

# Three Questions About Immunity to Error Through Misidentification

Giovanni Merlo<sup>1</sup>

Received: 16 September 2015 / Accepted: 22 July 2016 / Published online: 4 August 2016  
© Springer Science+Business Media Dordrecht 2016

**Abstract** It has been observed that, unlike other kinds of singular judgments, mental self-ascriptions are immune to error through misidentification: they may go wrong, but not as a result of mistaking someone else’s mental states for one’s own. Although recent years have witnessed increasing interest in this phenomenon, three basic questions about it remain without a satisfactory answer: what is exactly an error through misidentification? What does immunity to such errors consist in? And what does it take to explain the fact that mental self-ascriptions exhibit this sort of immunity? The aim of this paper is to bring these questions into focus, propose some tentative answers and use them to show that one prominent attempt to explain the immunity to error through misidentification of mental self-ascriptions is unsuccessful.

When one forms a judgment, there are various kinds of errors one can make. One may judge *that someone is crossing the street* when, in reality, nobody is crossing the street. One may judge *that the person crossing the street is wearing blue trousers*, when, in reality, the person crossing the street is wearing black trousers. Or one may judge *that Paul McCartney is crossing the street*, when, in reality, some one is crossing the street, but it is not Paul McCartney. Errors of the last sort—whereby one mistakes someone for someone else—are often referred to as cases of *error through misidentification*.

It was Wittgenstein who first pointed out, in a famous passage of the *Blue Book*, that in the case of mental self-ascriptions “no error [through misidentification] is possible” (Wittgenstein 1958, 67). For example, “there is no question of recognizing a person when I say that I have a toothache” and so it is “impossible

---

✉ Giovanni Merlo  
giovanni.merlo@gmail.com

<sup>1</sup> Phlox Research Group, University of Hamburg, von-Melle-Park 6, 20146 Hamburg, Germany

that in making the statement ‘I have a toothache’ I should have mistaken another person for myself” (ibid.). From this observation Wittgenstein drew the radical conclusion that (unlike statements like ‘Paul McCartney is crossing the street’) neither ‘I have a toothache’ nor other mental self-ascriptions involve any *judgment* about a particular person: what mental self-ascriptions do is to *express* a mental condition, not to *report* it (as Wittgenstein puts it “to say ‘I have pain’ is no more a statement *about* a particular person than moaning is” (ibid.)).<sup>1</sup>

Many philosophers these days reject Wittgenstein’s expressivist account of mental self-ascriptions, but agree with the observation that motivated that account. On the resulting position, although every mental self-ascription involves a truth-evaluable judgment about oneself, the judgments in question (or, at least, a significant portion of them) exhibit some sort of *immunity to error through misidentification*: they may go wrong, but not as a result of mistaking someone else’s mental states for one’s own.<sup>2</sup>

Since Wittgenstein’s remarks were first published (1958) and Shoemaker brought the point to everyone’s attention (1968), a lot has been said and written about this phenomenon.<sup>3</sup> However, three basic questions about it remain without a satisfactory answer.

First of all, what is exactly an *error through misidentification*? Several definitions have been proposed, but, as we shall see, all of them admit of counterexamples. Without a formal definition, intuitions are all we have to go by when deciding whether this or that judgment should be described as a case of error through misidentification—not exactly an ideal situation, given that our intuitions about a number of key cases are rather shaky. It would be nice if we could extrapolate from a few central cases (where our intuitions are clear) a general definition of error through misidentification, which we could then use to classify more peripheral cases (where our intuitions are not so clear).

The second question is: what is exactly *immunity to error through misidentification*? Wittgenstein speaks of the “impossibility” of a misidentification, and so do many other philosophers. But (even bracketing the fact that error through misidentification hasn’t been properly defined) it has not been made explicit what kind of impossibility is supposed to be at stake in this context: is it the impossibility that a particular judgment involve any misidentification (*de re* impossibility) or, rather, the impossibility that any judgment possessing certain features involve any misidentification (*de dicto* impossibility)? Moreover, is it really just the impossibility of a misidentification (in either the *de re* or the *de dicto* sense) that Wittgenstein was concerned with, or is there something more to the phenomenon he identified?

The third question is: what does it take to *explain* the fact that certain judgments are immune to error through misidentification? Different (non-expressivist) accounts have been provided of this phenomenon, but the sense in which they purport to *explain* it remains unclear, not only because the phenomenon itself hasn’t

<sup>1</sup> See Wright (2000) for a defence of this interpretation.

<sup>2</sup> I will hereafter reserve the term ‘mental self-ascriptions’ for the judgments themselves, rather than their linguistic expressions.

<sup>3</sup> Recent years have witnessed increasing interest in this topic. See, for instance, Prosser and Recanati (2012).

been properly characterized, but also because the criteria that an account should meet in order to count as explanatorily successful haven't been fully spelled out.

In this paper, I will bring these questions into focus, propose some tentative answers and use them to cast doubt on one prominent attempt to explain the datum that Wittgenstein was puzzled by—the so-called “Simple Account” or “Simple Explanation” of immunity to error through misidentification. My overall aim is twofold. First, to suggest that we lack as yet a fully convincing alternative to Wittgenstein's (admittedly, unconvincing) way of dealing with the datum. Second, to put forward a framework in which the search for such an alternative could be more effectively pursued.

What the alternative might look like is a question on which I will remain neutral. The broader significance of immunity to error through misidentification has been a matter of debate between what one might call ‘deflationists’ and ‘inflationists’—the former arguing that an adequate account of this phenomenon requires no more than a proper understanding of the epistemology of mental self-ascriptions; the latter using Wittgenstein's observations to motivate substantive (and, sometimes, revisionary) views about, e.g., the semantics of ‘I’ or the metaphysics of the self.<sup>4</sup> While the deflationist camp includes several proponents of the account I will criticize in the last section of this paper, my intention is not to take sides in the debate between ‘inflationists’ and ‘deflationists’. Insofar as there is a special feature of mental self-ascriptions which makes them interestingly different from otherwise similar judgments in other domains, it seems to me that this fact might, in principle, have significant implications for our general conception of the mental—not just in epistemology, but also in metaphysics and the philosophy of language and psychology. In order to be able to work out these implications, however, we need to get sufficiently clear on what the feature in question is. It is mainly to this preliminary and clarificatory task that my paper aims to contribute to.

## 1 What is an Error Through Misidentification?

Intuitively, an error through misidentification (EM, hereafter) occurs when, in forming a judgment that is otherwise correct, someone *misidentifies* something with something else or (perhaps more neutrally) *mistakes* something for something else. For instance, consider again the example we started with:

*Paul McCartney* – I look at the man who is crossing the street. The man looks impressively like Paul McCartney, so I think to myself that *Paul McCartney is crossing the street*. But I am wrong: although someone is crossing the street, it is not Paul McCartney.

In this case, one would ordinarily say that I *misidentified* the guy who is crossing the street with Paul McCartney, or that I *mistook* him for Paul McCartney. Our intuitive grip on these notions is reasonably good (at least, it is as good as our intuitive grip on other notions that are of interest to epistemologists, such as *evidence* or

<sup>4</sup> ‘Inflationist’ arguments can be found in Anscombe (1975) and McDowell (1998). Contemporary ‘deflationists’ include, among others, Coliva (2006, 2012), Morgan (2012) and Wright (2000).

*justification*). In looking for a formal definition, it is these intuitive notions that we are trying to precisify: we want our definition to be specific enough to capture what is distinctive of the mistake I commit when I judge *that Paul McCartney is crossing the street* when someone else is, but also broad enough to encompass the full range of judgments whose ‘epistemic badness’ is sufficiently similar to that of my judgment about Paul McCartney to justify the use of a single theoretical concept.

With these simple criteria in mind, let us look at some recent attempts to define EM. A good starting point for our discussion is the definition of EM offered by François Recanati in his recent “Immunity to Error through Misidentification: What It Is and Where It Comes From” (2012). Recanati suggests that EM occurs when “a subject S judges that some object *a* is F, because S has grounds for believing that some object is F and wrongly believes that *a* is one such object”, viz. “an object satisfying ( $\lambda z$ ) (S has grounds for believing that *z* is F)” (Recanati 2012, 180). More schematically:

- (Recanati) A judgment J is an error through misidentification if and only if:
- (i) J is a singular judgment that some object *a* is F;
  - (ii) the author of J has grounds for believing that some object is F;
  - (iii) the author of J wrongly believes that she has grounds for believing that *a* is F.<sup>5</sup>

There are, I think, several problems with this definition, the decisive one being that Recanati’s conditions are not *necessary* for a judgment to be a case of EM. First of all, it seems to me that, contrary to what condition (iii) demands, a judgment can be a case of error through misidentification even if its author has no second-order beliefs about her grounds for making the judgment. Secondly, even if the author of the judgment were to form some second-order beliefs, these would not have to be *wrong* for the judgment to be a case of EM. To see this, consider again my judgment in *Paul McCartney*. Plausibly, I have plenty of good grounds for believing that Paul McCartney is crossing the street: I’m looking at the guy over there, he is crossing the street and he looks impressively like Paul McCartney. So if I were to form a second-order belief to the effect that I have good grounds for forming my judgment, my second-order belief would *not* be wrong.

One may suggest that the mistake in Recanati’s definition is the following: what’s distinctive of EM is *not* the fact that author of the judgment wrongly believes herself to have grounds for the judgment, but rather the fact that the grounds she has for the judgment are somehow misleading. For example, in the Paul McCartney case, I do have grounds for making my judgment, but the grounds in question are also grounds for believing the false proposition *that the guy I’m looking at is Paul McCartney*, or, more simply, *that that guy is Paul McCartney* (where “that guy” demonstratively picks out the guy I’m looking at). Building on this suggestion, one might be led to define EM in terms of Pryor’s (1999) notion of ‘*de re* misidentification’:

<sup>5</sup> Notice that condition (iii) is equivalent to Recanati’s requirement that the author of the judgment S believe the object to have the property “( $\lambda z$ ) (S has grounds for believing that *z* is F)” (Recanati 2012, 180).

- (Pryor-1) A judgment J is a case of *de re* misidentification if and only if:
- (a) J is a singular judgment that some object a is F;
  - (b) the author's grounds for J are grounds for believing that some object b is F and that b is identical to a;
  - (c) however, unbeknownst to the subject, a and b are not identical.<sup>6</sup>

This definition applies smoothly to my judgment in *Paul McCartney*: although my grounds are grounds for believing *that that guy is Paul McCartney*, that guy is *not* Paul McCartney and this seems to capture the sense in which my judgment—being based on grounds that justify a false identity claim—is an EM.

As Pryor himself noticed, however, some cases make trouble for *de re* misidentification:

*The skunk* – I smell a skunky odor and see several animals rummaging around in my garden. Approaching closer and sniffing, I form the belief, of the smallest of these animals, *that it is a skunk in my garden*. The belief is mistaken. There are several skunks in my garden, but none of them is the small animal I see.<sup>7</sup>

Consider my judgment or belief, of the smallest of the animals rummaging in my garden, that it is a skunk. The evidence I have for making that judgment does not seem to include or justify any identification of the form 'b is identical to a', contrary to what condition (b) demands: for my judgment to be justified, there need not be any particular skunk such that I can reasonably believe, of the smallest of the animals rummaging around in my garden, that it is identical with *that skunk*.<sup>8</sup> Despite this, it seems natural to say that my judgment is an EM of some sort: after all, the judgment is wrong and it is wrong because I mistook the smallest of the animals in my garden for one of the skunks in my garden. Based on this kind of considerations, Pryor argues that not all instances of EM are also cases of *de re* misidentification and, along with many others, I agree with him on this point.<sup>9</sup>

However, Pryor suggests that *de re* misidentification constitutes at least an interesting *subspecies* of EM, and this is where I think I want to part company with him. The crucial point is that for a judgment *that a is F* to qualify as a case of *de re* misidentification there doesn't even have to be *something* which is F. Hence cases like the following count as cases of *de re* misidentification:

*The Lion King* – An otherwise very trustworthy person tells me that Al Pacino and Harrison Ford are the same person. She also tells me that Harrison Ford starred in "The Lion King". On this testimonial basis, I come to justifiedly (and yet wrongly) believe *that Al Pacino starred in "The Lion King"*.

<sup>6</sup> See Pryor (1999, 274–275).

<sup>7</sup> The example is reported, almost verbatim, from Pryor (1999, 281).

<sup>8</sup> In fact, Pryor makes the stronger point that I might not yet be in a position to hold any *de re* beliefs whatsoever about any of the things I'm smelling (1999, 282).

<sup>9</sup> For a related case, involving misidentification of location without *de re* misidentification, see also Pryor (1999, 297).

My judgment *that Al Pacino starred in “The Lion King”* satisfies all the criteria for being a case of *de re* misidentification (it is singular and the grounds on which it is based are grounds that justify a false identification). But I’m inclined to think that it should *not* be classified as an EM. In any standard case of EM, the author of the judgment, while wrong about the fact that such-and-such individual is so-and-so, is at least right about the general fact that *some* individual is so-and-so—that’s why we say that EM occurs when someone is only wrong about *which* person or thing is so-and-so.<sup>10</sup> Now, since “The Lion King” is an animation movie, *nobody* starred in it (or so, at any rate, we may suppose). This means that my judgment *that Al Pacino starred in “The Lion King”* cannot be described as the result of mistaking the person who starred in “The Lion King” for Al Pacino: there is no such person. But if there is no such person, there is no real *misidentification* going on—so why talk of error *through misidentification*? Sure enough there is a false identification in the background (the proposition *that Al Pacino is Harrison Ford*). But the label “error through misidentification” should not mislead us into thinking that the presence of a false identification in the background of a judgment is, all by itself, sufficient to make it a case of EM. Committing what Pryor calls ‘*de re* misidentification’ is, in my opinion, neither necessary *nor* sufficient for committing an EM.

Pryor defines another notion of misidentification, which doesn’t suffer from the same difficulty. He calls this other notion ‘which-object misidentification’ (or ‘wh-misidentification’ for brevity):

- (Pryor-2) A judgment J is a case of wh-misidentification if and only if:
- (I) J is a singular judgment that some object a is F;
  - (II) The evidence supporting J puts the subject in a position to know that something is F;
  - (III) J is false.<sup>11</sup>

<sup>10</sup> The basic idea is that, in a case of EM, “one reaches for the *right* predicate, but misidentifies the object in question when applying it” (McGlynn 2016, 25; my emphasis). Against this, it might be objected that committing an EM is compatible with making errors of other sort, as in the following variation of *Paul McCartney*, which appears to involve *both* an EM *and* an error of mispredication:

*Paul McCartney\** – A look-alike of Paul McCartney is crossing the street. He is wearing blue shorts which, in this light, look black. Seeing him, John thinks to himself *that Paul McCartney is wearing black shorts*.

But the objection seems to me to overlook a subtle distinction between *being* and *involving* an EM. I agree John’s judgment *involves* an EM (plausibly, in judging *that Paul McCartney is wearing black shorts* John judged also *that Paul McCartney is wearing shorts*, and the latter certainly *is* an EM). Still, I don’t think we should say that John’s judgment *is* an EM, because that would suggest (falsely) that there’s a person wearing black shorts that John mistook for Paul McCartney. Arguably, the notion of a judgment *being* an EM (which constitutes the focus of this section) can be used to define the notion of a judgment *involving* an EM: roughly, a judgment J *involves* an EM if and only if, in forming J, one also forms a judgment J’ (possibly distinct from J) which *is* an EM.

<sup>11</sup> See Pryor (1999, 282). Pryor suggests a “small amendment” to condition (II), but the amendment does not matter too much for my present purposes. Instead of speaking of evidence that “puts the subject in a position to know” that p, Pryor speaks of grounds “offering knowledge” that p, where “certain grounds G you have ‘offer you knowledge that p’ iff p is true and, in virtue of having grounds G, you satisfy all the other conditions for knowing that p, except for those conditions that require you to believe that p” (Pryor 1999, 281). I adopt the former construction because it bears an immediate connection with the more familiar notion of “being in a position to know” (discussed, among others, by Williamson (2000, 95)).

My judgment in *The Skunk* doesn't count as a case of *de re* misidentification, but it does count as a case of wh-misidentification: it is a judgment, of a certain animal, that it is a skunk in my garden and, while it is a false judgment, the evidence supporting it puts me in a position to know that there is some skunk in my garden. On the other hand, my judgment in *The Lion King* (which counted as a case of *de re* misidentification) does not count as a case of wh-misidentification, because the evidence on which my judgment *that Al Pacino starred in "The Lion King"* is based does *not* put me in a position to know that *someone* starred in "The Lion King". All this is just as it should be.

But there are problems with the notion of wh-misidentification, too. First of all, this notion fails, once again, to cover all the intuitive cases of EM. Consider the following example:

*The red shirt* – I look at the man who is crossing the street and think to myself 'Paul McCartney is wearing a red shirt'. As a matter of fact, the man I'm looking at is not Paul McCartney. As it happens, Paul McCartney *is* wearing a red shirt at the time of my judgment.

Notice that my judgment *that Paul McCartney is wearing a red shirt*, being true, involves no wh-misidentification. Now, this is fair enough as far as it goes (after all, one might insist that what is true shouldn't be called an *error* through misidentification). Still, it's clear that the 'epistemic badness' displayed by my judgment *that Paul McCartney is wearing a red shirt* makes it pretty much *akin* to my judgment in *Paul McCartney*. And it is also clear that, to the extent that the judgments expressed by the statements Wittgenstein talks about ('I have a toothache', 'I am in pain', etc.) are immune to error through misidentification, they are also immune to the 'epistemic badness' in question. This suggests that, unlike the notion of wh-misidentification, a good working notion of EM should be broad enough to encompass my judgment in *The Red Shirt*.

But the problems don't end here, because, in other respects, the notion of wh-misidentification is also too broad. To see this, reflect on the following example:

*The oak-tree* – Walking in a forest, I come across two trees, A and B. My body of evidence includes two propositions: I see *that B is an oak-tree* and I notice *that A looks pretty much like B*. Based on such evidence, I conclude *that A is an oak-tree*. But A is not an oak-tree.

I think you'll agree that my judgment *that A is an oak-tree*, though wrong, is not a case of EM: it's true that I arrived at the wrong judgment by comparing A to B, but it is not as if I mistook A for B or viceversa. Notice, however, that both the proposition *that B is an oak-tree* and the proposition *that A looks pretty much like B* form part of the evidence I have for believing that A is an oak-tree. Given that I know both of these propositions (and, in particular, the first), it is therefore true, in this case, that the evidence I have for the judgment *that A is an oak-tree* puts me in a position to know the general fact *that something is an oak-tree*. So my judgment that

A is an oak-tree satisfies all the requirements for being a case of wh-misidentification.<sup>12</sup>

Where do we go from here? What I want to do is suggest two amendments to Pryor's notion of wh-misidentification. The first amendment concerns condition (III). Cases like *The Red Shirt* can be taken to show that a judgment need not be false to exhibit the 'syndrome' of misidentification. But if it is not falsity, what is the bad-making feature of the syndrome? One might be tempted to say that the relevant judgments, though perhaps true, are not justified. But, in any typical case of EM, the subject *is* justified in forming the judgment he or she forms (this was one of the chief reason to reject Recanati's definition of EM). So the bad-making feature is not lack of justification either. One natural possibility remains: the landmark of EM is not lack of truth or lack of justification, but rather lack of knowledge. My judgment *that Paul McCartney is wearing a red shirt* may well be true, but it doesn't amount to knowledge: it's not Paul McCartney that I'm looking at, so the perceptual evidence I have does not put me in a position to know that Paul McCartney is wearing a red shirt (and this is so even if, as it happens, it is true that Paul McCartney is wearing a red shirt). In this respect, my judgment in *The Red Shirt* deserves to be put on the same level as my judgment *that Paul McCartney is crossing the street*: both are cases in which I fail to know the content of my judgments in a distinctive way. So the first suggestion I want to put forward is the following: instead of requiring that a judgment be false in order to count as a case of EM, we should just require that it fail to constitute knowledge.

The second amendment concerns condition (II). Cases like *The Oak Tree* argue for a qualified reading of this condition. There's certainly a viable conception of evidence on which both my belief *that B is an oak-tree* and my belief *that A looks pretty much like B* form part of the evidence I have for believing that A is an oak-tree. However, there seems to be also a notion of *minimal evidence* on which neither of these two beliefs enter into the minimal evidence I have for my (wrong) singular judgment about A, because their content partially outstrips what is strictly speaking *relevant* for my coming to (justifiedly) form that judgment. If we think of evidence in terms of propositions and think of proposition in terms of sets of possible worlds, we can think of the minimal evidence a subject has for making a certain judgment as the *largest* set of possible worlds that justifies the subject in making that judgment. If we conceive of propositions as more fine-grained than sets of possible worlds, we could think of the minimal evidence for a judgment as the *weakest* proposition that provides the subject with good grounds for that judgment (where for a proposition to be weaker than another it will be sufficient, but not necessary, that it hold true under a larger set of possible circumstances). I offer these as possible ways of glossing a notion that (for present purposes, at least) I am happy to treat as a primitive and on which we have—I think—a reasonably good intuitive grip. For example, I take it that, intuitively, the minimal evidence I have for thinking that A is an oak-tree doesn't involve anything about B. Rather, it reduces to something along the following lines: *A looks pretty much like what an oak-tree would look like (if there were any oak trees)* or—better, perhaps—*A looks oak-tree-ish*. Now, clearly the mere observation *that A looks oak-tree-ish* does not justify me in believing that there are any oak-tree candidates other

<sup>12</sup> Thanks to Sven Rosenkranz here.



than A.<sup>13</sup> So, since A is *not* an oak-tree, the minimal evidence I have for believing that A is an oak-tree does *not* put me in a position to know that something is an oak-tree. And this seems to be why *The oak-tree* is *not* a case of error through misidentification. So my second suggestion is that, for a judgment to count as a case of EM, we should require that the *minimal evidence* (and not just any kind of evidence) supporting the judgment put its author in a position to know the relevant existential generalization.<sup>14</sup>

If we implement the two amendments I've suggested, the definition of EM we arrive at is the following:

A judgment J is an error through misidentification iff:

- J is a singular judgment that some object a is F;
- the minimal evidence supporting J puts the subject in a position to know that something is F;
- J fails to constitute knowledge;

This definition classifies correctly all the cases discussed in this section, striking a good balance between specificity and broadness.<sup>15</sup> Needless to say, those who feel nervous about calling 'errors' judgments that are perfectly correct (e.g. my judgment in *The Red Shirt*) can reject my amendment to condition (III) and work with a more narrow notion of error through misidentification *strictu sensu*:

A judgment J is an error through misidentification *strictu sensu* iff:

- J is a singular judgment that some object a is F;
- the minimal evidence supporting J puts the subject in a position to know that something is F;
- J is false;

<sup>13</sup> Perhaps, if certain versions of externalism about mental content are true, the observation that I possess the concept of something looking 'oak-tree-ish' would justify me in believing that there actually are some oak-trees. But of course the observation that something looks oak-treeish does not justify me in believing that I possess the concept of something looking oak-tree-ish. So the observation that something looks oak-tree-ish does not justify me in believing that there actually are oak-trees, even if externalism about mental content is true.

<sup>14</sup> Importantly, saying that the minimal evidence M supporting a judgment that *p* puts the subject in a position to know that *q* is *not* saying that there is a way of supplementing M with further evidence M' such that the combination of M and M' would no longer justify the subject in believing that *p* while still justifying him or her in believing that *q*. The minimal evidence I have for believing that Paul McCartney is crossing the street offers me knowledge that someone is crossing the street, but (for all we are allowed to assume about the case) it might also make me so confident that it is precisely Paul McCartney who is crossing the street that any additional evidence to the contrary would leave me uncertain whether anybody at all is crossing the street (this would be so if, say, the probability of my being victim of a hallucination conditional on *e+* were greater than the probability of there being someone who is crossing the street conditional on *e+*, where *e+* is the combination of my minimal evidence with the piece of information that it is not Paul McCartney who is crossing the street). Reflection on this point suggests that the conditions spelled out by Pryor (1999, 284) might not be strictly speaking necessary for a judgment to be vulnerable to misidentification.

<sup>15</sup> Notice that the definition I'm proposing (like the other definitions of EM examined in this section) implies that only singular *de re* judgments can be cases of EM. This requirement strikes me as relatively innocuous, given that what we are ultimately interested in is the immunity to EM of mental self-ascriptions and mental self-ascriptions are generally deemed to be singular first-person judgments. At any rate, it's easy to see how the definition could be modified so as to allow for cases of EM that are neither singular nor *de re*.

Alternatively, one can reserve the label of ‘error through misidentification’ for the notion just defined, and describe the type of case defined above as one of ‘ignorance through misidentification’.<sup>16</sup> Once the distinction is clear, the terminology doesn’t matter too much.

## 2 What is Immunity to Error Through Misidentification?

As we’ve seen at the beginning, it was Wittgenstein who first pointed out that, in the case of mental self-ascriptions, “no error [through misidentification] is possible” (Wittgenstein 1958, 67). And it is first and foremost the impossibility of any misidentification that Shoemaker (1968) had in mind when he invented the label of ‘immunity to error through misidentification’.<sup>17</sup> Now that we know what an error through misidentification is (or at least have a credible hypothesis as to what it is), we should try to get clearer on the nature of this impossibility.

Let us begin by reminding ourselves that, like all other modal claims, impossibility claims can be *de dicto* or *de re*. It is one thing to say that it is impossible that a morally perfect being act wrongly (*de dicto* impossibility) and it is another thing to say that it is impossible that God act wrongly (*de re* impossibility). So consider a situation in which I judge *that I’m thirsty* on the basis of introspection and consider Wittgenstein’s claim that, in this and other cases, “no error [through misidentification] is possible”. Is this supposed to imply that my judgment couldn’t have been an error through misidentification (*de re* impossibility)? I suppose not. My judgment *that I’m thirsty* can be seen as a particular mental episode or event, but when we consider that judgment in the context of discussing immunity to error through misidentification (IEM, hereafter) we are not interested in what could (or could not) have been true of that mental episode in various counterfactual circumstances. In particular, we are not interested in whether or not that mental episode could have been a case of error through misidentification had it had completely different features than the ones it actually has.<sup>18</sup> Presumably, then, the kind of impossibility that underlies IEM is some kind of *de dicto* impossibility. More precisely, the point must be that there’s some kind  $\Phi$  that my judgment *that I’m thirsty* actually exemplifies such that:

(*Impossibility*) Necessarily, for any judgment J, if J is of kind  $\Phi$ , then J is not a case of error through misidentification.

<sup>16</sup> The label was suggested to me, in conversation, by Aidan McGlynn.

<sup>17</sup> “The statement ‘I feel pain’ [is immune] to error through misidentification [...]: *it cannot happen* that I am mistaken in saying ‘I feel pain’ because, although I do know of someone that feels pain, I am mistaken in thinking that person to be myself” (Shoemaker 1968, 557).

<sup>18</sup> An obvious reply is that that it shouldn’t be assumed without argument that my judgment could have had completely different features than the ones it actually has. But, even so, the point remains that questions of *de re* modality seem completely off topic in this context. One could be an *haecceitist* (i.e. embrace the view that any entity can possess radically different properties in different possible worlds) and yet recognize a sense in which my judgment *that I’m thirsty* is immune to EM. Symmetrically, one can be an *essentialist* (i.e. embrace the view that every feature of any entity is essential to it) and yet recognize that certain judgments which do not happen to be cases of EM are nonetheless not *immune* to EM.

Notice that *Impossibility* is formulated as a schematic principle: ‘ $\Phi$ ’ is a schematic letter that needs to be replaced by a kind term to obtain a true (or false) claim. In this way, no assumption is made about the nature of the kind in question, leaving room for different hypotheses for why, in the case of my judgment *that I’m thirsty*, “no error [through misidentification] is possible”.

Now, I said that when Shoemaker invented the label of ‘immunity to error through misidentification’ what he had in mind was, first and foremost, the impossibility of a misidentification. But if it is to be a distinctive feature of mental self-ascriptions (and, perhaps, other selected categories of judgments), immunity to error through misidentification cannot consist merely in the impossibility of a misidentification. To see why, consider ERROR-FREE, the kind including all and only the judgments that are not cases of error through misidentification:

x belongs to ERROR-FREE iff x is a judgment that is not an EM

Notice that ERROR-FREE satisfies *Impossibility*: necessarily, if a judgment belongs to ERROR-FREE, it is not a case of EM. But is this a good reason to speak of ‘immunity to error through misidentification’? It seems not. For example, suppose that, seeing my friend Fred walking along the river, I judge *that Fred is walking along the river*. Even if my judgment is correct and belongs to ERROR-FREE, one wouldn’t say that it is “immune” to error through misidentification in the same sense in which the judgments I form about various aspects of my mental life are. In fact, my judgment *that Fred is walking along the river* does not seem to be “immune” to error through misidentification in *any* philosophically interesting sense (and this is not just because the judgment, regarded as a particular mental episode or event, belongs to ERROR-FREE only contingently—presumably, we wouldn’t describe it as “immune” to error through misidentification, even if we believed that all the properties of a judgment, including its being error-free or not, are essential to it).

Here’s another illustration of the same point. Consider KNOWLEDGE, a kind defined as follows:

x belongs to KNOWLEDGE iff x is a judgment that amounts to knowledge

Again, KNOWLEDGE satisfies *Impossibility*: necessarily, if a judgment belongs to KNOWLEDGE, it is not a case of EM. But we would not say that, whenever a judgment constitutes knowledge, it is “immune” to error through misidentification in the same sense in which the judgments I form about my mental life are. In fact, many judgments that constitute knowledge do not seem to be “immune” to error through misidentification in any philosophically interesting sense (and this is not just because they belong to KNOWLEDGE only contingently—presumably, we wouldn’t describe them as “immune” to error through misidentification, even if we believed that all the properties of a judgment, including whether or not it amounts to knowledge, are essential to it).

These examples seem to me to point in the same direction: for there to be IEM more is needed than just the satisfaction of *Impossibility*. What is the additional ingredient? The suggestion I want to put forward is inspired (though not explicitly suggested) by the following passage from Campbell (1999):

A judgment like “Bill spoke”, when made on an ordinary basis, is subject to error through misidentification, in that *you could have a ground for doubt about the correctness of the judgment* which did not undermine your right to claim to know, on that basis, the existential proposition, “Someone spoke”. (Campbell 1999, 89; my emphasis)

Part of what Campbell says in this passage is that, in the case of many ordinary judgments, the author of the judgment cannot rule out the possibility of his or her judgment being a case of misidentification. The implicit suggestion is that this does *not* happen in any paradigmatic case of immunity: where there is immunity to error through misidentification *the author of the judgment* (and not just any well-informed onlooker) *can* rule out that possibility. Phenomenologically, the suggestion strikes me as rather plausible: in judging *that I’m thirsty*, I am not completely sure to be right (after all, sometimes one thinks one is thirsty even when one is not thirsty), but—in a yet to be clarified sense—I can ‘rest assured’ that I am not committing an error *through misidentification*, (it would make hardly any sense *for me* to wonder whether I’m mistaking someone else’s thirst for mine). The idea, then, is that IEM doesn’t consist just in the impossibility that an error through misidentification might occur, but involves also some sort of *epistemic assurance*, on the part of the author of the judgment, against such a type of error.

What does this epistemic assurance consist in exactly? As a first stab, one might try to define it as follows:

(*Assurance-1*) Necessarily, for any judgment J, if J is of kind  $\Phi$ , then the author of J is in a position to know, while he or she makes J, that, in so doing, he or she is not committing an error through misidentification.

But there are at least two problems with *Assurance-1*. The first and most obvious is that one should not expect every author of a paradigmatic immune judgment to possess the concept of an EM (let alone to have all the right philosophical beliefs about EM). This means that—on a fairly natural understanding of ‘being in a position to know’—the author of a paradigmatic immune judgment might *not* be in a position to know that her judgment is not a case of EM. The second problem is that assurance against error through misidentification seems to depend (in part, at least) on our capacity to know certain features of our own judgments on the sole basis of introspection. This suggests that the presence or absence of epistemic assurance against EM should be made to depend only on cases where the subject knows what introspection allows us to know about our own judgments—cases of blatant self-ignorance should be screened out as irrelevant. Taking these two points into account, we arrive at:

(*Assurance*) Necessarily, for any judgment J, if J is of kind  $\Phi$ , then, if the author of J knows what introspection allows us to know about our own judgments, what he or she knows is incompatible with J being an error through misidentification.

Notice that *Assurance* is, just like *Impossibility*, a schematic principle: ‘ $\Phi$ ’ is a schematic letter that needs to be replaced by a kind term to obtain a true (or false)

claim. In this way, we leave room for different hypotheses as to the nature of the kind in question. Also, unlike *Assurance-1*, *Assurance* does not require that, whenever the judgment is of kind  $\Phi$ , the author of the judgment be in a position to know that his or her judgment is not an error through misidentification. It only requires that, whenever the judgment is of kind  $\Phi$  and the author knows what introspection allows us to know about our own judgments, he or she know something which is *incompatible* with the judgment involving an EM.

Some might object that this is too weak to solve the problem posed by ERROR-FREE and KNOWLEDGE. For example, consider again my judgment that *Fred is walking along the river*. Maybe, just as perception allows me to know, of the particular surface I am looking at, that it is *this colour* (where ‘this colour’ refers to the maximally determinate shade of colour that the surface in front of me happens to possess), introspection allows me to know, of my judgment, that *that it is thus* (where ‘thus’ refers to the maximally specific way in which the judgment I am making happens to be). But being ‘thus’ is incompatible with being an EM: if ‘thus’ expresses the maximally specific way in which the judgment I am making happens to be, nothing could be ‘thus’ and be an EM. So, even if my judgment is *not* immune to error through misidentification, introspection allows me to know something about it (namely, *that it is thus*) which is incompatible with the presence of any EM. This shows that immunity to error through misidentification cannot be reduced to the sum of *Impossibility* with *Assurance*.

I think that one can respond to this objection by drawing a distinction between two propositions being mutually *inconsistent* and two propositions being mutually *incompatible*. Inconsistency is just the contrary of compossibility: for p and q to be mutually inconsistent is just for p and q to exclude one another, i.e. to hold true in disjoint sets of possible worlds. Incompatibility—as understood here—involves more than that. When p and q are mutually incompatible, anyone who knows p and is sufficiently trained with the concepts necessary to grasp q will be able to work out, from his or her knowledge of p, that q doesn’t hold. Incompatibility is—one might say—manifest inconsistency, or something near enough. Applying this distinction to the case of my judgment that *Fred is walking along the river*, one could say the following: although what I know on the basis of introspection is *inconsistent* with my judgment being an EM (for being ‘thus’ requires, among other things, not involving any EM), it is not *incompatible* with my judgment being an EM (because not even someone sufficiently trained with the concept of an EM could work out that my judgment is error-free from the mere observation that it is ‘thus’). By contrast, when I judge that *I’m thirsty* on the basis of introspection, I seem to know something about my judgment that is incompatible, and not just inconsistent, with it being an EM—the absence of any error is fully (or almost fully) ‘in view’ for its author.

Admittedly, the notion of incompatibility at play here is somewhat vague and imprecise.<sup>19</sup> But I take it that it is clear enough for present purposes—mainly, to

<sup>19</sup> To make it more rigorous, we would have to be explicit about the resources one can draw on when one works out one piece of knowledge from another—can one appeal to some very general pieces of background empirical knowledge? Or is it only one’s conceptual competence that one can rely upon?

make *Assurance* less demanding than *Assurance-1* without rendering it entirely vacuous. Different versions of *Assurance* can certainly be distinguished, depending on how liberal or strict we want to be about what propositions should count as mutually incompatible. The suggestion I want to put forward is simply that IEM arises when a kind satisfies both *Impossibility* and (some version of) *Assurance*—some hypothesis along these lines is needed to explain why ERROR-FREE and KNOWLEDGE do not constitute cases of IEM.<sup>20</sup>

### 3 What Does it Take to Explain Immunity to Error Through Misidentification?

Now that we know (or at least have some credible hypotheses as to) what an error through misidentification is and what immunity to these errors consists in, we can finally get clearer on what it takes to *explain* various cases of IEM and, in particular, the IEM of mental self-ascriptions.

Consider any claim of the following form:

Mental self-ascriptions are immune to error through misidentification because they belong to  $\Phi$ .

Given what we said in the last section, there are two things we should require for a claim of this sort to count as a satisfactory explanation of the phenomenon discussed by Wittgenstein in the *Blue Book*. The first and most obvious is that it be true that mental self-ascriptions belong to  $\Phi$  while paradigmatically EM-vulnerable judgments don't. The second and less obvious is that  $\Phi$  satisfy both *Impossibility* and *Assurance*: since IEM involves not only the impossibility of any error through misidentification, but also the epistemic assurance of the author of the judgment against such a type of error,  $\Phi$  should satisfy the two principles at the same time.<sup>21</sup>

Those who think that IEM is, essentially, a ground-relative phenomenon might suggest a further requirement, namely that  $\Phi$  be defined in terms of the epistemic

<sup>20</sup> The case of KNOWLEDGE raises another difficulty: philosophers who think that if one knows that  $p$  then one knows that one knows that  $p$  will say that KNOWLEDGE satisfies both *Impossibility* and *Assurance*. There are, I think, three ways of reacting to this. One possibility is to say that the philosophers in question are wrong—knowing that  $p$  does not entail knowing that one knows that  $p$ —and this explains why KNOWLEDGE is not a species of IEM. A second possibility is that, intuitions to the contrary notwithstanding, KNOWLEDGE *should* be regarded a species of IEM (or, at least, should be regarded as a species of IEM by philosophers who think that knowing that  $p$  is sufficient for knowing that one knows that  $p$ ). The third possibility would be to say that, since we are interested in a notion of IEM on which a judgment can enjoy IEM and yet be false, whatever kind satisfies *Impossibility* and *Assurance* should not be defined in such a way as to exclude false judgments. Note that, since at least some ERROR-FREE judgments are false, this last move would not undercut the need to supplement *Impossibility* with *Assurance*.

<sup>21</sup> Arguably, there is also a third requirement, namely that the claim be genuinely explanatory, in particular that it be informative and non-circular: if someone were to specify  $\Phi$  as “whatever kind the judgments in question belong to that satisfies *Impossibility* and *Assurance*”, the resulting claim would hardly constitute a satisfactory account of IEM.

grounds on which mental self-ascriptions are based.<sup>22</sup> In my opinion, the suggestion should be resisted. As I already pointed out in the last section, one important advantage of formulating *Impossibility* and *Assurance* schematically is that this allows us to remain neutral on the specific nature of  $\Phi$ . When initially describing what it takes for mental self-ascriptions to exhibit IEM, we can limit ourselves to saying that there is *some* kind they belong to which satisfies *Impossibility* and *Assurance*—this much virtually every participant in the debate on IEM will agree on. It is only when explaining *why* mental self-ascriptions exhibit IEM that different hypotheses will be put forward as to *what* the kind in question is—this is where disagreement will begin. The task of characterizing the *explanandum* can thus be kept distinct from the task of finding the right *explanans* for it. Dialectically speaking, this strikes me as the right way to proceed: to assume from the start that  $\Phi$  should be defined in terms of some *epistemological* feature of mental self-ascriptions begs the question against those accounts on which something different explains their exhibiting IEM—for example, the kind of *content* they have or the *method* through which they are formed.<sup>23</sup>

In fact, once it becomes clear that IEM can be characterized *without* assuming that it is a ground-relative phenomenon, it becomes less obvious that IEM *is* a ground-relative phenomenon. What I want to do in the remainder of this paper is precisely show that one prominent attempt to explain the immunity to error through misidentification of mental self-ascriptions in purely epistemological terms—the so-called “Simple Explanation” of IEM—is unsuccessful or, at least, much more problematic than it is usually thought to be.

In its most general form, the Simple Explanation says that for a class of judgments to be immune to error through misidentification it is sufficient that they not be based on any identification:

The *Simple Explanation* just consists in the observation that a judgment will be immune to error through *misidentification* when it is not based on an *identification*. The *Simple Explanation* may not appear very exciting- or perhaps even very explanatory. However, what it says does at least seem true. [...] No one should deny the truth of the *Simple Explanation*. (Morgan 2012, 106)

<sup>22</sup> The assumption that “beliefs and judgments are immune to error through misidentification in virtue of the grounds on which they are based” (Bermudez 1998, 6) is widespread. It is implicit in the characterization of IEM offered by Pryor (1999, 485) and McGlynn (2016, 50) and explicit in the discussion of Wright (2000, 19), Coliva (2006, 415), Smith (2006, 274–275) and many others (see the contributions in Prosser and Recanati (2012)).

<sup>23</sup> Content-based accounts of IEM typically claim that mental self-ascriptions have subjectless content (see, for instance, Recanati 2012). Campbell (1999) offers a method-based account appealing to ‘dedicated ways’ of finding out about oneself. One standard argument against these accounts is that IEM is common to mental self-ascriptions and, e.g., some perceptual demonstrative judgments that are neither subjectless in content nor formed through the same method as mental self-ascriptions. However, once it is appreciated that there can be as many species of IEM as are the kinds satisfying *Impossibility* and *Assurance*, the argument loses its force—independent evidence is called for that exactly the *same* kind of IEM displayed by mental self-ascriptions is also displayed by, e.g., some perceptual demonstrative judgments.



Though hints of it can be found in Shoemaker (1968), the origins of the Simple Explanation are probably to be traced back to Gareth Evans's *The Varieties of Reference*. Evans begins by distinguishing two kinds of singular knowledge:

When knowledge of the truth of a singular proposition, 'a is F', can be seen as the result of knowledge of the truth of a pair of propositions, 'b is F' (for some distinct idea, b) and 'a = b', I shall say that the knowledge is *identification-dependent*. [...] We might say that knowledge of the truth of a singular proposition is *identification-free* if it is not identification-dependent. (1982, 180)

After showing that the distinction holds for judgments more in general, he claims that:

Clearly, [identification-free judgments] are immune to a kind of error to which [identification-dependent] judgments are liable. *Since they do not rest on any identification, they are immune to error through misidentification.* (1982, 182; my emphasis)

In its broad outline, Evans's account has been accepted by many participants in the debate surrounding the IEM of mental self-ascriptions. Implicitly or explicitly, it has been defended or endorsed by Peacocke (1999, 270), Wright (2000; 2012), Coliva (2006), Nida-Rümelin (2011), Stanley (2011), Morgan (2012), and Garcia-Carpintero (2013), among others. The reasons for this popularity are not difficult to see. The Simple Explanation is not only simple. If successful, it would provide us with a perfectly general account of IEM (one that could in principle be applied to other kinds of judgments that have been thought to be immune to EM, such as perceptual demonstrative judgements). What's more, it is "a pleasingly deflationary account" that "liberates us from any need for metaphysical or semantic extravagance in trying to account for [this] phenomenon" (Wright 2012, 255).

Suppose we call "BASIC" the kind including all and only the judgments that (in Morgan's terminology) are not "based" on any identification or (in Evans's jargon) cannot "be seen as a result of an identification" and are "identification-free". Then the Simple Explanation can be put as follows:

Mental self-ascriptions are immune to error through misidentification because they belong to BASIC

To evaluate the success of this account, we need to determine whether BASIC satisfies *Impossibility* and *Assurance*. Now, one crucial lesson we learned in § 1 is that not all cases of EM are cases where the evidence that justifies the subjects in making the judgment includes an identification: my judgment in *The Skunk* is justified by identification-free evidence and, yet, it constitutes a case of EM. This means that, when they speak of judgments that are not "based" on any identification or that are "identification-free", friends of the Simple Explanation cannot have in mind judgments that are justified by identification-free evidence—if BASIC were defined in this way, it would *not* satisfy *Impossibility* and the Simple Explanation would in trouble.



What's the right way of defining BASIC, then? One proposal often discussed in this connection is the following:

a judgment belongs to BASIC iff it is a judgment whose grounds *and background presuppositions* do not involve any identification

The notion of *background presupposition* requires some unpacking. As we know, the grounds of a judgment are nothing else than the evidence that justifies the subject in making that judgment. By contrast, the background presuppositions of a judgment are supposed to be those tacit or implicit commitments that, if called into question, would make it rational to withdraw the judgement, because it is only against the backdrop of such commitments that one took oneself to have evidence (or grounds) for the judgement. To illustrate, the content of my visual experience is the evidence I have for judging *that there is table in front of me*, whereas the proposition *that my visual apparatus is working properly* is, arguably, a background presuppositions of my judgment (if someone were to tell me that my visual apparatus is not working properly I would no longer take the content of my visual experience to support my belief *that there is a table in front of me*).<sup>24</sup>

Why adopting this narrower characterization of BASIC? The crucial thought is that “[the background] presupposition of a judgment may in a particular case include an identification. Should that presupposition fail, the judgment at issue may suffer error through misidentification even if it is [supported by identification-free evidence]” (Wright 2012, 270). My judgment in *The Skunk* is supposed to provide an illustration of this. According to Coliva (2006), if someone were to call into question the proposition *that this animal (I can now see) is the animal (in my garden) which is actually responsible for this odor I can smell*, I would be rationally obliged to withdraw my judgment. So that proposition—a proposition identifying the animal I am seeing with the animal whose odor I am smelling—should be regarded as a background presupposition of my judgment. The general hypothesis is that “error through misidentification [...] will *always* depend on the presence of a false identification component as part of [either the grounds or the background presuppositions of] a given judgment” (Coliva 2006, 417).<sup>25</sup> Define BASIC so as to exclude any identification from *both* the grounds *and* the background presuppositions of the judgment and the Simple Explanation will be safe.

I think there is much to be said in favour of this way of developing the Simple Explanation. The distinction between grounds and background presuppositions of a judgment is central to much current theorizing in epistemology, and it is certainly a nice feature of this proposal that it puts this distinction to new use, rather than responding to the problem posed by *The Skunk* in some *ad hoc* way. Nevertheless, there seem to me to be two problems with this strategy that friends of the Simple Explanation have either underestimated or entirely overlooked.

<sup>24</sup> The distinction between evidence and background presuppositions of a judgment is explicitly drawn by Coliva (2006), although (as she points out) hints of it can be found already in Pryor (2000) and Wright (2002).

<sup>25</sup> The same move can be found in Garcia-Carpintero (2013).

The first problem becomes apparent if we focus on the kind of reconstruction Coliva proposes of *The Skunk*. Consider again the proposition *that this animal (I can now see) is the animal (in my garden) which is actually responsible for this odor I can smell*. Saying that this proposition is a background presupposition of my judgment is tantamount to saying that, when I formed my judgment, I somehow presupposed (perhaps implicitly or tacitly) that there was a unique animal in my garden which was causally responsible for my olfactory experience (at any rate, that is what the definite description “*the animal in my garden which is actually responsible for this odor I can smell*” suggests, on one natural way of interpreting it). But why should I have presupposed such a thing? Could I not have presupposed (perhaps implicitly or tacitly) that my olfactory experience was collectively caused—or perhaps even overdetermined—by the odour of many distinct skunks? It certainly seems so. In fact, it seems that it would *not* have been very rational of me to presuppose that there was a unique skunk whose odor was responsible for my olfactory experience. And even if it was rational to presuppose such a thing, it would not have been very rational of me to let that presupposition play any remotely justificatory role in the formation of my judgment. But then why thinking that there actually was, among the background presuppositions of my judgment, any (true or false) proposition worth describing as an “identification”?<sup>26</sup>

As far as I can see, the problem that these questions raise is rather general and does not have anything to do with the specific way in which Coliva formulated her proposal. The challenge that the Simple Explanation originally faced was that of accounting for cases of EM where the evidence justifying the judgment is identification-free. Taking the identification to be part, not of the grounds of the judgment, but rather of its background presuppositions does not seem to be a satisfactory way of meeting that challenge. Friends of the Simple Explanation should—I think—come to terms with the idea that a judgment can be a case of EM even if its entire justificational architecture (and not just the part of it we associate with the judgment’s grounds) is identification-free. It may be possible to do this without giving up the Simple Explanation. But, at the very least, BASIC has to be defined in even narrower terms than Coliva and Wright have suggested—how much narrower it is difficult to see.

This takes me to the second problem. It seems to me that, the narrower and more sophisticated our characterization of BASIC is, the greater chances BASIC will have to satisfy *Impossibility* without satisfying *Assurance*. This is already evident if we bracket for a moment the first problem and define BASIC in the way suggested above, i.e. as including all and only judgments whose grounds and background presuppositions that do not involve any identification. The author of a judgment

<sup>26</sup> In order to sidestep these worries, one could take the background presupposition to be *that this animal (I can now see) is an animal (in my garden) which is actually responsible for this odor I can smell*. The general idea would be to relax the notion of an identification so as to include not only propositions of the form ‘*x is the y that is so-and-so*’, but also any proposition of the form ‘*x is a y that is so-and-so*’. At that point, however, it would no longer be clear that mental self-ascriptions (or, for that matter, any other singular judgments) would still be classifiable as belonging to BASIC. Arguably, the proposition *that I am an individual who has a headache* is a background presupposition of my judgment *that I have a headache*, at least in the sense that it would be rational for me to withdraw my judgment were I told from a reliable source that I am *not* an individual who has a headache. Morgan (manuscript) makes a closely related point in his discussion of the notion of a background presupposition.

does *not* typically have access to what, from the objective point of view of epistemology, can be seen as the background presuppositions of his or her judgment. In particular, when one performs a judgment one may very well not be able to tell whether or not the background presuppositions of his or her judgments include an identification (at least not in general and not if one knows nothing of epistemology). So the risk looms large that some judgments may be BASIC (in the sense defined above) *without* their authors being epistemically assured against EM.

The point can be seen most clearly in the case of my judgment *that I'm thirsty*. In § 2, I suggested that, while in making that judgment I cannot be sure to be right, something about my judgment tells me that I am not committing any error *through misidentification*: given what I know about my judgment on the basis of normal self-knowledge, there is simply no question of having mistaken someone else's thirst for mine. Now, finding out that this is so does not require much effort or concentration or reflection on my behalf: to convince myself of the fact, I need only cast a quick introspective glance at the judgment I am performing. Certainly I do not have to evaluate elaborate hypotheses about the 'background presuppositions' of my judgment. If that's what I had to do in order to rule out the possibility of being committing an EM, I doubt that I would enjoy the kind of epistemic assurance against EM that I actually enjoy. The worry, then, is that, even if being BASIC makes for the impossibility of an error through misidentification, it does not make for any kind of epistemic assurance of the author of the judgment against EM: when one makes a BASIC judgment, what one knows may well be compatible with the judgment being an EM, even if one possesses adequate self-knowledge.

An advocate of the Simple Explanation can respond to this problem in at least three different ways.

The first would be to insist that the fact that a judgment is BASIC is, indeed, 'fully in view' for its author, in the minimal sense that an ideal subject exercising self-knowledge in optimal circumstances would be in a position to know that fact. The observation would not give us a very strong kind of epistemic assurance. But perhaps it would give us epistemic assurance of some kind. The question is whether we can content ourselves with such a weak version of the principle or whether we need a more robust sense in which we—imperfect subjects operating in less-than-optimal circumstances—are epistemically assured against EM when we make various judgments about what we believe, hope, desire, feel or experience.

The second response is more subtle and trades on the idea that, whenever a judgment is BASIC, its author may not know that this is so, but will know of some *other* feature of the judgment which happens to be incompatible with the judgment's involving an EM. Thus *Assurance* against EM is guaranteed, even for imperfect subjects operating in less-than-optimal circumstances. But the proponent of this line cannot get away with the vague suggestion that some feature or another of BASIC judgments allows their subjects to rest epistemically assured against EM. If it is agreed that it is not their being BASIC, one would like to know what the features in question are—in fact, one would expect the Simple Explanation to be re-formulated directly in terms of those features.<sup>27</sup>

<sup>27</sup> Could the relevant feature be, quite simply, that the judgment is IEM? But for a judgment to be IEM is for it to belong to some kind that satisfies *Impossibility* and *Assurance*, and for judgments of a certain kind

The third and last response would be to reject the considerations I offered in § 2 for regarding *Assurance* as part and parcel of IEM. This leaves the proponent of the Simple Explanation with only two options. The first is to show that, contrary to what I've argued there, IEM can and should be reduced to the mere impossibility of any error through misidentification. On this approach, in so far as it succeeds in showing that BASIC satisfies *Impossibility*, the Simple Explanation has succeeded in explaining all there is to explain. The second option is to offer some alternative hypothesis as to the additional ingredient that, together with *Impossibility*, makes certain judgments immune to error through misidentification.

This is not the place for a detailed examination of these responses. All I want to say is that—as things stand and given the two problems I outlined here—I find little justification for the optimism surrounding the Simple Explanation. I do agree that what the Simple Explanation says “does at least *seem* true”, as Morgan puts it. But I also think that, when the notion of a judgment's being “based on” (or “resting on” or being “the result of”) an identification is subjected to closer scrutiny, this appearance of truth vanishes and the prospects of the Simple Explanation start to look much less promising.

**Acknowledgments** Thanks to Manuel Garcia-Carpintero, John Hawthorne, Guy Longworth, Aidan McGlynn, Daniel Morgan, François Recanati, Sven Rosenkranz, Assaf Weksler and Crispin Wright for their helpful comments on earlier drafts of this paper. Thanks also to the participants of the LOGOS Colloquium in Barcelona, the second HUJI Graduate Conference in Jerusalem and the 2013 NIP / Institut Jean-Nicod Workshop “Thinking of Oneself” in Paris. The research leading to this paper has received funding from the European Community's Seventh Framework Programme FP7/2007–2013 under grant agreement no. FP7-238128, and from the Swiss National Science Foundation Sinergia Project ‘Grounding – Metaphysics, Science, and Logic’ (Project 147685).

## References

- Anscombe, E. (1975). The first person. In S. Guttenplan (Ed.), *Mind and language* (pp. 45–65). Oxford: Clarendon Press.
- Bermudez, J. L. (1998). *The paradoxes of self-consciousness*. Cambridge: MIT Press.
- Campbell, J. (1999). Immunity to error through misidentification and the meaning of a referring term. *Philosophical Topics*, 26(1–2), 89–104.
- Coliva, A. (2006). Error through misidentification: Some varieties. *The Journal of Philosophy*, 103(8), 403–425.
- Coliva, A. (2012). Which ‘key to all mythologies’ about the self? A note on where illusions of transcendence come from and how to resist them. In S. Prosser & F. Recanati (Eds.), *Immunity to error through misidentification: New essays* (pp. 22–45). Cambridge: Cambridge University Press.
- Evans, G. (1982). *The varieties of reference*. Oxford: Clarendon Press.
- Garcia-Carpintero, M. (2013). Self-conception: Sosa on de se thought. In J. Turri (Ed.), *Virtuous thoughts: The philosophy of Ernest Sosa*. Philosophical studies series (Vol. 119, pp. 73–99). Springer. doi:10.1007/978-94-007-5934-3.

---

Footnote 27 continued

to satisfy *Assurance* is for them to possess some feature that is both incompatible with their being cases of EM and epistemically accessible to their authors. So to suggest the subject is epistemically assured against EM by knowing that her judgment is IEM is (roughly) to say that she is epistemically assured against EM by virtue of knowing that *something* she knows about her judgments is incompatible with its being an EM. It's hard not to think that this suggestion requires an answer to the prior question of what that ‘something’ is. Thanks to an anonymous referee for pressing me on this point.

- McDowell, J. (Ed.). (1998). Reductionism and the first-person. In *Mind, value and reality* (pp. 359–382). Cambridge: Harvard University Press.
- McGlynn, A. (2016). Immunity to error through misidentification and the epistemology of de se thought. In M. Garcia-Carpintero & S. Torre (Eds.), *About oneself: De se thought and communication* (pp. 25–55). Oxford: Oxford University Press.
- Morgan, D. (2012). Immunity to error through misidentification: What does it tell us about the de se? In S. Prosser & F. Recanati (Eds.), *Immunity to error through misidentification: New essays* (pp. 104–123). Cambridge: Cambridge University Press.
- Morgan, D. (manuscript). Thinking about the body as subject.
- Nida-Rümelin, M. (2011). The conceptual origin of mind-body dualism. In A. Coliva (Ed.), *Self and self-knowledge*. Oxford: Oxford University Press.
- Peacocke, C. (1999). *Being known*. Oxford: Oxford University Press.
- Prosser, S., & Recanati, F. (Eds.). (2012). *Immunity to error through misidentification: New essays*. Cambridge: Cambridge University Press.
- Pryor, J. (1999). Immunity to error through misidentification. *Philosophical Topics*, 26(1–2), 271–304.
- Pryor, J. (2000). The Skeptic and the Dogmatist. *Noûs*, 34(4), 517–549.
- Recanati, F. (2012). Immunity to error through misidentification: What it is and where it comes from. In S. Prosser & F. Recanati (Eds.), *Immunity to error through misidentification: New essays* (pp. 180–201). Cambridge: Cambridge University Press.
- Shoemaker, S. (1968). Self-reference and self-awareness. *The Journal of Philosophy*, 65(19), 555–567.
- Shoemaker, S. (1970). Persons and their pasts. *American Philosophical Quarterly*, 7(4), 269–285.
- Smith, J. (2006). Which immunity to error? *Philosophical Studies*, 130, 273–283.
- Stanley, J. (2011). *Know how*. New York: Oxford University Press.
- Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.
- Wittgenstein, L. (1958). *The blue and brown books*. Oxford: Blackwell.
- Wright, C. (2000). Self-knowledge: The Wittgensteinian legacy. In C. Wright, B. Smith, & C. McDonald (Eds.), *Knowing our own minds* (pp. 13–45). New York: Oxford University Press.
- Wright, C. (2002). (Anti)-sceptics, simple and subtle: Moore and McDowell. *Philosophy and Phenomenological Research*, 65, 330–348.
- Wright, C. (2012). Reflections on François Recanati's 'Immunity to error through misidentification: What it is and where it comes from'. In S. Prosser & F. Recanati (Eds.), *Immunity to error through misidentification: New essays* (pp. 247–280). Cambridge: Cambridge University Press.