. The subject does not attribute the right property to the wrong person, but rather attributes the wrong property to the right person. For there to be a misidentification, the source of information on which the predication is based has to be different from the target to which the predicate is applied. That is to say, in anarchic hand syndrome, the information on which the denial of agency is based would have to derive from some other person’s agency or non-agency. And that clearly is not the case.¹⁵

3.3 Somatoparaphrenia

Somatoparaphrenia is the disowning of one of one’s body parts together with the attribution of the body part in question to another person. Timothy Lane & Caleb Liang (2011) argue that a particular case of somatoparaphrenia disproves the Immunity Thesis. In this case, a stroke patient (FB) suffering a lesion in the right cerebral hemisphere attributes her left hand to her niece (Bottini et al. 2002). Moreover, FB seems not to feel touches in her left hand (hemianesthesia) unless the touches are announced as delivered to the niece’s hand. The phenomenon has been assessed by touching the blindfolded patient on her left hand in two conditions. In condition one, the touch was announced as

¹⁴ Thanks to Gottfried Vosgerau for suggesting this objection.

¹⁵ If this rejection seems to miss out on some important aspect of the case, note that in section 4 I will come back to discuss how anarchic hand syndrome may be indirectly related to the Immunity Thesis.
being delivered to FB’s left hand; in condition two, the touch was announced as being delivered to the niece’s hand. FB was asked to report with ‘yes’ and ‘no’ whether she felt any touch. In condition one, FB never reported feeling touched; in condition two, FB reliably reported the touch.

Now, what does this case show? Lane & Liang take it to show that FB represents the tactile sensation, of which she is obviously aware in condition two, to be experienced not by herself but by her niece. That is, according to Lane & Liang, FB misrepresents the owner of the tactile experience, not merely the owner of the hand or the location where she (FB) feels the touch. Note how this interpretation builds on a distinction between *being* the subject or owner of an experience (or simply: being the one who experiences) and *representing oneself as* the subject of the experience. While Lane & Liang (at least implicitly) accept that it is FB who is in fact experiencing the sensation, they claim that FB does not *represent* herself as the subject of the experience, but instead misrepresents her niece as the subject. The putative counterexample to the Immunity Thesis hence is FB’s judgment ‘my niece is feeling touched’.

What is interesting and novel about Lane & Liang’s argument is how they turn what looks at first sight like a misattribution of a bodily state into a misattribution of a mental state. However, this move comes at a cost, for they have to assume a very questionable interpretation of the case. What Lane & Liang claim is that FB misrepresents *who is experiencing* the tactile sensation. But a more conservative interpretation is available that explains the results equally well: FB does not misrepresent *who* is having the sensation, rather FB misrepresents *where she (FB) experiences* the tactile sensation. To see what is meant, consider being touched first on your right and then on your left hand. Both times, you are aware that it is you who is experiencing the touch. But you experience the first touch in your right hand and the second in your left hand. Similarly, the conservative explanation holds, FB is well aware that *she* is experiencing the touch. She just misrepresents *where* she is feeling the touch.16

However, even if we accept Lane & Liang’s interpretation, the case does not refute the Immunity Thesis. For, clearly, the judgment in question (‘my niece is feeling touched’) does not satisfy the target criterion. The property in question (feeling touched) is not self-ascribed but attributed to the niece.17 What the case does show, given Lane & Liang’s interpretation, is that it is possible to misattribute a mental state of which one is introspectively aware to the wrong person. The fact that this possibility does not challenge Immunity will be elaborated in section 4.

In an attempt to rescue the case against Immunity, it could be assumed that FB makes an implicit self-ascription. That is, she *implicitly* judges that she does not feel touched. This implicit judgment (‘I do not feel touched’) does satisfy target and (arguably) source criteria. However, this judgment fails as a counterexample for the same reason as the cases discussed above, which should by now be familiar: the judgment does not involve a misidentification. When FB implicitly judges that she does not feel touched, she does not ascribe the right property to the wrong person, but simply ascribes the wrong property to herself. The information on which the self-concerning judgment is based does not derive from some other person, and therefore the judgment does not involve a misidentification.

3.4 Intermediate Summary
Let me quickly recap the discussion of the putative counterexamples. None of the cases can be construed in a way so as to involve a judgment (or feeling) that satisfies both scope criteria but suffers from misidentification (see fig. 1). Hence, none of the cases refutes the Immunity Thesis.

16. For a similar reply see Rosenthal (2010).
17. A similar objection is raised by de Vignemont (2012), who discusses somatoparaphrenia as a possible counterexample to bodily immunity.
Immunity and Self-Awareness

Max Seeger

Immunity and Self-Awareness

Sydney Shoemaker famously claimed the opposite, namely that “in being aware that one feels pain one is, tautologically, aware, not simply that the attribute feel(s) pain is instantiated, but that it is instantiated in oneself” (Shoemaker 1968, pp. 563f.; Shoemaker’s emphases). Let us call this claim Self-Awareness: If one is introspectively aware of a mental state, then one is necessarily aware of that state as one’s own state. Evans professes a corresponding claim about the proprioceptive awareness of bodily states:

There just does not appear to be a gap between the subject’s having information (or appearing to have information), in the appropriate way, that the property of being $F$ is instantiated, and his having information (or appearing to have information) that he is $F$; for him to have, or to appear to have, the information that the property is instantiated just is for it to appear to him that he is $F$. (1982, p. 221; Evans’s emphases)

On the face of it, Self-Awareness may seem to be refuted by the pathological cases of alienation.

But what does Self-Awareness have to do with Immunity? First, note well that Self-Awareness does not say the same as Immunity (this will be elaborated below). However, many writers assume a close connection between Self-Awareness and Immunity in roughly the following way: Self-ascriptions of mental states are immune to error through misidentification because they are identification-free in the sense that the subject does not need to figure out whose mental states she introspects. Identification-freedom, in turn, is guaranteed by Self-Awareness: introspection necessarily presents one’s mental states as one’s own.

Now, the pathological cases seem to show that there is something wrong with this picture. Patients do go wrong in ascribing mental

<table>
<thead>
<tr>
<th></th>
<th>source</th>
<th>target</th>
<th>misidentification</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>thought insertion</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘Chris is the author’</td>
<td></td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>‘I am not the author’</td>
<td></td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td><strong>anarchic hand</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘I am not the agent’</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td><strong>somatoparaphrenia</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘my niece feels touched’</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>‘I do not feel touched’</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

Fig. 1: assessment of putative counterexamples

The results are by no means a coincidence, but reflect the logic of Immunity. They reveal a tight relation between the notions ‘target’, ‘source’ and ‘misidentification’. Once we assume that the pathological judgments are based on the subjects’ own experiences and thereby satisfy the source criterion, we get the following dependency relation: If the judgments are self-ascriptions, the target criterion is satisfied, but there is no coming apart of source and target and hence no misidentification. If the judgments are not self-ascriptions, but ascriptions to an external entity, we do get a misidentification, but the target criterion is breached.

4. Self-Awareness and the Explanation of Immunity

I argued that the pathological cases do not refute the Immunity Thesis, for they either do not involve a misidentification (as in ‘I am not the author’ and ‘I am not the agent’) or do not satisfy the target criterion (as in ‘Chris is the author’ and ‘My niece feels touched’). However, it might be argued that they do involve a kind of mistake that challenges Immunity indirectly. The mistake in question is that subjects are claimed to be introspectively aware of a thought, an action or a sensation, but fail to...
states of which they are introspectively aware to the wrong person. Doesn’t that show that introspection-based self-ascriptions of mental states are not identification-free after all? And if so, doesn’t it follow that these self-ascriptions are not IEM either? After examining in more detail whether the cases of alienation do refute Self-Awareness (section 4.1) and how Self-Awareness is often assumed to explain Immunity (4.2), I show that a possible refutation of Self-Awareness does not imply anything for Immunity (section 4.3).

4.1 Self-Awareness

Consider again the case of FB. According to Lane & Liang’s interpretation, she is introspectively aware of her tactile sensation, but misrepresents the sensation as her niece’s sensation. Lane & Liang take this to show that FB is not aware of this sensation as her own sensation, although she is introspectively aware of it. This would directly contradict Self-Awareness, which holds that introspective awareness of a sensation implies awareness of that sensation as one’s own.

It may be worth noting that, although Lane & Liang nominally claim to be challenging Immunity, they also put their claim this way: “Our main thesis is: awareness that mental states are instantiated does not entail awareness that said states are instantiated in self [sic]” (2011, p. 83). Or, put negatively, they challenge the claim that “[e]very mental state is, from the first-person point of view, represented as experienced by the one who is introspecting the state” (op. cit., p. 87; Lane & Liang’s emphasis). If we follow these passages, they are actually criticizing Self-Awareness rather than Immunity.

I must at this point reiterate my reservations about Lane & Liang’s interpretation of the case. According to the more conservative interpretation, FB merely misrepresents where she (FB) feels the touch, not who feels the touch. On this interpretation, the case is perfectly compatible with Self-Awareness. However, putting the interpretation of this particular case to the side, Lane & Liang’s interpretation after all suggests that it is conceptually possible to be aware of a touching sensation as somebody else’s sensation.

Similar considerations apply to anarchic hand syndrome. In anarchic hand syndrome, Marcel assumes, the patient is first-personally aware of an action but is not aware of that action as his own. Given this description, the case is incompatible with Self-Awareness. And, just like Lane & Liang, Marcel nominally claims to be challenging Immunity, but actually describes a thesis very close to Self-Awareness in framing his target explicitly. “If one is aware through internal proprioceptive awareness of an action, of a posture, or of a sensation, one might think that it is impossible to be [phenomenally] mistaken about whose it is” (Marcel 2003, p. 80; Marcel’s emphasis). Marcel goes on to equate this claim with Immunity.

But again, the assumption that subjects have first-person awareness of the hand’s action is questionable (see section 3.2). What they are aware of first-personally is merely the hand’s movement. It may be objected that the movement is the action and that, therefore, there is a sense in which subjects are first-personally aware of the action, namely via first-personal awareness of the movement. The point is that they are not first-personally aware of the action as an action; they lack first-person action awareness. And that is what would be needed to challenge Immunity.

Finally, it is thought insertion which is most convincing as a counterexample to Self-Awareness. Here, it seems, subjects are introspectively aware of thoughts without being aware of these thoughts as their own.

A possible defense of Self-Awareness could appeal to the distinction between phenomenal self-awareness and doxastic self-ascription. It could be objected that the cases of alienation primarily show that subjects believe certain states not to be their own, while Self-Awareness is most naturally construed as a thesis regarding the phenomenal experience. It could then be argued that alienated subjects actually do phenomenally experience the thoughts, actions and sensations as their own, and only fail to self-ascribe them due to other factors, such as delusional beliefs, which override this phenomenal awareness.
It is hard to tell whether this response is empirically adequate. Marcel’s description of anarchic hand syndrome certainly speaks against it. According to Marcel, subjects primarily lack the feeling of agency: they may even acknowledge that it is their action, but maintain that it doesn’t feel like theirs (cf. 2003, pp. 79f.). So, to say the least, the cases of alienation put a lot of pressure on Self-Awareness.

Yet another way to defend Self-Awareness would be to construe it in normative rather than descriptive terms. Roughly, this would be the claim that if one is introspectively aware of a state, then one is justified in self-ascribing that state or one would be irrational not to do so. Plausibly, the pathological cases discussed here do not undermine that claim.

Indeed, it may be suggested that the whole notion of immunity to error through misidentification should be understood in normative rather than descriptive terms. The Immunity Thesis, normatively construed, could roughly be this: If one is introspectively aware of a mental state, one cannot rationally wonder whose state it is, or one cannot be justified in doubting that it is one’s own. My descriptive reading, in contrast, does not construe Immunity as a claim about what is rationally or justifiably believable. Rather, it construes it as the claim that it is metaphysically impossible to make a certain kind of mistake in introspection-based self-ascriptions. I will not engage in the exegetical question which one of the two readings is historically more adequate. In many cases it is simply unclear whether the main protagonists in the debate intend their claims to be understood descriptively or normatively. Rather, I am deliberately putting the normative reading of Immunity aside and limiting my discussion in this paper to the descriptive reading. So, while I acknowledge that the two theses, Immunity and Self-Awareness, may be closely connected in their normative readings, I argue that they are logically independent in their descriptive readings.

4.2 Identification-Freedom

Let us now take a closer look at the explanation of Immunity in terms of Self-Awareness and identification-freedom. The idea that identification-freedom implies immunity is expressed by Evans in the following passage:

Clearly, judgements of the first kind [identification-free] are immune to a kind of error to which judgements of the second kind [identification-dependent] are liable. Since they do not rest upon an identification, they are immune to error through misidentification. (1982, p. 182; Evans’s emphasis)

Note that Evans construes identification-dependence as a judgment’s being based on an identity proposition (cf. 1982, pp. 180f.). Pryor (1999) criticizes that this notion of identification-dependence leaves out the kind of identification in which a subject singles out an object as the bearer of a property, issuing into judgments that are vulnerable to so-called which-object misidentification. A broader notion of identification-freedom, one that encompasses freedom from which-object identification, could be construed as a judgment’s not being based on any kind of identification criteria. Although Annalisa Coliva is critical of Pryor’s notion of which-object identification, her explanation of Immunity seems to appeal to the broader notion of identification-freedom:

The reason why one cannot make an error through misidentification when one is self-ascribing a mental property on the basis of one’s introspective awareness of that mental property is that, minimally, introspective awareness is a form of awareness that does not involve either observation or inference. [...] The important point is that because the self-ascription is not based on the observation of oneself, then it cannot be grounded on any
identification component and, therefore, it cannot be affected by [error through misidentification]. (2002, p. 28)

So, in a nutshell, Self-Awareness is taken to explain Immunity as follows: Since introspection necessarily presents mental states as one’s own, one need not identify the owner of an introspected mental state, and one therefore cannot make an error through misidentification (see also Smith 2006).

The pathologies of alienation undermine this explanation. Depending on where the disruption is located (on the level of self-awareness or on the level of self-ascription), either they show that it is possible to be introspectively aware of a mental state without being aware of it as one’s own, or they show that being introspectively aware of a state as one’s own is not sufficient for and hence does not explain its correct self-ascription. The pathological cases may seem, then, to indirectly cast doubt on Immunity. They seem to show that there is some sense in which introspection-based self-ascriptions involve an identification. And if there is identification, there is room for misidentification (see Shoemaker 1968, pp. 56f.).

Before we move on, note that the difference between de re identification and which-object identification and between the corresponding notions of identification-freedom (cf. Lafraire 2013) is not pertinent to the following discussion. Roughly, I argue that the idea of identification-freedom concerns the question how one knows that the states one introspects are one’s own. Explaining Immunity, in contrast, means explaining why the states one introspects necessarily are one’s own. (This will become clearer shortly.) What matters to my critique is that Self-Awareness and the explanation in terms of identification-freedom are construed as descriptive claims. Things will look different on a normative reading. The normative explanation in terms of identification-freedom could be read as saying that there is no gap between a subject’s being justified in believing that a property is instantiated and the subject’s being justified in believing that this property is instantiated in herself. But, again, I will not explore this reading.

4.3 Explaining Immunity
I now show that Self-Awareness (and identification-freedom) did not properly explain Immunity to begin with and propose a different explanation which remains untouched by the pathological cases. To do so, it will be helpful to elaborate why Immunity and Self-Awareness are distinct claims. Self-Awareness holds that if one is introspectively aware of a state, then one is aware of that state as one’s own. Immunity holds that if one self-ascribes a state based on introspection, then this self-ascription cannot be in error through misidentification. One difference to note is that the two theses concern distinct levels: Self-Awareness is a thesis about phenomenal awareness, whereas Immunity is a thesis about doxastic self-ascription. Self-awareness and self-ascription can come apart, as noted in section 4.1.

It may seem that one way to argue why Self-Awareness does not explain Immunity has to do with this gap, i.e. with the fact that awareness of a state as one’s own does not guarantee the self-ascription of that state. This is not my argument. I will argue instead that, even under the assumption that self-awareness and self-ascription go hand in hand, Self-Awareness does not explain Immunity. In assuming that self-awareness and self-ascription go hand in hand, I am ignoring cases in which subjects are phenomenally aware of a state as their own yet fail to self-ascribe it, and cases in which subjects self-ascribe a state without being phenomenally aware of it as their own. I am making this assumption only to bring out a much more fundamental difference between Immunity and Self-Awareness, one that has often been overlooked in the debate.

Immunity, to repeat, holds that if one self-ascribes a state of which one is introspectively aware, then that self-ascription is IEM. To bring out the difference I am after, it will be helpful to slightly rephrase the thesis. Given the notion of error through misidentification introduced above (a judgment is in error through misidentification if and only if the source of the predication information differs from the target object of the ascription), we can translate the Immunity Thesis as follows: If
one self-ascribes a state of which one is introspectively aware, then that state is one’s own state.\footnote{Note that in cases of misprediction, the self-ascribed state is not the same as the introspected state. The term ‘that state’ in the consequent must then be understood to refer to the introspected state. To illustrate, if I judge that I am angry when really I am afraid, my self-ascription of anger does not imply that the anger is my own anger (for there is no anger), but that that the fear on which the judgment is based is my own fear (which I mistook for anger). For the sake of perspicuity, I will ignore the possibility of misprediction in what follows.}

\textbf{(Self-Awareness)} introspection ($F$) $\rightarrow$ self-awareness ($F$)

\textbf{(Immunity)} introspection-based self-ascription ($F$) $\rightarrow$ ownership ($F$)

The following discussion hinges essentially on this definition of Immunity in terms of ownership, and this definition of Immunity in turn depends essentially on my metaphysical notion of error through misidentification. Things may play out differently if the notion of error through misidentification is construed differently, for instance in terms of rationality or justification, but I am not going to explore that.

Now, considering what would falsify each thesis will show that they are mutually independent. Self-Awareness is false if it is possible to be introspectively aware of a state without being aware of it as one’s own. That is exactly what is claimed to be the case in pathologies of alienation. Immunity, in contrast, holds that if one self-ascribes a state based on introspection, then that state really is one’s own. The failure to be self-aware of and self-ascribe an introspected state, which refutes Self-Awareness, does not damage Immunity, for in that case the antecedent of Immunity is false (and hence the implication holds). Instead, Immunity would be falsified if it were possible that an introspected and self-ascribed state is actually somebody else’s state. This case, however, is not ruled out by Self-Awareness.

Let me illustrate the difference by applying the idea to the case of awareness of tactile sensation. Self-Awareness holds that if I am introspectively aware of a tactile sensation, then I am aware of that sensation \textit{as my own} sensation. In other words, Self-Awareness rules out the case that I am introspectively aware of a sensation \textit{as not being my own}. Crucially, Self-Awareness does not imply anything about the question whether that sensation actually is my own or not. Immunity, in contrast, rules out the case that I am introspectively aware of a sensation and take it to be my own, when actually it is not mine.

Speaking more generally, Self-Awareness claims a connection between introspection and awareness of a state as one’s own. It is a claim about whose states the introspected states appear to be. Relatedly, the idea of identification-freedom concerns the question how we know the introspected states to be our own. Immunity, in contrast, claims a connection between introspection-based self-ascription and a state’s being one’s own. It is a claim about whose states the introspected states actually are.

We now begin to see why, contrary to what many authors hold, Self-Awareness never explained Immunity to begin with. Explaining Immunity means explaining why the following implication holds: If I self-ascribe a state of which I am introspectively aware, then that state is my own. Self-Awareness may explain why the antecedent is often true: we do (normally) self-ascribe mental states of which we are introspectively aware. But Self-Awareness does not even begin to explain why the implication holds, for it says nothing about whose states the introspected states actually are. Self-Awareness is perfectly compatible with my being introspectively aware of and self-ascribing somebody else’s mental states.

What actually explains Immunity, then, is the fact that I cannot be introspectively aware of somebody else’s mental states. I can be introspectively aware only of my own states.\footnote{To my knowledge, an explanation along these lines has first been given by Coliva (2002, pp. 28f.). However, Coliva at the same time accepts the explanation in terms of identification-freedom. She offers the ontological explanation in terms of ownership only as a supplement to explain why introspective self-ascriptions (in contrast to proprioceptive self-ascriptions) are logically (rather than de facto) IEM.}

\textit{Immunity and Self-Awareness}
Explains the immunity of bodily self-ascriptions. If one is proprioceptively aware of a bodily state, it is (normally) one’s own state. The more general explanation of Immunity, then, is provided by the assumption of a link between information channel and ownership of the state in question. Plausibly, this link can be of different strengths, depending on the information channel and property in question. Spelling out a comprehensive classification is beyond the scope of this paper and not necessary for the argument. However, to make the claim more tangible, I shall suggest how to treat some of the cases that play a central role in the debate on Immunity.

1. Phenomenal states of which one is introspectively aware are one’s own with conceptual necessity (see the example regarding pain above). Hence, the according self-ascriptions are IEM in the strongest possible sense.

2. Propositional attitudes of which one is introspectively aware are one’s own with at least nomological necessity. It may seem conceivable, but not nomologically possible, that somebody, say, telepathically inserts a desire into somebody else’s mind. There may then be a sense in which that desire is not truly the receiving subject’s own desire. If this is correct, introspective self-ascriptions of propositional attitudes are nomologically IEM.

3. Bodily states that one perceives proprioceptively are one’s own only contingently, namely due to the fact that proprioceptive information channels are hard-wired to one’s own body. Most writers allow for the conceptual possibility of cross-wiring scenarios: it is conceivable that two subjects are wired up in a way that one person proprioceptively perceives the other person’s bodily states. Given this assumption, proprioceptive self-ascriptions of bodily states are only de facto IEM.
External perception delivers information about one’s own relation to other objects, given that normal conditions obtain. For instance, when one sees a table as if in front of oneself, it is usually oneself who is facing the table. But with the aid of virtual-reality goggles, illusions can be created in which what one is seeing as if in front of oneself is really in front of somebody else (see e.g. Petkova & Ehrsson 2008). Hence, exteroceptive judgments about one’s relation to other objects are IEM only given normal conditions.

I am not committed to any particular evaluation of these cases, and the argument does not depend on them. I offer them to illustrate how the information-ownership link may be of different strengths in different cases and to highlight how the suggested explanation of Immunity dovetails with the widely accepted idea that different kinds of judgments enjoy different kinds of immunity.

Finally, let me quickly sketch how the suggested approach handles the putative counterexamples. It has been established above that the judgments in question are IEM. The question now is whether my proposed explanation of Immunity works for these cases as well. This depends on whether the information-ownership link holds in these cases.

First, we need to get clear on the ontological side of the issue: Are the states in question (i.e. the inserted thought, the anachric movement and the disowned tactile experience) the subjects’ own or not? If they are, the information-ownership link holds. This is clearly so in the case of FB’s tactile sensation, as Lane & Liang implicitly agree. However, it might be argued that inserted thoughts and anachric actions are not the subjects’ own. Since they are not anybody else’s either, it must then be assumed that there can be thoughts without authors and actions without agents. Given this assumption, the information-ownership link seems to be in trouble: Subjects are, by assumption, first-personally aware of a thought or an action that is not their own.

However, the link can be saved in both cases by taking a second look at its motivation. As a reminder, the original way of putting the information-ownership link was to say that if one is first-personally aware of a state, then that state is one’s own. This claim gets into trouble only under the assumption that there can be unowned mental states. However, the initial motivation of the information-ownership link as an explanation of Immunity was really to say that if one is first-personally aware of a state, then that state cannot be somebody else’s state. Taking into account the possibility of unowned mental states, the information-ownership link thus has to be amended as follows: If one is first-personally aware of a state, then that state is either one’s own or unowned (but it certainly is not somebody else’s). This way of putting the information-ownership link stays true to the original idea, but is not vulnerable to cases of unowned mental states. Hence, the amended version can explain the immunity of all three cases.

Conclusion

Three pathologies of alienation have been discussed as counterexamples to the Immunity Thesis. I argued that the cases either do not involve misidentifications or are not self-ascriptions, and hence do not refute Immunity. However, the pathologies show that it is possible to be introspectively aware of a mental state without self-ascribing it. This, in turn, seems to challenge the idea that introspection-based self-ascriptions of mental states are IEM in virtue of being identification-free. I argued that identification-freedom does not explain Immunity anyhow and presented an alternative explanation according to which immunity to error through misidentification is the result of a tight link between first-personal information channels and ownership of the states in question.

Although none of the suggested cases actually succeeds as a counterexample to Immunity, the analysis reveals what actual counterexamples must look like. Let’s focus on introspective self-ascriptions of mental states. A judgment refutes Immunity if and only if the judgment is a self-ascription of a mental state, the judgment is based on
introspective awareness of that state, and yet that state is actually somebody else’s. Are such cases possible? Let me suggest a case that is inspired by the discussion of inserted thoughts.

Assume it were possible to actually insert thoughts into somebody else’s mind, say, by transcranial magnetic stimulation or simply by sticking an electrode into that person’s brain. A spiritual variant of this scenario is considered by Descartes in asking: “Is there not a God, or whatever I may call him, who puts into me the thoughts I am now having?” (Meditations 2[3]). Let’s assume further that the thought does not appear strange in any way or inserted to the subject, so that she will self-ascribe it. Now, depending on one’s definition of thought-authorship, there may be a good sense in which the thoughts that are put into the subject’s mind are actually authored by the one inserting these thoughts. In that case, the receiving subject would self-ascribe authorship for these thoughts, of which she is introspectively aware, but which are actually somebody else’s. Whether this case is conceptually coherent essentially depends on what it means for a thought to be one’s own — a question that has been neglected in the debate on thought insertion.  

I conclude by addressing the allegation, sometimes raised in discussions on Immunity, that the thesis is no more than a tautology. The worry can be spelled out more precisely within the framework introduced above according to which a misidentification occurs if and only if source and target come apart. Here is how Immunity may seem to be tautological: The target criterion requires that the target of the predication is the judging subject herself, and the source criterion requires that the information on which the predication is based derives from the subject herself. Hence, trivially, the judgments that satisfy these two criteria cannot involve a misidentification in the sense that source and target come apart. Immunity, the objection goes, is an instance of the following trivial conditional: When a judgment of the form ‘x is F’ is based solely on information that derives from x, then that judgment cannot be wrong in ascribing the property of being F to the wrong object.

But actually, this objection misconstrues the framework set out above. In particular, it misconstrues the source criterion. The source criterion does not state that the information has to derive from the subject herself. Rather, it specifies that the information has to be delivered by certain information channels, namely introspection and proprioception. Of course, these channels do typically deliver information that derives from the subject herself. But whether they always and necessarily do so is a non-trivial matter. The debate on cross-wiring testifies to this.

There is a non-trivial gap between a judgment’s being based on information deriving from the subject herself and the judgment’s satisfying the source criterion. The source is metaphysically characterized as the source from which the predication information actually derives. The source criterion, in contrast, is an epistemic specification of the information channels that fall within the scope of the thesis. The interesting question, then, is whether it is possible, and if so in which way, for first-personal information channels (introspection and proprioception) to deliver information that derives from another person. Answering that question means spelling out the information-ownership link in more detail, a task that is neither trivial nor tautological.

References


de Vignemont, Frédérique (2012): “Bodily Immunity to Error”. In: Simon Prosser & François Recanati (eds.): Immunity to Error through

