Arithmetic Judgements, First-Person Judgements and Immunity to Error Through Misidentification

Michele Palmira

University of Barcelona and LOGOS

Abstract

The paper explores the idea that some singular judgements about the natural numbers are immune to error through misidentification by pursuing a comparison between arithmetic judgements and first-person judgements. By doing so, the first part of the paper offers a conciliatory resolution of the Coliva-Pryor dispute about so-called "de re" and "which-object" misidentification. The second part of the paper draws some lessons about what it takes to explain immunity to error through misidentification. The lessons are: First, the so-called *Simple Account* (see Wright 2012) of which-object immunity to error through misidentification to the effect that a judgement is immune to this kind of error just in case its grounds do not feature any identification component fails. Secondly, wh-immunity can be explained by a *Reference-Fixing Account* to the effect that a judgement is immune to this kind of error just in case its grounds are constituted by the facts whereby the reference of the concept of the object which the judgement concerns is fixed. Thirdly, a suitable revision of the Simple Account explains the *de re* immunity of those arithmetic judgements which are not

1

wh-immune. These three lessons point towards the general conclusion that there is no unifying explanation of *de re* and wh-immunity.

1 Introduction

In the *Blue Book*, Wittgenstein observes that present-tense self-ascriptions of psychological properties, such as 'I am in pain', 'I'm thinking this thought', and so on exhibit what Shoemaker (1968) has later dubbed *immunity to error through misidentification* (IEM for short, I'll use this label to refer to the property of being immune to error through misidentification as well). If I judge 'I am in pain' on first-person grounds, I can be mistaken about whether it really is a pain or an itch, but I cannot be mistaken that it is *I* who am undergoing that sensation.

One of the main bones of contention in the debate about IEM is its scope. Several philosophers (Evans 1982, McDowell 1998, Shoemaker 1968, Wright 1998, 2012) have variously argued that past-tense self-ascriptions of physical and psychological properties such as 'I was in Scotland five years ago' and 'I was edgy', self-ascriptions of bodily properties such as 'My legs are crossed', demonstrative judgements such as 'That ball is about to hit me', spatial and temporal judgements such as 'It's raining here' and 'It's raining now', second- and third-person judgements such as 'You are next to me' and 'He's a long way off', when made on the appropriate grounds, are IEM.

Discussion of the scope of immunity to error through misidentification has proven to have wide-ranging implications for longstanding debates in the philosophy of mind, such as debates about Cartesian *vs.* non-Cartesian metaphysical conceptions of the self (see Evans 1982), the semantic status of the first-person pronoun 'I' (see

Wittgenstein 1958), the existence of a distinctively first-person perspective (see Cappelen and Dever 2013), the animalist *vs.* neo-Lockean conception of personal identity (see McDowell 1998).

Recently, some authors (Coliva 2017, Wright 2012) have ventured the hypothesis that some singular judgements about the natural numbers can be IEM. This paper takes up this largely unexplored hypothesis and pursues a twofold general aim: to establish the existence of various cases of arithmetic judgements that are IEM;[1] to show how such cases should be brought to bear on the question of what it takes for judgements to be IEM.

By pursuing this twofold aim, the paper touches on several important issues revolving around IEM, chief amongst them the question of how to characterise the phenomenon of error through misidentification. As is well-known, there has been some debate about this issue. In a 1999 paper, James Pryor draws a distinction between two types of error through misidentification: what he calls "*de re*" misidentification and "which-object" misidentification (henceforth wh-misidentification). In a 2006 paper, Annalisa Coliva argues that wh-misidentification is a spurious phenomenon. By discussing whether, and if so which, arithmetic judgements are IEM I also attempt at resolving the Coliva-Pryor dispute by reconciling their respective positions.

The paper is organised as follows. In section 2 I elaborate on Coliva (2017)'s discussion of arithmetic judgements and IEM. In section 3 I focus on the Coliva-Pryor dispute about wh-misidentification. I argue that discussion of two cases of arithmetic judgements show that we can admit the existence of wh-misidentification and wh-IEM while, at the same time, retaining the main substantive points of Coliva's

---

[1] Henceforth my focus will be on singular judgements about the natural numbers, that I will call – for ease of expression – 'arithmetic judgements'.

position. In section 4 I move on to draw some lessons on what it takes to explain IEM.
I offer some concluding thoughts in section 5.

2 Arithmetic Judgements and De Re *IEM*

Coliva's discussion of arithmetic judgements and IEM revolves around the phenomenon of *de re* misidentification (see Coliva 2017). By adopting Pryor's definition, *de re* misidentification occurs when:

(dr$_i$) There is some singular proposition about x, to the effect that it is F, that a subject believes or attempts to express. […]

(dr$_{ii}$) The subject's justification for believing this singular proposition rests on his justification for believing, of some y, and of x, that y is F and that y is identical to x. […]

(dr$_{iii}$) However, unbeknownst to the subject, y ≠ x.

(Pryor 1999: 274 f.)

To illustrate this definition with an example, suppose that, upon looking in a mirror, I form the judgement that I am making a mess. Yet, unbeknownst to me, the person I perceive in the mirror is my twin sibling NN. The judgement 'I am making a mess', when based on perceptual grounds, is vulnerable to *de re* misidentification.

Let us move on now to the case of arithmetic judgements. Consider the following:

(1) '4 is a square root of 16'.

Suppose that S judges (1) on the basis of number theoretical facts, viz. facts about both negative and nonnegative integers. Plausibly, S's grounds for (1) should be specified as follows:

(G1) A square root of a number n is a number m such that $m \times m = n$.

(G2) $4 \times 4 = 16$

(G3) All integers have a square root.

(G4) For any two integers m and n: $m \times n = \text{-}m \times \text{-}n$.

By discussing the similar example of '3 = $\sqrt{9}$', Coliva (2017: 247) claims that judgements like (1) are *de re* IEM. I agree with her verdict and I turn now to elaborate on it in two directions: first, I specify the structure of the grounds of (1) to make sure that it exhibits the pattern relevant to *de re* misidentification. Secondly, I focus on the question of what kind of immunity is required by the phenomenon of IEM.

We should bear in mind that in order for (1) to be *de re* IEM, it must be the case that: conditions (dr_i) and (dr_{ii}) can be met while condition (dr_{iii}) cannot. It is easy to grant that (dr_i) can be met, for (2) expresses a singular proposition. To show that (dr_{ii}) can also be met, let us add some details and suppose that S maintains that (G2) holds in virtue of the following grounds:

(G5) $\text{-}4 \times \text{-}4 = \sqrt{4}$.

(G6) $\text{-}4 \times \text{-}4 = 4 \times 4$.

(G6) is entailed by (G4). So, let us say that S's justification for (1) rests on (G1), (G3), (G4), (G5) and (G6). To clarify: this modification of the example is needed in

order for (1) to follow the *de re* misidentification pattern of judgement. Since (G5) is an instance of y is F, and (G6) is an instance of y = x, we have that (dr_{ii}) can be met, thereby preserving the pattern of judgement relevant to *de re* misidentification.

Let us now ask: can (dr_{iii}) be met? That is to say, is (G6) such that (dr_{iii}) can be met? An affirmative answer to this question makes (1) vulnerable to error through *de re* misidentification, whereas a negative answer makes (1) IEM.

To take up this question, I deem helpful to unpack the often invoked idea that IEM involves the impossibility of making a mistake in the ascription of the property to a witness while judging on the relevant kinds of grounds (see e.g. Coliva 2006: 403, Cappelen and Dever 2013: 130, Shoemaker 1968: 557). So, to establish whether (1) is *de re* IEM we can ask: are S's grounds for (1) such that they preserve the modal force implicit in the notion of immunity? More specifically: is (G6) such that it preserves the modal force implicit in the notion of immunity?

Coliva (2017: 243-4) suggests that cases such as (1) are instances of *de re* IEM since they are based on an identity component which is necessarily and a priori true. While I agree that (G6) is necessarily true, if true at all, I believe that the necessary truth of (G6) is not necessary to determine the *de re* immunity of (2), for it would give rise to a mistaken interpretation of the modal force of immunity.[2] Let me explain.

Since the work of Kripke, it has become customary to distinguish between two modal profiles of sentences and judgements: the metaphysical profile and the epistemic profile. Roughly put, to establish whether something is metaphysical possible we

---

[2] Coliva's of the "Hesperus/Phosphorus" case (2017: 243) makes it clear that the necessity of (G6) to be sufficient for *de re* IEM. However, I take it to be important to emphasise that the modal force of immunity has nothing to do with metaphysical modality in the sense that the necessary truth of (G6) is not even necessary to establish the *de re* immunity of (1).

should consider counterfactually what would have been the case had a possible world *w* obtained. By contrast, to establish whether something is epistemically possible, we should consider actually what is the case if a possible world *w* obtains.

The well-tried two-dimensional semantic framework enables us to make this idea more precise by distinguishing between two kinds of *possibilia*: *possible worlds*, which are the familiar Kripkean metaphysically possible worlds, and *scenarios*, which can be seen as maximally specific ways the world might be for all we know *a priori* (or, alternatively, as centered possible worlds, that is, an ordered triple of a possible world, an individual and a time in that world, see Chalmers 2006 for more details).

Expanding on this distinction, I contend that in order to establish whether it is metaphysically possible for S to make a mistake while judging *j* we should consider a possible world p_w as counterfactual and ask: would S be mistaken in judging *j* had p_w obtained? In order to establish whether it is epistemically possible for S to make a mistake we should consider a scenario p_s as actual and ask: if p_s is the case, is S mistaken in judging *j*?

To illustrate, take S's judgement 'Water is H_2O' and consider Twin Earth as counterfactual: had the oceans and lakes been filled with XYZ would S's have been mistaken in judging 'Water is H_2O'? The answer, Kripke and Putnam taught us, is no, for water is H_2O in all metaphysically possible worlds. By contrast, if you consider Twin Earth as actual, you have to ask: if the actual world is such that oceans and lakes are filled with XYZ, is S's mistaken in judging 'Water is H_2O'? The answer is yes, in that for all we know a priori the actual world could be such that water is XYZ.

Take:

(2) 'I am the thinker of this thought'.

Suppose that S judges (2) on introspective grounds. This is a classical example of a judgement which is IEM. If the modal force of the notion of immunity were to be equated to the metaphysical impossibility of making an error through misidentification relative to certain grounds, we would have to establish whether S could have been mistaken (in the relevant sense) in judging (2) had introspection been different. For instance, a question like the following would be relevant to establishing whether, and if so how, (2) is IEM: had introspection been a faculty such that S judges 'I am φ' (where 'φ' stand for a psychological property) on introspective basis only if S grounds her judgement on the testimony of T telling S that S is identical to the individual who is φ, could (2) have been IEM? And if so, why?

However, these questions do not bring to bear on the reason why (2) is IEM. On closer inspection, the leading question we are after while reflecting on cases such as (2) is the following: what is it about introspection – as opposed to, say, getting to know which psychological properties S instantiates on testimonial grounds (e.g. a brain scan carried out by a neurologist) – that make it the case that S' actual judgement (2) made on such grounds be about herself in such an error-less way?

It won't help addressing this question by reflecting on whether (and if so why) (2) could have been IEM had introspection worked in a very different way from how it actually does. On the contrary, taking into account the metaphysical possibility that introspection could have worked in the way suggested in the counterfactual situation depicted above seems to make the leading question irrelevant. To illustrate. In the counterfactual situation depicted above, introspection works pretty much like testimony. And yet, if introspective-based self-ascriptions of psychological properties were not supposed to be vulnerable to error through misidentification even in this

counterfactual situation in order to count as IEM, what would be the point in asking what it is that makes introspective-based and testimony-based self-ascriptions of psychological properties different?

These observations, in my view, strongly support the contention that it is not necessary for a judgement to be IEM that its grounds guarantee the absence of error in all metaphysically possible worlds. Therefore, the modal force of the notion of immunity has to be cashed out in terms of the epistemic impossibility of making a mistake.[3]

Summing up. To establish whether a judgement is *de re* IEM, we have to establish whether it is epistemically possible for S's grounds to satisfy (dr_{iii}). When the judgement's grounds do not feature any identification component, (dr_{iii}) cannot be satisfied *a fortiori*. However, the foregoing discussion shows that (dr_{iii}) cannot be satisfied even if the judgement's grounds do feature an identification component whose falsity is epistemically impossible in the sense of being ruled out a priori.

We can now get back to (1) and focus on (G6) in order to establish whether it is epistemically possible for S to satisfy condition (dr_{iii}). I contend that it is not, for (G6) is true (if at all) in all maximally specific ways the world might be for all we know a priori. To put it differently, it cannot be the case that -4 × -4 is not equal to 4 × 4 is not *a priori*. Thus, when S judges (1) on the basis of (G1), (G3), (G4), (G5) and (G6), (1) is *de re* IEM since (dr_{iii}) cannot be satisfied in the relevant modal sense required by the notion of immunity.

---

[3] I believe that the interpretation of the modal force of immunity I have argued for here is assumed by various authors in the literature. For instance, García-Carpintero (2015: 14) qualifies the impossibility of error by using the expression "counteractually, i.e., considering alternative worlds as actual". Cappelen and Dever (2013: 130) write: "[I am in pain] is *epistemically* privileged, by virtue of a certain kind of error being impossible, in the sense of a priori ruled out".

3 Wh-IEM and the Coliva-Pryor Dispute

In his 1999 paper, James Pryor distinguishes between two varieties of error through misidentification: *de re* misidentification and wh-misidentification. Wh-misidentification occurs when:

(wh$_i$) A subject has some grounds G that offer him knowledge of the existential generalization ∃x Fx [and G is independent of any justification the subject has for believing of some particular object that it is F].[4]

(wh$_{ii}$) Partly on the basis of G, the subject is also justified, or takes himself to be justified, in believing of some object a that it is F.

(wh$_{iii}$) But in fact a is not F. Some distinct object (or objects) y is F, and it's because the grounds G "derive" in the right way from this fact about y that they offer the subject knowledge that ∃x Fx. [5]

(Pryor 1999: 282)

Pryor illustrates wh-misidentifcation via the following case (1999: 281):

(SKUNK)

---

[4] Pryor does not officially add the independence condition in the definition, but he refers to it at various points (see pp. 282-3).

[5] In his 1996 Whitehead lectures, Wright characterised IEM as follows: "A claim made on a certain kind of ground involves immunity to error through misidentification just when its defeat is not consistent with retention of grounds for existential generalization" (Wright 1998: 19). This passage encapsulates Pryor's definition of wh-misidentification.

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

In (SKUNK), while what I smell allows me to being warranted in the existential judgement 'There's a skunk nearby', the distinctive odour I'm smelling still leaves it open whether the smallest of these animals I see is a skunk. Thus, I fail to correctly identify which thing it is that is a skunk in the first place, as opposed to correctly identify which thing is a skunk and then mistakenly re-identifying as the smallest of these animals I see (as it would have been the case had the judgement been a case of *de re* misidentification).

Coliva (2006) has objected to the existence of a distinctive phenomenon of wh-misidentification by arguing that (SKUNK) can be shown to exhibit the same pattern of judgement involved in *de re* misidentification. In her view, the grounds for the singular judgement 'This animal (I can now see) is a skunk' can be cashed out as follows (Coliva 2006: 412):

(i) The animal (in my garden) which is actually responsible for this odour I can smell is a skunk.

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

Since (i) is of the form 'y is F' and (ii) of the form 'x = y',[6] Coliva concludes that the 'This animal (I can now see) is a skunk' exhibits the pattern of judgement relevant to *de re* misidentification.

While Coliva's alternative analysis of (SKUNK) is – no doubt – plausible, various authors have raised doubts about the generalizability of her strategy. Consider the following two examples:

(DESERT)

I am lost in a sandy desert and, attempting to walk out, come across footprints which I misidentify as my own, concluding somewhat desperately "I am going round in circles". (Wright 2012: 256)

(ORANGES)

I peel a few oranges, separate the segments, put them on a dish and go out to buy some sugar and port to make a dessert; when I come back, all the segments have vanished. I judge someone ate my oranges, without assuming a unique culprit; several family members are around, and for all I know a few of them might have each eaten some segments. Then I see Mary with some orange peel in her hand, and judge Mary ate my oranges. (García-Carpintero 2015: 10)

Coliva's recipe for showing that condition (dr_{ii}) is satisfied is the following: introduce a rigidified definite description standing for the unique object that is suitably related

---

[6] Importantly, Coliva (2006: 413) maintains that the concept 'the animal (in my garden) which is actually responsible for this odour I can smell' that features in (ii) is a singular but not *de re* in the sense of not requiring any identifying knowledge of the objects.

to S's grounds for judging that something is F. Wright (2012 258) follows Coliva's recipe in (DESERT) and obtains:

(I) The person who has actually caused these footprints in the sand has passed this way already.

(II) The person who caused these footprints in the sand = me.

According to Wright, the fact that someone has already passed this way is the very cause of the footprints suggests that a judgement like (I) is very similar to something like 'the unique object that caused the grounds G is responsible for grounds G'. For this reason, Wright contends that (I) is "near enough, a tautology" (2012: 258). That is to say, Wright maintains that (I) is too little informative to be plausibly taken to be part of the grounds of the judgement 'I'm going round in circles'. Notice that the claim here is not that (I)-(II) could never be regarded as grounds for 'I'm going round in circles'; the point, rather, is that a more plausible specification of this judgement's grounds does not involve specifying any identification component.

(ORANGES), by contrast, is such that we can't even come up with a possible – let alone plausible – rigidified definite description guaranteeing that condition (dr_{ii}) is satisfied since it is part of the case that the grounds for the existential judgement are purely general.[7]

Still, Coliva might reply by arguing that what matters about (DESERT) is that it be *possible* – and not necessarily plausible – to reconstruct the grounds for 'I'm going round in circles' in such a way as to obtain the pattern relevant to *de re* misidentification. As for cases such as (ORANGES), she might maintain that they are

---

[7] See also Recanati (2012) and Wright (2012) for similar examples.

such that subjects make predication mistakes (see Coliva 2006 and 2017 for this suggestion).

Even granting the soundness of this twofold reply, however, I believe that there are two reasons to treat *de re* and wh-misidentification as varieties of the *same* phenomenon. First, both *de re* and wh-misidentification are such that S's grounds for the singular judgement 'a is F' are defeated and S is nonetheless warranted in judging the correspondent existential generalisation '∃x Fx'. This is definitional of wh-misidentification. As for *de re* misidentification, notice that condition (dr$_{ii}$) says that one is warranted to believe a judgement of the form 'y is F' and this clearly entails that one is warranted to believe '∃x Fx'. Relatedly, it is important to emphasise that not any mispredication results in wh-misidentification. Suppose that S judges, on perceptual grounds, 'That man is wearing a yellow coat' of somebody wearing a green coat. Clearly, S's judgement is an instance of mispredication; yet, S's perceptual grounds in no way warrant the existential judgement 'Somebody is wearing a yellow coat'. In my view, these two points jointly support taking wh-misidentification to be a genuine phenomenon of misidentification just like the *de re* case discussed in section 2.

Having introduced and supported the existence of wh-misidentification, let us ask: is (1) wh-IEM?

I contend that (1) is not wh-IEM, for (1) can be defeated without the existential judgement 'Some *x* is a square root of 16' be defeated. To illustrate. S could acquire evidence which defeats (G2). Suppose, for instance, that a great mathematician tells S that recent mathematical inquiry has established the following surprising theorem: there are some nonnegative integers which do not obey the usual definition of

multiplication. That is to say, this great mathematician provides S with defeating evidence against the following:

(G4) For any two integers m and n: $m \times n = -m \times -n$.

The piece of testimonial evidence just mentioned defeats S's grounds for judging *that 4 is a square root of 16*. To see why, notice that since (G4) is true as a matter of metaphysical necessity (if true at all), the great mathematician's testimony counts as misleading evidence. Yet, misleading evidence can nonetheless play the role of evidence which, in this case, defeats (G4). Moreover, I take it that, however strange it may be, it is certainly possible that a great mathematician wholeheartedly and reasonably – even though falsely, for the alleged proof of the result would be mistaken given the necessity of (G4) – believe that it is true that there are some nonnegative integers which do not obey the usual definition of multiplication and sincerely report this surprising result to a mathematical novice. Finally, it seems perfectly normal to regard our relying on the expert testimony of a mathematician while making a mathematical judgement as part of our grounds for that judgement.

Having clarified this, let us ask: does the piece of evidence provided by the great mathematician defeating S's grounds for judging '4 is a square root of 16' equally defeat the grounds for the existential judgement 'Something is the square root of 16'? No, it doesn't. To see why, let us go through Pryor's three conditions for wh-misidentification.

First, S is warranted in judging that something is a square root of 16 on the basis of both (G1) and (G3). So, (wh$_i$) is satisfied. Secondly, (wh$_{ii}$) is also satisfied: S judges (1) partly on the basis of the grounds for the existential judgement, in that S judges (1)

on the basis of (G1). Finally, (wh$_{iii}$) is satisfied: since the other ground S has for (1) is (G2), and since (G2) is defeated by evidence concerning the multiplication of *nonnegative* integers only, such evidence does not defeat (G1) and (G3), for they are about *both* negative and nonnegative integers. Thus, S is still justified in judging that something is a square root of 16. Importantly, the justification for the existential judgement is not packed into the piece of defeating evidence, for neither (G1) nor (G3) are justified by the surprising theorem defeating (G4). This respects Pryor's widely accepted point (Pryor 1999: 284) that wh-misidentification must be characterised by taking the relevant defeating evidence to be undercutting rather than additive.

I have argued that (1) is vulnerable to wh-misidentification. I want to conclude this section by presenting a case of wh-IEM. Consider the following (also discussed in Coliva 2017: 243):

(3) '3 is the successor of 2'.

Suppose that S judges (3) on the basis of arithmetic facts. The grounds for (3) should therefore be specified as follows:

(G1*) For every number n, the successor of a n S(n) is a number.

(G2*) For all numbers m and n, $m = n$ iff S(m) = S(n).

(G3*) For every n, S(n)=0 is false.

(G4*) 3 = the successor of the successor of the successor of the least element of the sequence of objects defined by Dedekind-Peano axioms.

(G5*) 2 = the successor of the successor of the least element of the sequence of objects defined by Dedekind-Peano axioms.

Given the meaning of 'successor', and given the place 2 and 3 occupy in the sequence of objects defined by Dedekind-Peano axioms, it cannot be the case that S has grounds for the existential judgement 'something is the successor of 2' and these grounds derive from a different object than 3. To put it otherwise, if (G1*)-(G5*) warrant the existential judgement, they can't but warrant the correspondent ascription of the property to a witness, i.e. 3. So, relative to (G1*)-(G5*), (3) is wh-IEM.

Coliva (2017: 243) discusses (3). Since she takes wh-misidentification to be a spurious phenomenon, she denies that (3) is wh-IEM. However, I believe that if we take into account and elaborate on other elements of Coliva's overall view we can resolve the Coliva-Pryor dispute on IEM in a conciliatory way and ultimately defend the contention that (3) is wh-IEM.

First, Coliva (2017: 243) takes (3) to be the analog of (2) 'I am the thinker of this thought'. She maintains that (2) is a definition of the first-person concept (or, slightly weakly, a way to individuate the first-person concept).

Secondly, on Coliva's view, if I possess the first-person concept and I am rational, whenever I deploy the first-person concept I cannot be mistaken about which person I am thinking of. Thus: The first-person concept enjoys what Coliva calls the "real guarantee":

Real Guarantee: The possession of the first-person concept guarantees that the subject knows which person that concept is a concept of. (Coliva 2003: 429).

Coliva writes (*ibid.*): "The real guarantee holds at the level of thought because the first-person concept is the concept of oneself and one cannot have it unless one knows which person one is".

Thirdly, Coliva (2017: 247) ventures the hypothesis that wh-IEM and the real guarantee can actually point towards the same feature of certain singular concepts, such as the first-person concept. However, she points out (*ibid.*) that if we took wh-IEM and the real guarantee to be identical, since all I-judgements enjoy the real guarantee we would have to accept that the judgement 'My hair is blowing in the wind' formed on perceptual grounds is wh-IEM.

As far as I can see, believers in the existence of wh-IEM should resist that perceptual-based judgements such as 'My hair is blowing in the wind' are wh-IEM, for we can clearly envisage cases in which S acquires evidence that undercuts S's grounds for the singular judgement while leaving intact the grounds for the existential one. However, this should not force us to forsake the possibility that the real guarantee and wh-IEM have something in common.

First, notice that a judgement is wh-IEM just in case S's grounds have a certain epistemic structure, i.e. they should leave no gap between the existential judgement '∃x Fx' and the singular judgement 'a is F'; or, to put it differently the possibility that such a gap be open is ruled out a priori.[8] It should also be kept in mind that S's judgement enjoys the real guarantee just in case S deploys a singular concept of the object which the judgement concerns such that S cannot have such concept unless S knows which object it is that is picked out by that concept. In my view, the kind of knowledge required in order for S to possess a concept enjoying the real guarantee

---

[8] Thus, both *de re* and wh-immunity involve a kind of epistemic impossibility that I have characterised above in terms of how things might be for all we know a priori.

exhibits the same epistemic structure that S's judgement's grounds have to enjoy in order for the judgement to be wh-IEM. That is to say, the kind of knowledge afforded by the real guarantee is such that it creates no gap between knowing the existential judgement and knowing the correspondent ascription of the property to a witness. To illustrate the point, notice that whatever puts me in a position to know the existential judgement 'There exists somebody who is thinking this thought' *guarantees* that I am in a position to know the singular judgement 'I am the thinker of this thought'. This, to my mind, suggests that the real guarantee and wh-IEM share the same epistemic roots: the kind of epistemic guarantee afforded by the real guarantee has the same epistemic structure that S's judgement's grounds must exhibit in order for the judgement to be wh-IEM.

To forestall misunderstandings, however, the common epistemic roots of wh-IEM and the real guarantee just highlighted should not lead us to take wh-IEM and the real guarantee to be identical. To see why, notice that the real guarantee is, primarily, a property of singular concepts (or, more precisely, of concept possession). By contrast, wh-IEM is a property of judgements' grounds. This also explains why we can say that a judgement such as 'My hair is blowing in the wind' made on perceptual grounds still enjoys the real guarantee but is not wh-IEM. Yet, acknowledging this fact does not undermine the point made in the previous paragraph. To put it in another way, the claim I have argued for is this: whenever S makes a judgement based on grounds that afford the kind of knowledge of which object it is that is picked out by a given singular concept guaranteed by the possession of that concept, the judgement is wh-IEM and enjoys the real guarantee.

Let us take stock. Coliva and Pryor disagree about the existence of wh-misidentification and wh-IEM. Cases such as (3) provide us with an opportunity to

resolve the Coliva-Pryor dispute in a conciliatory way. In my view, Coliva is right in emphasising that (3) is the arithmetic analog of the first-person judgement (2) 'I am the thinker of this thought'. Moreover, I agree with her that possession of certain singular concepts, such as the first-person concept, comes with a certain guaranteed knowledge – perhaps to be further unpacked in terms of acquaintance – of which object that concept is a concept of. However, Pryor and other authors defending wh-misidentification are surely onto something: both *de re* and wh-misidentification result in the fact that S keeps on being warranted in judging '∃x Fx' while being mistaken in judging 'a is F'. This fact cannot be overlooked. On the basis of these points, I suggest reconciling Coliva's and Pryor's positions as follows.

It seems correct to maintain that (G1*)-(G5*) leave no gap between the existential judgement 'Something is the successor of 2' and the singular judgement '3 is the successor of 2': any defeater that undercuts S's grounds for believing the latter *ipso facto* undercuts S's grounds for believing the former. The same holds when I judge (2) on first-person (say, acquaintance) grounds. Thus, Coliva's analogy between (2) and (3) can be maintained even if we accept the existence of wh-IEM. The grounds securing wh-IEM and the kind of knowledge-which guaranteed by the possession of the relevant singular concept is the same: the fact that possession of the concept [3][9] is such that S cannot fail to assent to the judgement (3) and the fact that when S is provided with (undefeated) (G1*)-(G5*) S can't but be warranted in judging both (3) and 'Something is the successor of 2' are two sides of the same coin.

Another reason for reconciliation comes from the analogy between the Evans-Shoemaker *senseless identification question test* (SIQT) for wh-IEM and what Coliva (2017: 235) calls a "stopping point of inquiry" (SPQ). Evans (1982) and Shoemaker

---

[9] I will use "[]" to refer to concepts.

(1968) maintain that if S judges 'I am F' on certain grounds, then the judgement is wh-IEM just in case the question 'Something is F, but am *I* F?' is senseless (relative to those grounds).[10] Coliva contends that that what is distinctive of indexical concepts such as the first-person concept is that since their possession comes with the real guarantee, they make a question such as 'Which person does have that property?' senseless. (SIQT) and (SPQ) resemble to one another, in that they both point towards the senselessness of a *which-question*, viz. a question about which object it is that S is thinking of having a given property F.

Relatedly, it is worth stressing that (3) passes (SIQT): the question 'Something is the successor of 2, but is it 3?' does not make much sense. By contrast, (1) does not pass the test, for it makes sense to ask: 'Something is a square root of 16, but is it 4?'. This lends further support to the thesis that (3) is wh-IEM whereas (1) is not.

4 Lessons on How to Explain IEM

In the previous sections I have argued for the existence of some cases of arithmetic judgements which are IEM. By discussing those cases, I have also taken up the more general question of whether IEM manifests itself in more than one way. I have done so by offering a conciliatory resolution of the Coliva-Pryor dispute on this question: my vindication of Pryor's distinction between *de re* and wh-IEM respects Coliva's

---

[10] To forestall misunderstandings: Evans and Shoemaker deployed (SIQT) long before Pryor advanced the *de re*/wh-IEM distinction, so they have never defended that (SIQT) is a test for wh-IEM in print. However, the fact that (SIQT) is a test for wh-IEM might be taken as evidence that they were after wh-IEM, but I am not here concerned with the Evans-Shoemaker dispute.

insights about the real guarantee and the analogy between judgement such as 'I am the thinker of this thought' and '3 is the successor of 2'.

I will now bring the foregoing discussion to bear on the correct explanation of IEM. I will draw three intermediate lessons that jointly support the more general conclusion that there is no unifying explanation of *de re* and wh-IEM.

4.1 The Simple Account of wh-IEM fails

The prominent *Simple Account* of IEM – inspired by Evans (1982) and recently defended in Guillot (2014) Morgan (2012), Wright (2012) – reads as follows:

> Simple Account
>
> A judgement is IEM just in case its grounds do not feature any identification component.[11]

Wright extends the Simple Account to inferentially-based judgements, such as (1) and (3). He maintains (Wright 2012: 260) that if a judgement *j* is inferentially based on general judgements and singular judgements which are themselves IEM (according to the Simple Account), then *j* is IEM.

 The Simple Account is meant to provide a unified explanation of both varieties of IEM (on this point see in particular Guillot 2014 and Wright 2012). The alleged

---

[11] Wright (2012) qualifies this definition by adding that no identification component must feature in the judgement's background presuppositions. However, I will ignore this qualification since my objection to the Simple Account is independent of whether the singular judgement upon which the target inferential judgement depends can be either in the grounds or in background presuppositions.

unifying power of the Simple Account might be regarded as particularly attractive for the following reason: since the common trait of *de re* and wh-misidentification is that S is warranted in judging '∃x Fx' while being mistaken in judging 'a is F', one might hope for a unifying account of how S's grounds should look like in order for such a situation not to obtain. As we shall see, however, these hopes will be dashed by our discussion of cases such as (1) and (3). In the remainder of this section I argue that the Simple Account of inferential wh-IEM incorrectly predicts that (1) is wh-IEM.

To begin with, notice that (1) has an identification component in its grounds, i.e. (G2). Is (G2) wh-IEM? On closer inspection, (G2) is a singular judgement which is itself grounded on the following identification component:

(G2**) 4 = the successor of the successor of the successor of the successor of the least element of the sequence of objects defined by Dedekind-Peano axioms.

So, in order to establish whether (G2) is wh-IEM, thereby establishing whether the Simple Account predicts that (1) is wh-IEM or not, we have to establish whether (G2**) is, in its turn, wh-IEM.

I submit that there is good reason to think that (G2**) is wh-IEM. Whatever grounds S has for the existential judgement 'something is the successor of the successor of the successor of the successor of the least element of the sequence of objects defined by Dedekind-Peano axioms, those grounds can't but warrant the correspondent property ascription to 4, for 4 is whatever object has that property of being the successor of the successor of the successor of the successor of the least element of the sequence of objects defined by Dedekind-Peano axioms.

This contention is further supported by the fact that (G2**) passes (SIQT). Indeed, it does not make sense to ask 'Something is the number which is the successor of the successor of the successor of the successor of the least element of the sequence of objects defined by Dedekind-Peano axioms, but is it 4?'.

Thus, since (1)'s only singular ground is (G*), and since (G2) is wh-IEM because its only identification component, i.e. (G2**), is itself wh-IEM, the Simple Account is bound to conclude that (1) is wh-IEM. And yet, as argued in section 3, this is the wrong conclusion. Therefore, the Simple Account of wh-IEM should be rejected.

At this point, one might reasonably wonder: how to explain wh-IEM?

4.2 The Reference-Fixing Account of Wh-IEM succeeds

In this section I argue that the respective immunity and vulnerability to wh-misidentification of (3) '3 is the successor of 2' and (1) '4 is a square root of 16' can be explained by a different account of wh-immunity, which I shall call the *Reference-Fixing Account*. The key insight of the Reference-Fixing Account is that a judgement is wh-IEM just in case its grounds are constituted by the facts whereby the reference of the concept of the object which the judgement concerns is fixed. In the literature, García-Carpintero (2015) has proposed a Reference-Fixing Account to capture the wh-immunity of *de se* and demonstrative judgements. So, my aim here is to expand on the key insight of the Reference-Fixing Account to show that what makes (1) wh-IEM are exactly those facts whereby the reference of numerical concepts is fixed.

Let us begin with the plausible assumption that concepts are mental representations which purport to refer to the objects they represent. Evans insightfully observes (Evans 1982: 106-7) that having the concept of a number amounts to "being able to

differentiate one number from another by their position in an infinite ordering". We can expand on Evans' point thus: by deploying a numerical concept, we purport to refer to a number in virtue of the position it occupies in the infinite ordering of objects which constitutes the sequence of the numbers. More specifically, the concept of 0 is such that it purports to refer to the object which satisfies the property of not being the successor of any number. The concept of 3 is such that it purports to refer to the object which satisfies the property of being the next to the next to the next to the least element of the infinite sequence of objects which start with 0. The same holds, *mutatis mutandis*, for other singular numerical concepts.

I hasten to clarify that the foregoing observations commit us only to the minimal claim that singular numerical concepts purport to pick out numbers in virtue of their intrinsic (order-theoretical) properties. Let me therefore forestall three possible misunderstandings:

(a)   The minimal claim does not say that, in order to uncover the *nature* of the numbers, we have to pay attention to their intrinsic properties only.

(b)   The minimal claim does not require the *existence* of numbers, in that it focuses on what it takes for a numerical concept to *purport to* refer to a number. A concept can purport to refer to a given object *o* even if *o* does not exist (and the concept does not refer).[12]

(c)   The minimal claim does not commit us to any view about the *interpretation* of arithmetic sentences.

---

[12] To give an example, suppose that S judges 'He's is chasing me' on the basis of a hallucinatory experience. S deploys the demonstrative concept [he] whereby S purports to refer to a male individual in particular without succeeding in doing so.

Let us now focus on our examples (1) and (3). While (3) is judgements in which S ascribes arithmetic properties to the number three, (1) is such that S ascribes the property of satisfying the function 'being a square root of 16', which is not squarely arithmetic in that it admits of a solution which is not part of natural numbers, to four. More importantly, while the grounds for (3) are such that they feature arithmetic facts only, the grounds of (1) feature a non-arithmetic fact, for (G1) is a fact about both nonnegative *and* negative integers and negative integers are not part of arithmetic. Thus, if we accept the minimal claim about numerical concepts, it follows that the S's grounds for (3) are exactly those facts whereby the reference of [3] is fixed, while S's grounds for (1) are such that they do not play any reference-fixing role for [4]. Hence, the reason why (3) is wh-IEM is that S's grounds for (3) are constituted by facts whereby the reference of [3] is fixed. This guarantees that there is no distinction in S's grounds between what justifies the existential generalisation that there exists the next to the next to the next to the least element of the infinite sequence of objects which start with 0 and what justifies that singular judgement that 3 is the next to the next to the next to the least element of the infinite sequence of objects which start with 0. However, (1) is not wh-IEM since S's grounds for (1) include facts which do not contribute to fixing the reference of [4]: this explains why S can be justified in making the existential generalisation without *ipso facto* being justified in making the singular judgement.

I would like now to emphasise the connections between the Reference-Fixing Account just deployed to make sense of wh-IEM and Coliva's take on the real guarantee. Bear in mind that Coliva takes judgement such as (2) and (3) to be definitions, respectively, of the concept [I] and [3]. Thus, the reference of [3] and the

reference of [I] are fixed by such judgements. The Reference-Fixing Account of wh-IEM draws on a very similar insight: it looks at how concepts are defined and maintains that when S judges *that a is F* on the basis of how concept [a] is defined, the judgement is wh-IEM. Importantly, definitions are not offered in the void, as it were. They rely on certain grounds. If we take (3) to be a definition of the concept [3], then (3)'s grounds are plausibly specified by (G1*) and (G5*). This is why, when made on these grounds, (3) is wh-IEM.

4.3 The Simple Account* of de re *IEM succeeds*

The Reference-Fixing Account is unable to explain why (1) is *de re* IEM, though. Bear in mind that (1) is *de re* IEM since (G6), i.e. '-4 × -4 = 4 × 4' is a priori true. This guarantees that (dr$_{\text{iii}}$) can't be satisfied. However, (G6) does not feature amongst the facts whereby the reference of [4] is fixed, for the reference of [4] is fixed by (G1*)-(G3*) and (G2**) alone. Hence, since there's no link between (G6) and the reference-fixing facts of [4], the Reference-Fixing Account cannot explain why (1) is *de re* IEM. This result is in line with what has emerged previously. Notice that Coliva has repeatedly (see Coliva 2003, 2017) argued that *de re* IEM and the real guarantee are different phenomena. Since *de re* IEM can't be explained by the Reference-Fixing Account, this reinforces the contention that the Reference-Fixing Account is an account of something, i.e. wh-IEM, that is a close kin of Coliva's real guarantee.

That being said, I suggest, that the following reformulation of the Simple Account is able to explain why (1) is *de re* IEM:

Simple Account*[13]

A judgement is *de re* IEM just in case: either its grounds do not feature any identification component, or its grounds feature an identification component which is a priori true.

Since (1) is grounded on the identity component (G6), and since (G6) is a priori true, it is epistemically impossible that (dr_{iii}) be satisfied. However, the Simple Account* cannot provide an overarching explanation of both *de re* and wh-IEM since the fact that (G2) and (G6) are a priori true would make it both *de re* and wh-IEM, thereby yielding – just like the standard Simple Account – the incorrect verdict that (1) is wh-IEM. Accepting a suggestion made in Coliva (2017), we might take the disjunctive nature of the Simple Account* as tracking a distinction between empirical and non-empirical judgements: an empirical judgement is *de re* IEM since its grounds do not feature any identification component; by contrast, a non-empirical judgement, such as (1), is *de re* IEM since its grounds features an identification component which is a priori true.

4.4 No Monistic Explanation: Towards a Pluralist Account *of IEM*

The discussion pursued so far establishes the following general conclusion. Since the Reference-Fixing Account explains wh-immunity but fails to explain *de re* immunity, and since the Simple Account* explains *de re* immunity but fails to explain wh-

---

[13] Again, the Simple Account* may - and perhaps should - be reformulated so as to include the judgement's background presuppositions. I ignore this for the sake of simplicity.

immunity, we can conclude that there is no single unifying explanation of the varieties whereby IEM manifests itself.

Thus, since the monistic explanatory aspirations of the Simple Account are not to be endorsed, we had better turn towards what I shall a *Pluralist Account* to the effect that different accounts need to be invoked to explain the different varieties of IEM. Schematically:

Pluralist Account of IEM

Wh-IEM is explained by the Reference-Fixing Account.

De re IEM is explained by the Simple Account*.

The Pluralist Account is a form of *Explanatory Pluralism* which has to be kept distinct from the idea that the phenomenon of (immunity to) error through misidentification comes in more than one variety. Let me explain. Let us call *Extensional Pluralism* the view that (immunity to) error through misidentification extends beyond its standard *de re* variety. Extensional Pluralism about (immunity to) error through misidentification has been endorsed by authors such as Guillot (2014), Morgan (2012), Pryor (1999), Recanati (2012), Wright (2012). In section 3, I have myself offered a vindication of Extensional Pluralism in my attempt at reconciling Coliva's and Pryor's. However, being an Extensional Pluralist about (immunity to) error through misidentification does not *ipso facto* commit one to the idea that the different varieties of IEM will have to receive different explanations. This is witnessed by the fact that Guillot (2014), Morgan (2012) and Wright (2012) all subscribe to the explanatorily monistic Simple Account. Thus, what I've called the

Pluralist Account is, to the best of my knowledge, the first formulation of an Explanatory Pluralist view of *de re* and wh-IEM.

The Pluralist Account is meant to hold in full generality. That is to say, the Reference-Fixing Account will have to be able to explain all cases of wh-IEM, whereas the Simple Account* will have to be able to explain all cases of *de re* IEM. The Simple Account*, given its disjunctive nature and following Coliva's suggestion, tells us that an explanation of *de re* IEM also comes in two varieties. In the empirical case, judgements are *de re* IEM since they have no identification component amongst their grounds. In the non-empirical case, judgements are *de re* IEM since their grounds feature identification components that are a priori true. Given, the analogies between arithmetic judgements such as (2) and (3), it is reasonable to expect that the Reference-Fixing Account will be able to deliver the correct explanation of why a certain class of first-person judgements are wh-IEM (as mentioned earlier, García-Carpintero 2015 has already offered an elaboration of the Reference-Fixing Account in this direction).

Showing how the Pluralist Account deals with most cases of IEM goes beyond the scope of this paper. However, I believe that the results of sections 4-4.4 should be welcomed as significant progress on the issue of what it takes to explain IEM, for they teach us that we should give up the unifying aspirations lurking behind the Simple Account to endorse and develop a genuinely pluralistic approach to IEM along the lines just sketched.

5 Conclusion

In this paper I have looked at the phenomenon of IEM from a largely unexplored perspective by focusing on the question of whether – and if so why – some arithmetic judgements are IEM. Following the lead of Coliva (2017), at various stages of the discussion I have compared arithmetic judgements and first-person judgements. Doing so has proven beneficial for several issues surrounding IEM. First, I have offered a conciliatory resolution of the Coliva-Pryor dispute on the existence of wh-misidentification. Secondly, I have argued that the prominent Simple Account of IEM incorrectly predicts the wh-immunity of some arithmetic judgements, such as '4 is a square root of 16'. The more general lesson to be drawn from the present discussion is that there is no single explanation of *de re* and wh-IEM. This, I submit, warrants endorsement and further development of the Pluralist Account of IEM.[14]

References

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

---

University Press.

Chalmers, D. (2006) "Two-Dimensional Semantics", in E. Lepore and B. Smith (eds.) *Oxford Handbook of the Philosophy of Language*, Oxford: Oxford University Press.

Coliva, A. (2003) "The First Person: Error Through Misidentification, the Split between Speaker's and Semantic Reference, and the Real Guarantee", *Journal of Philosophy* 100(8): 416-431.

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

Coliva, A. (2017) "Stopping Points: 'I', Immunity, and the Real Guarantee", *Inquiry*, 60(3): 233-252.

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

García-Carpintero, M. (2015) "*De Se* Thoughts and Immunity to Error Through Misidentification", *Synthese* online first, DOI: 10.1007/s11229-015-0917-y.

Guillot, M. (2014) "Identification-Free at Last. Semantic Relativism, Evans's Legacy and a Unified Approach to Immunity to Error Through Misidentification", *Teorema* 33(3): 7-30.

McDowell, J. (1998) "Reductionism and the First Person", in *Mind, Value and Reality*

(pp. 359-82), Cambridge Mass. Harvard 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.): pp. 104-123.

Prosser, S. and 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: 271-304.

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

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

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

Wright, C. (1998) "Self-Knowledge: The Wittgensteinian Legacy", in C. Wright, B.C. Smith, and C. Macdonald (eds.) *Knowing Our Own Minds* (pp. 13-45), Oxford: Clarendon.

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.): pp. 247–280.