Philosophers' Imprint

VOLUME 19, NO. 20
JUNE 2019

The Essential Non-Indexical

Léa Salje

University of Leeds

§1. Introduction

Lewis's two Gods knew exactly which world they were in, but they didn't know everything.[1] Neither God knew *which God they were*, truths they would have canonically expressed using words like 'I' or 'me'. We're lucky; we aren't like Lewis's Gods. We do know things we canonically express using first-person pronouns. The advantage we have over Lewis's two Gods is one way of capturing what is at issue in the debate around the so-called *essential indexical thesis*, the thesis that indexical attitudes have an irreducible role in intentional action explanations.

Imagine now a third God, omniscient and a competent first-person thinker but for the following limitation: she is unable to think of herself, thought of first-personally, as identical to anything encountered under a non-first-personal way of thinking. As with Lewis's Gods, this impairment (as well as her omniscience) sets her apart from us. Not only do thinkers like us know things about ourselves canonically expressible using words like 'I' and 'me', but, to our further advantage, we also know things about ourselves we canonically express *non*-first-personally — by the use of descriptions, for instance, or demonstrative expressions. We typically have no trouble slipping between these distinct forms of self-directed thought with minimal effort or attention, an ability anchored in our recognition of what it is for an identity between their referents to be true.

Recently there has been growing resistance to the essential indexical thesis, previously considered orthodoxy.[2] This paper doesn't join that resistance, but I do think that our preoccupation with the essential indexical thesis has led to neglect of the importance of our *non*-indexical ways of thinking of ourselves for our agency. That is the focus of this paper. Of course, thinking of oneself as identical with the referent of a non-indexical concept is precisely what God number three can't do — so, with this third God in view, the question of this paper becomes: what advantage do we have as agents over God

1. Lewis (1979), p. 520

2. See, e.g., Cappelen and Dever (2013), Magidor (2015)

number three? My answer is going to be that thinkers like her would be excluded from participation in some of the domains of agency we value most as part of a full human life — *viz*., domains of agency associated with our social identities.

It will be useful to have God number three's details up front. Let I be a first-person concept, and δ be an arbitrary concept other than a first-person concept. God number three is a thinker defined by two conditions:

> (i) She cannot entertain identities of the form $\ulcorner I = \delta \urcorner$; and
>
> (ii) she is otherwise omniscient and conceptually unlimited.

God number three shouldn't be conflated with another possible thinker — God number four — who cannot think of what is in fact themselves under a non-first-personal (or, equivalently as I am using the term, non-indexical) concept. God number four plays no part in the arguments to come.

Some points of further clarification about God number three. Since the question that ultimately interests us is one about the significance of certain conceptual capacities of ours — that is, the conceptual capacity to recognise oneself, thought of first-personally, as identical to something thought of non-first-personally — the limitation in (i) should be thought of as a stipulated conceptual deficiency on God number three's part, rather than (say) a glitch in her neural hardware. Since we don't want the downstream deficiencies we end up finding to be effects of any other limitations than this, make her omniscience in (ii) as fine-grained as you like. If knowing that *Hesperus is bright* and knowing that *Phosphorus is bright* is knowing two things, then she knows them both. By 'otherwise' in (ii), I mean other than limitations stipulated or implied by (i).

By *concept*, I mean a hyperintensional constituent of thought, individuated by its reference rule or the way it gets to have the reference that it has. The first-person concept is individuated by the reference rule that any token use of the concept refers to its user. This rule states the conditions under which someone counts as the referent of a token use of the concept; it does not provide descriptive content that a thinker must entertain in thinking a first-person thought. Any concept whose reference ascriptions are governed by this rule of token-reflexive reference is a first-person concept (I). Any concept whose reference ascriptions are not governed by this rule is not (δ).[3]

Both the limitation described in (i) and the powers ascribed to her in (ii) will make God number three's psychology unlike ours in various ways. One worry here might be that her minimal description makes it hard to say exactly what *happens* to God number three, psychologically speaking, when she attempts thoughts of the form $\ulcorner I = \delta \urcorner$ — and that this risks derailing her usefulness as a thought experiment. After all, so long as she has the concept I, the concept δ, the concept of identity, and the conceptual capacity for concatenation, then perhaps (the worry goes) this suffices to secure her capacity to entertain thoughts of the form $\ulcorner I = \delta \urcorner$. So what exactly goes wrong when she attempts to think them? Luckily, there are plenty of more familiar cases that will help to make God number three's psychology seem somewhat less incredible. For example: I have full mastery of the concepts *Caesar*, *prime number* and *is*, and the conceptual capacity for concatenation, but *Caesar is a prime number* is plausibly not a string that I can meaningfully entertain — at least, not in the same way that I can entertain intra-categorial strings. Of course, the explanation of my difficulties in entertaining this string will clearly be different from the stipulative explanation of God number three's difficulties in entertaining thoughts of the form $\ulcorner I = \delta \urcorner$. This isn't important: the important point is that this comparison gives us a way of understanding what happens to God number three when she attempts these thoughts. Psychologically speaking, let's say that God number three's attempts to entertain $\ulcorner I = \delta \urcorner$ thoughts are

3. This framework excludes those who either don't think the first-person concept refers (e.g. Anscombe 1975) or think it gets its reference in some way other than by the token-reflexive rule (e.g. Evans 1982); I take these to be minority views among contemporary philosophers of mind, but see Morgan (2015) for an exception.

broadly comparable to what it is like for us when we attempt to entertain cross-categorial strings like *Caesar is a prime number*.[4] She just can't make out what it would take for the thought to be true.

The aim of this paper, broadly speaking, is to raise a case for the importance of our non-indexical ways of thinking of ourselves for our agentive lives. I will pursue this aim by considering the predicament of God number three. The argumentative strategy comes in two steps. First I will establish that God number three's stipulated limitation in (i) precludes her from properly forming a certain range of *I am F* beliefs, despite her omniscience as otherwise given in (ii). The normative force of 'properly' here and throughout is that of epistemic justification. I argue for this first step in §§2–3. The second step is to show that this restriction to her properly formed *I am F* beliefs in turn restricts her agentive range—and more specifically, that it limits her participation in the domains of agency associated with her social identity. I argue for this second step in §4. §5 concludes.

§2. De jure absolute vulnerability to error through misidentification

The first step in the argument is to establish the claim that God number three's stipulated limitation in (i) entails that some properly formed *I am F* beliefs will be out of her reach, despite the fact that she is otherwise omniscient and conceptually unlimited. The key to establishing this claim is the theoretical notion of *de jure absolute vulnerability to error through misidentification relative to a use of the first-person concept*. The task of this section is to set out that notion; to do that, we will need to begin with the more familiar notion of *immunity to error through misidentification (IEM)*. After setting out these key notions in this section, the rest of the argument for the first argumentative step comes in the next section (§3).

4. Note that there are parallel questions about how psychologically realistic Lewis's two Gods are, if we think that acting intentionally entails *de se* knowledge of what one is doing and if we think that Lewis's Gods are intentional actors (see O'Brien [1994] for the beginnings of such an argument). This isn't normally taken to undermine their usefulness in highlighting the significance of our capacity for *de se* thought.

IEM is a modal property had by some judgment-grounds pairs. There is ongoing disagreement about how best to formulate it, but the following will do for our purposes:

> IEM. A judgment, *a is F*, made on grounds *g*, is IEM relative to *a* iff the following mistake is not possible: *g* provides proper grounds for knowledge that *something* is *F*, but the judgment is in error only because the subject has mistakenly identified *a* as the thing that she thereby knows to be *F*.[5]

There is yet more ongoing disagreement about which grounds get to count as issuing in first-person judgments with this property. Almost everyone agrees on introspection, but the rest is controversial. Other popular candidates in the literature include proprioception (*I have crossed legs*) and the distinctive way we have of knowing about our own actions in virtue of being their agent (*I am peeling a potato*). Gareth Evans adds to this list our faculty of perceptual self-location (*I am in front of a burning tree*) and, along with others after him, argued that episodic memory serves to preserve the IEM of first-person judgments into their past-tensed analogues (*I was on a ship*). Others have argued against the inclusion of one or more of these grounds.[6] We

5. This formulation is largely taken from Shoemaker (1969) pp. 556–7, with his talk of 'statements' updated to talk of 'judgments' to make it applicable to the mental realm, and with the standard addition (since Evans 1982) of relativisation to grounds. It approximates the *which-object* side of Pryor (1999)'s *which-object / de re* divide, but since this is the stronger category of the two, nothing hangs on putting things this way. By 'grounds' I mean to include both explicit inferences leading to the formation of the judgment, but also implicit presuppositional background (see Wright 2012 and Coliva 2006 for two theorists who think presuppositional structure should be included in assessments of IEM). Even where I drop the relativisation for ease of expression, I always mean IEM (and VEM) *relative to a use of the first-person concept*. As I have formulated it, IEM captures the impossibility of a certain kind of *false positive* error pertaining to a first-person thought that has already in fact been formed on certain grounds, rather than the impossibility of having grounds that lead improperly to the formation (or non-formation) of corresponding singular judgments; thanks to a reviewer for pressing me on this point, and see Salje (2016) for more on this distinction.

6. For examples of discussions of the claimed IEM of proprioception, see, e.g., Evans (1982), Cassam (1997), Hamilton (2009), Peacocke (2008), Coliva (2012), Morgan (2012),

needn't enter into these ground-level disagreements here — important for our purposes is merely that these epistemic grounds form a highly restricted class; only a few select grounds will get in. Any first-person judgment formed in any way other than by them will be *vulnerable* to error through misidentification (VEM) relative to a use of the first-person concept.

What is the significance of claiming that a given epistemic source or grounds (I use these terms interchangeably) is a member of this restricted group? The immunity isn't particularly valuable in itself — we're hardly especially anxious about making mistakes of just this kind. The significance, rather, is that the immunity tells us something *about the way in which first-person judgments are formed* on those grounds. If there is no possibility of error through misidentification relative to a use of the first-person concept, that is because the judgment's formation, including its presuppositional background, did not involve an identity between oneself thought of first-personally and oneself thought of non-first-personally. After all, the presence of such an identification would have brought with it the possibility of a *mis*identification. But if that's right, then it must mean that these epistemic sources are such that they *directly sustain first-person thought* — the deliverances of introspection (and the rest) are the special forms of self-knowledge through which I directly encounter myself *as* myself, without drawing on any additional identificatory information.

The term 'immunity to error through misidentification' is from Sydney Shoemaker's 'Self-Reference and Self-Awareness' (1969). There Shoemaker offered a category of IEM judgments now generally thought to be defunct: the category of *absolute* IEM judgments.[7] These are judgments, grouped by their predicative content, that aren't IEM relative only to a given set of circumstances in which they are formed, but *in all circumstances*. Shoemaker thought that mental-state self-ascriptions were like this. The judgment *I am bored*, for instance, is IEM in all circumstances of formation, or so the idea goes. The main reason this category has since been largely rejected is that it always seems easy enough to cook up circumstances in which a possibility of a misidentification error is introduced into the judgment's grounds. Perhaps I visit my therapist, who assures me that I am bored, and — trusting her judgment — I come to form the corresponding belief about myself. As it turns out, she's reading from the wrong client file. While this epistemic transaction is a good way of knowing that *someone* is bored, the judgment is in error solely because I have taken myself to be the proper witness to the warranted existential claim. So although the judgment *I am bored* is *typically* IEM relative to its use of the first-person concept, here is a set of circumstances in which it would be VEM relative to its use of the first-person concept. So much for the claim that all mental-state judgments are absolutely IEM.

Individual cases like this, together with natural extrapolatory assumptions, do enough to undermine Shoemaker's category of absolute IEM. However, I also think that there is a more principled reason to reject that category in the offing. The case just given typifies those normally designed to demonstrate the point in its involvement of a faulty testifier, and this is no accident. That's because testimony is a highly content-neutral epistemic source. This means that whatever the property, one can always be *told* that one is instantiating it — and whenever one comes to self-ascribe a property by being told about it, there is always a risk of an error through misidentification having been introduced somewhere in the epistemic grounds. Testimony, that is to say, is firmly on the VEM side of the divide, and this alone suffices to establish that Shoemaker's category of absolute IEM is an empty one.

I don't now intend to defend Shoemaker's category of absolute IEM. I do, however, think he was right about something — there is space for an "absolute" category of possibility of error through misidentification in the vicinity; it's just that Shoemaker was looking on the wrong side of the divide. There aren't judgments that are IEM in all circumstances,

Recanati (2012), Salje (2017); of episodic memory, see Shoemaker (1969), Evans (1982), Coliva (2006), Pryor (1999), Bermúdez (2013), Fernández (2014), Shoemaker (1970); of agent's awareness, see O'Brien (2007); of self-location, see Evans (1982), Cassam (1997), Pryor (1999), Hamilton (2009), Peacocke (2014).

7. p. 564

but there are judgments (categorised by predicative content) that are always VEM relative to a use of the first-person concept in all circumstances. Call this the *absolute* VEM *thesis*.

Why think the absolute VEM thesis is true? As it happens, we've already met the answer. That is, we have already seen that being an IEM source of first-person judgments is a tightly restricted class — however the ground-level disagreements are resolved in the final tally, it's clear that there aren't very many candidate epistemic grounds that will make it in; IEM is a high bar. Now, from this it follows that there is also a restriction on the kinds of judgment that can be properly formed on the basis of epistemic grounds falling in this restricted group. That's because there will be some judgments (categorised by predicative content) that aren't of the right kind to be properly formable in this narrow group of ways. Those judgments will be VEM relative to the first-person concept in all circumstances, or *absolutely* VEM.

Examples will help to make the point. Take the judgments *I have brown eyes* and *I am prime minister*. Also, suppose for the sake of argument that *all* the candidate IEM epistemic grounds mentioned above turn out to be sources of IEM judgments in the final tally. Even so, I will never be able to properly form either of these judgments on the basis of introspection, proprioception, action-awareness, perceptual self-location or episodic memory alone. These just aren't ways of knowing about eye colour or positions in elected office. No amount of action-awareness or knowledge of limb position will put me in a position to know what colour my eyes are or whether my party won the election. To properly form judgments about my eye colour or elected roles, rather, I must plausibly turn to sources on the VEM side of the divide — mirror-involving vision or testimony for the first, and presumably some process involving testimony for the second. If that's right, then there are some judgments that fall into the category of absolute VEM.

There is a further distinction among these judgments that will be useful for our purposes — and that will also help to show *why* these two example judgments are absolutely VEM. The distinction is between two kinds of absolutely VEM judgments: those that fall in the category as a matter of contingent fact (*de facto* absolutely VEM judgments), and those that fall in it as a matter of conceptual necessity (*de jure* absolutely VEM judgments).

I have brown eyes is an example of a *de facto* absolutely VEM judgment. There are no ways of properly forming this judgment through an IEM epistemic source as things are for thinkers like us in a world like this. That's because eye colour is a visible property, so, given how things are for us, we can find out about it only either by *seeing* it or by being told about it by someone else who has seen it, or perhaps by inferring it from a reliably correlated non-visible property — all epistemic grounds that introduce the risk of an error through misidentification into the final judgment. But this is just an accident of our physiology. We might have been set up with eye colour on the inside of our retinas, in such a way that we could visually access our own eye colour and no one else's "from the inside". Or we could have been biologically furnished with a sort of fleshy headset involving blinkers and reflective surfaces, such that our entire visual range from birth to death was limited to gazing into our own eyes. Or perhaps we could have been equipped with an additional sense-modality, a form of electro-sensory perception that delivers a constant stream of detailed information about micro-changes in one's own eye colour in response to light variations in the immediate environment. In any of these cases, we would have had what Michael Martin and others have called a *sole-* or *single-object* faculty for knowing about eye colour; a way of knowing about one's own eye colour that is dedicated to one's own eyes and no one else's.[8] Where there is single-object dedication like this, there is IEM relative to a use of the first-person concept — these would be ways of knowing about *one's own* eye colour, if it were a way of knowing about anyone's eye colour at all. There could be no possibility of a misidentification error on the basis of a single-object dedicated source of knowledge, because there is only one object in question. So

8. Martin (1995, 1997); see also Campbell (1999)

although *I have brown eyes* is absolutely VEM for creatures like us in a world like this, things might have been otherwise.

De jure absolutely VEM judgments are not like this. These are judgments that don't have their VEM property contingently, but as a matter of conceptual necessity. The second judgment above, *I am prime minister*, is an example of this kind. I will argue for this claim in the next section, but the present aim is to introduce the category of *de jure* VEM judgments by example — and for that, it is enough to point out the starting plausibility of the idea that I couldn't properly judge this to be true of myself in solipsistic isolation, by focussing hard enough through appropriately trained introspection, or self-location, or proprioception. To properly judge that *I am prime minister*, rather, it's plausible that I must see myself "from the outside" as others do, as occupying a certain role embedded in a social context. Now, unlike the eye-colour example from the last paragraph, this doesn't seem like an accident of our physiology. It's not as if we could imagine some other way we might have been built that would make it easier to see how introspection could be a suitable way of grounding *I am prime minister* judgments, or that could provide us with a different single-object dedicated epistemic faculty that would have done the job. Whatever adjustments we might imagine, falling under the concept *prime minister* will always involve occupying a publicly recognised social role, and this makes it natural to think that it will always turn out that proper formation of judgments about being prime minister will involve epistemically apprehending oneself as others do. If this is right — and I will argue in the next section that it is — then it will be a matter of conceptual necessity that the judgment is VEM. It will be *de jure* absolutely VEM.

We now have before us the key notion of *de jure absolutely* VEM *judgments* relative to a use of the first-person concept.[9] These are first-person judgments whose proper self-ascription could never be exclusively made on the basis of a single-object dedicated epistemic source

9. As with IEM, I will mostly drop (but always mean) the relativisation to a use of the first-person concept.

as a matter of conceptual necessity. It is this category of judgments that will be of primary interest in the rest of this paper. That's because it is these properly formed judgments that will be systematically unavailable to God number three.

§3. What God number three can't properly believe

A reminder of the two definitive characteristics of God number three, repeated from above:

> (i) She cannot entertain identities of the form $\ulcorner I = \delta \urcorner$; and
>
> (ii) she is otherwise omniscient and conceptually unlimited.

Notice now that the identity mentioned in (i) is just the sort of identity involved in the proper formation of self-ascriptive VEM judgments. (Recall, the significance of saying that a self-ascriptive judgment is IEM relative to a use of *I* is that there is no such identity involved in the judgment's formation; conversely, the significance of saying that a self-ascriptive judgment is VEM relative to a use of *I* is that there *is* such an identity involved. It is the presence of this identification that explains the possibility of a *mis*identification.) Where there are judgments, grouped by content, that can be formed in both IEM and VEM ways (e.g. *my legs are crossed*), the deficiency in (i) needn't pose any problem for God number three — she will still be able to properly form these judgments; it's just that she will have to stick to the non-identification-involving routes to get there. Trouble for God number three comes when there are no IEM options. That's to say, God number three will not be able to properly form any absolutely VEM judgments.

This trouble does not run very deep for self-ascriptive judgments that are merely *de facto* absolutely VEM, like *I have brown eyes*. This might not be a properly formable judgment for God number three so long as she is otherwise just like us, but we don't need to change very much about her to make it properly available to her. We need only add to the case that God number three differs from us physiologically in

one of the ways described in the last section, such that she has a single-object dedicated way of forming the eye-colour judgment even if we don't. (In her world, say, eye colour is on the other side of the retina and visually accessible "from the inside".) So as yet, there's no reason to suppose that these sorts of judgments, as categorised by their predicative contents, will be out of reach to a thinker like God number three, even if there are some ways of properly forming them that are.

Real trouble for God number three comes in the form of *de jure* absolutely VEM judgments, judgments like *I am prime minister*. Judgments like this aren't just typically or contingently VEM relative to a use of the first-person concept; these judgments are always VEM for all thinkers in all worlds. As a result, no matter how we invite God number three to reach the judgment, or what physiological tweaks we imagine on her behalf, there will be no way of properly forming self-ascriptive judgments in this category that don't involve an identity of the form ⌜ $I = \delta$ ⌝ — an identity of just the kind that she is *ex hypothesi* barred from entertaining. So properly formed *de jure* absolutely VEM judgments are out of range for God number three. These are judgments that can be properly reached only in an identity-involving way as a matter of conceptual necessity, and God number three is precluded from entertaining identities of the relevant kind as a matter of stipulation.

The broad aim of this paper is to raise a case for the importance of our non-indexical ways of thinking of ourselves. That there is a category of properly formed self-ascriptive judgments that is available to us only by virtue of our capacity to think of ourselves, thought of as ourselves, as identical to the referents of non-first-personal concepts already largely answers that aim. It turns out that this capacity is what funds our ability to properly form this range of *I am F* beliefs: I learn *I am that way* by learning of the person who is in fact myself, thought of in a way that someone else might equally use to think of me, that *they* are that way. The next step, however, is to say something substantive about which beliefs fall in this range. I suggest that an important group of judgments falling into it are self-ascriptive social identity judgments, judgments that involve the self-ascription of a social kind predicate. These non-exhaustively include self-ascriptions of gender (*I am a woman*), religion (*I am Buddhist*), religious-cultural groupings (*I am Jewish*), disability status (*I am able-bodied*), professional role (*I am a plumber*), societal role (*I am a prisoner*), caste (*I am a Brahmin*), nationality (*I am French*), race (*I am Hispanic*) and subcultural groupings (*I am a hipster*).

The epistemology of social kind judgments is clearly varied and rarely obvious.[10] Those partly associated with perceptible biological characteristics (e.g. race) will plausibly involve perceptual grounds. Some (e.g. gender) will surely involve introspection. Others (e.g. nationality, professional role) don't seem to involve perception or introspection at all, but rather seem largely dependent on direct testimony. In common to them all, however, is that they all involve the ascription of social kind predicates — and this means that their proper formation will involve both knowing about a certain range of natural (or, more neutrally, non-social) properties, and a proper appreciation of those properties as relevant to ascriptions of the appropriate social kind property. Let's call this view — on which the proper ascription of social kind predicates involves both the proper recognition of the instantiation of non-social properties, and proper recognition of their social relevance — the *two-component model* of the epistemology of social kind judgments.

This two-component model forms a natural epistemic counterpart to views of social ontology falling under what we might call the

10. Despite growing recent discussion of social ontology (cf. Epstein [2018] for a recent survey of the social ontology literature) and an ongoing separate debate about the semantics of social kind terms (descended from Burge 1979), it's arguable that the epistemology of social kinds has been somewhat left behind. Amie Thomasson stands as an exception to the under-discussion of the epistemology of social kinds, in her interest in the question whether there could be gross error or ignorance throughout a community about the constitutive rules on social kinds (see Thomasson 2003). Another notable exception is Haslanger, who addresses the question whether we could be ignorant of the rules making up our *operative* (as opposed to our more formally or explicitly defined *manifest*) social kind concepts (Haslanger 2005, Haslanger and Saul 2006). However, both of these questions differ from questions of the form 'How do I know that this token entity is a five-pound note?', which, as far as I can see, hasn't yet been directly addressed in these literatures.

two-dimensional model of social kinds.[11] Included under this model are any views on which facts about the metaphysics of social kinds are determined by a specification of certain non-social facts on the one hand, and a specification of how those facts ground the associated facts about social reality on the other. The most influential two-dimensional model of social ontology comes from John Searle (1995, 2006, 2010), on whose view non-social facts ground social facts through the *collective imposition of status functions* of the form *x counts as y* relative to given contexts. Under the corresponding two-component model of the epistemology of social kind judgments, it's not just the metaphysics of the domain that is fundamentally social. To properly judge a social kind predicate to apply, not only must I be in a position to know that the object has a certain range of non-social properties, but I must also be in a position to properly recognise those properties as relevant to our communally determined social kind categories. If this is right, then there is something deeply social about the proper basis for applying social kind predicates.

What exactly does this mean for the way we come to properly apply social kind predicates — how do I know that this is a five-pound note, this a Spaniard, this a woman?[12] According to the two-component model, I must know (a) that this object has a certain range of non-social properties, and (b) that objects with these (non-social) properties count as five-pound notes in a UK context in 2018. The relevant question for our purposes now is whether it would be possible to gain knowledge of this kind through a single-object dedicated epistemic source.

So, is it possible? Well, yes and no. What's conceivable is that I could have a single-object dedicated source of knowledge of many of this entity's non-social properties: its colouring, texture, size, surface-reflectance properties and so on. (We might imagine a scenario in which I have all my exteroceptive sensory faculties locked exclusively onto this item from birth.[13]) That is the "yes" part of the answer, and it amounts to the claim that the knowledge mentioned in (a) could be gained in a single-object dedicated way.

The "no" part corresponds to the knowledge mentioned in (b). It amounts to the claim that no conceivable single-object dedicated channel will do as a way of knowing that the object so perceived falls under the social kind *five-pound note* in the relevant context. This is a strong claim, but one that we have good reason to accept. That's because what it is to fall under a social kind is to have a *publicly recognisable* status as having what it takes to count as an instance of the kind. This means that the properties featuring in our communally agreed rules on social kinds must be in-principle publicly accessible properties. And it is these publicly accessible properties that I must know a given worldly entity to instantiate before I can know it to be an instance of that kind. If this is right, then while there could conceivably be a single-object dedicated way of knowing about some of this entity's non-social properties (a), there is no possibility of a single-object dedicated way of knowing that entities with those properties count as falling under a certain social kind category (b). Both components are needed for a properly formed social kind judgment, so we should conclude that there is no single-object dedicated way of properly arriving at the judgment *this is a five-pound note*.

There are two premises in this argument in need of support: first, the claim that the properties featuring in the constitutive rules for social kinds must be in-principle publicly accessible; and second, the claim that it wouldn't suffice for knowledge that something is an instance of a social kind that those publicly accessible properties are accessed in a single-object dedicated way. I'll take each in turn before

11. This use of the term 'two-dimensional model' is from Brouwer (2018); at the intended level of generality here, the model is intertranslatable between fact-talk and entity-talk.

12. I frame the following discussion in terms friendly to Searle's framework, as the currently dominant account in the literature, but analogues of the argument to follow will apply to any two-dimensional account of social ontology that appeals to communally established constitutive rules on social kinds, even where the agreement is less explicit than on Searle's account (see Epstein 2013, p. 14, for a helpful survey of options).

13. It's harder to know how I could have a single-object dedicated way of knowing facts about its provenance — perhaps an exercise of single-object dedicated abductive inference.

turning back to social categories more pertinent to the predicament of God number three than the example of a five-pound note.

First, the claim that properties featuring in the constitutive rules for social kinds must be in-principle publicly accessible. Earlier (p. 12), I implied that this claim follows from what it is to be a social kind. There are, of course, a good many ways of spelling out the notion of a social kind, and we don't yet seem to have reached a clear consensus about how best to do it.[14] Minimally, however, social kinds are theoretical posits in our best theories in the relevant domains, posits that earn their keep (as posits do) by their explanatory and predictive power. But, for a social kind to explain and predict observable social phenomena, it couldn't be the case that its instances are in principle unrecognisable as such to the individuals involved in the explanatory-target phenomena.[15] For instance, it would clearly be a bad idea to posit a human-society currency whose constitutive condition included the property of being composed of 1,638,278 billion atoms. Individuals in that society wouldn't be able to recognise whether they had any money in that currency or not, so it couldn't be used to explain or predict their patterns of behaviour at an appropriate level of explanation. What this example demonstrates is that the properties that feature in the constitutive rules for social kinds are plausibly constrained by the condition of (in principle) public accessibility.

Next, why couldn't these properties *also* be accessible through a single-object dedicated faculty, and why couldn't such single-object access be enough for a knowledgeable ascription of the associated social kind predicate? For example, why isn't it enough for knowledge that this is a five-pound note to use information coming in through my locked-in senses that it has all the right properties to count as such, so long as the properties I have single-object access to *could* also in principle be accessed in some public multiple-object way?

The problem here is a classic opacity issue. The difference between these two ways of apprehending the properties impedes direct recognition that the publicly accessible properties ($a_1 \ldots a_n$) featuring in the relevant constitutive rule are the same as the properties ($b_1 \ldots b_n$) apprehended through the single-object dedicated epistemic channel. If I can access this range of properties only in a single-object dedicated way, but they feature in the constitutive rule under their publicly accessible guise, then I must find a way of telling that $a_1=b_1$, $a_2=b_2$ and so on, before I can rightly be said to know that this object has what it takes to count as a five-pound note. No problem for those of us who have both ways of accessing (and so comparing) the same properties, but this would be much more of a problem for an epistemic agent like God number three, who has access *only* through the single-object dedicated route. So, to know that this is a five-pound note, not only must I have a way of telling that the item instantiates the relevant range of properties, but I must know it in a way that facilitates recognition that these are the very same publicly accessible properties that feature in the constitutive rule that says that objects with properties $a_1 \ldots a_n$ count as five-pound notes in a UK context in 2018. In other words, I must epistemically apprehend the object's properties in the very same way that others do. If this is right, then there could be no single-object dedicated way of knowing that *this is a five-pound note*.

Turn now to social kind categories that apply to people, like *woman*, *black* or *Spaniard*. As it goes for five-pound notes, so it goes for genders, racial categories and nationalities too. Again, it's perfectly conceivable that we could have had single-object dedicated ways of properly self-ascribing certain (non-social) biological properties typically associated with these social kinds. There is no difficulty in conceiving of a physiologically tweaked God number three who has an IEM way of coming to properly self-ascribe a skin colour, a sex organ or a DNA profile. But, of course, to properly self-ascribe a skin colour, sex organ or DNA profile is not yet to properly self-ascribe social kind predicates

14. Cf. Epstein (2018), s.2

15. At least, in the majority of cases. There could conceivably be derivative cases of instances of the kind whose relevant properties were not accessible to others, but only once the kind were established by paradigm instances in which the properties *were* publicly accessible; hence the 'in principle' qualifier.

of race, gender or nationality. In different settings, the same skin colour is racialized in different ways, the truth of gender claims varies independently of the truth of claims about sex organs, and there is a vanishingly low correlation between DNA profiles and nationality. Rather, to know that this is a woman, this person is black or this is a Spaniard is to know that the individual has the cluster of in-principle publicly accessible properties that feature in a communally imposed status function for those social categories in the local context (e.g. people with properties $a_1 \ldots a_n$ count as black in a UK context in 2018).[16] Now, the moral of the last few paragraphs is that for token social kind ascriptions to count as well-grounded, I must be in a position to recognise the individual's properties as being the very same ones as those featuring in this constitutive rule on the social kind, without running into opacity issues. This means that to properly ascribe these categories, I must apprehend the individual's relevant properties, $a_1 \ldots a_n$, in the same way that others do. That is the case even when the target of the ascription is myself.

The upshot of this argument is that it is built in to social kind concepts that they cannot be properly ascribed in a single-object dedicated way. It is part of the nature of these concepts that their proper application depends on an awareness of the object's properties that feature in the relevant communal beliefs and attitudes. Insofar as I know about certain properties of mine only in a single-object dedicated way, I cannot be in a position to recognise them as the properties that others are responding to in their ascriptions of the social kind. It is only insofar as I know about my own properties in the same way that others do that I am justified in treating these as the very properties that feature in the relevant collective status function imposition. Social kind self-ascriptions are *de jure* absolutely VEM relative to uses of the first-person concept.

This brings us back to the central question of this section: what can't God number three properly believe — God number three, who cannot entertain identity beliefs of the very kind needed to form VEM judgments? I have argued that included in the category of *de jure* absolutely VEM judgments are self-ascriptive social identity judgments. So, included in the first-person judgments that systematically elude God number three are judgments about her social identity. No matter what counterfactual interventions we try, God number three could never properly come to form beliefs about her own gender, race, social role, profession and the rest; this range of properly formed *I am F* beliefs is beyond her otherwise omniscience. This is no small lack on her part — these are self-ascriptive beliefs that matter to us. They pick out aspects of our identity that make us "who we are", in the important non-philosophical sense of that phrase. God number three, then, already looks to be deeply debilitated, and in answer to the broad question of this paper, the importance of our conceptual capacity to think of ourselves as ourselves under non-indexical concepts must already be correspondingly promoted in its significance.

That concludes the first argumentative step of the paper. In the next section, I argue that there are further ramifications for what God number three can't *do*.

16. This proposal is neutral on what these properties are; I don't rule out that they could include such self-determined properties as public avowals.

§4. What God number three can't do

The stated aim of this paper was to identify the advantage we have over God number three *as agents*. The debility mentioned at the end of the last section doesn't yet give us an answer to that question, because it tells us only what self-ascriptive beliefs she can't properly form, not what she can't *do* as an agent. But it gets us most of the way there. That's because the answer is going to be that the domains of agency that God number three is systematically excluded from are (at least) those whose explanation involves essential mention of properly formed self-ascriptive social identity beliefs on her part.

How exactly should this proposal be fleshed out? I'll start with a simple suggestion, which I only mention to put to one side: the actions that God number three is precluded from intentionally performing include all those that would have been motivated by properly formed

self-ascriptive social identity beliefs. For example, given that God number three can't properly self-ascribe a gender predicate, it will be beyond her to intentionally tick the box marked "W" on a passport application. The "domains of agency" that God number three is excluded from on this simple suggestion are (at least) made up of all the individual actions like this that her doxastic debility from the last section prevents her from intentionally performing.

The problem with the simple suggestion is that an improperly formed belief can explain intentional action just as well as a properly formed one. If I take hallucinogenic drugs that cause me to form the belief that an angry pink elephant is stalking me with a hunting knife, this improperly formed belief will obviously feature in a personal-level explanation of why I'm currently hiding under the table. Likewise, even if God number three's stipulated limitation prevents her from properly forming self-ascriptive beliefs about her social identity, there's nothing to stop her forming them improperly (by gut feeling, for instance, or by use of psychoactive drugs) and then intentionally acting by their lights — forming the spontaneous belief that she is a woman and intentionally ticking the box, for instance. Their improper formation would make these beliefs ill-grounded, but ill-groundedness doesn't obviously undermine explanatory force with respect to intentional actions performed on their basis.

A better suggestion is one that shifts from thinking about God number three's capacity for individual intentional action performances to her capacity to produce reliable and coherent patterns of agency over time. That's because even if one-off intentional action performances can be adequately explained by appeal to God number three's improperly formed beliefs about her social identity, it's not true that reliable and coherent patterns of action can be likewise explained. My ill-formed pink-elephant belief might enter centrally into a good explanation of why I am intentionally acting as I am at the moment, but its ill-groundedness makes it unstable; as soon as I sober up, I'll come out from under the table. Such drug-induced beliefs are too fickle to motivate reliable and coherent patterns of action over time. Likewise with

God number three's ill-formed social identity beliefs. Even if she can improperly form one-off beliefs about her social identity, they won't be stable enough to sustain reliable and coherent patterns of action over time.

Why are properly formed (or, equivalently, well-grounded or epistemically justified) beliefs better suited to sustaining reliable and coherent patterns of action over time than their improperly formed counterparts? This is really a new guise for the old question of what justification adds to mere true belief. And the answer I am drawing on here is likewise an old one:

> So long as they stay with us, true beliefs are a fine possession, and effect all that is good; but they do not care to stay for long, and run away out of the human soul, and thus are of no great value until one ties them down by working out the cause. (Meno 97e–98a)

Improperly formed beliefs are unstable beliefs, whereas proper formation makes for stability over time. And it is only stable beliefs over time that can reliably motivate coherent patterns of action over time.

I say that the explanation of reliable and coherent patterns of intentional action over time requires that the motivating beliefs are well-grounded. I don't say it requires their *truth*. A well-grounded pattern of false beliefs can enter into the explanation of a reliable and coherent pattern of intentional action just as effectively as a well-grounded pattern of true ones. An individual who is misinformed early on about their birth date will reliably and coherently intentionally act in ways made appropriate by their properly formed beliefs about how old they are, despite the falsity of those beliefs. An individual who internalises a misascribed gender category as a child will be reliably motivated to act coherently in ways made appropriate by those gender beliefs, despite those beliefs being false. The key notion for the production of reliable and coherent patterns of intentional action over time is not truth but well-groundedness.

We are now in a position to give the paper's proposal in full. In §§2–3 I argued that God number three's stipulated limitation in (i) precludes her from properly forming self-ascriptive social identity beliefs. That was the first argumentative move. I now propose that this doxastic limitation in turn precludes her from producing reliable and coherent patterns of intentional action associated with her social identity. That is the second move. Among others, included in this range will be patterns of activity characteristically involved in raising a family, in having a career or otherwise performing a professional role, in carrying out civic or political roles, in engaging in religious life, in developing and maintaining meaningful interpersonal relationships, and in deliberately developing modes of social-identity expression. God number three's agency would be severely compromised, and compromised in ways that matter deeply. Whatever one's conception of the good, counted among these domains of agency are clearly those we value most as part of a full human life.

§5. Conclusion

It's hard to deny our advantage over thinkers like Lewis's Gods. Unlike them, we know a great many things about ourselves that we naturally express using first-person pronouns. Plausibly (and importantly) this capacity facilitates intentional action. I must realise that the bear is after *me* before I can flee. I couldn't be properly motivated to right the sugar sack unless I notice that *I* am the one making a mess. The aim of this paper has been not to deny the essential indexical thesis, but to argue for a complementary thesis:

> Our capacity to think of ourselves, as ourselves, as identical to the referents of non-first-personal concepts is essential to the explanation of our capacity for stable and coherent patterns of intentional action associated with our social identities. (*Essential non-indexical thesis*)

Indexical and non-indexical ways of thinking of ourselves are both important for our agency. They are just important for different reasons.

Even if the essential non-indexical thesis doesn't contradict the essential indexical thesis, it does stand counter to another influential philosophical tradition. There have been many articulations, by philosophers of different traditions, of the idea that apprehending oneself *as oneself* is in some sense incompatible with apprehending oneself *as an object* — indeed, for Schopenhauer, this is "the most monstrous contradiction ever thought of".[17] If the essential non-indexical thesis is true, then not only is it possible to think of oneself as oneself as an object, but our capacity to think of ourselves this way is indispensable to our capacity to participate fully in some of the most value-laden domains of intentional activity.[18]

Bibliography

Anscombe, G. E. M., 1975: 'The first person', in Samuel D. Guttenplan (ed.), *Mind and Language*, Oxford University Press

Bermúdez, José Luis, 2013: 'Immunity to error through misidentification and past-tense memory judgements', in *Analysis* 73:2

Brouwer, Thomas, 2018: 'Social inconsistency' (ms)

Campbell, John, 1999: 'Immunity to error through misidentification and the meaning of a referring term', in *Philosophical Topics* 26:1–2

Cappelen, Herman and Josh Dever, 2013: *The Inessential Indexical: On the Philosophical Insignificance of Perspective and the First Person*, Oxford University Press

Cassam, Quassim, 1997: *Self and World*, Oxford University Press

Coliva, Annalisa, 2006: 'Error through misidentification: Some varieties', in *The Journal of Philosophy* 103:8

Coliva, Annalisa, 2012: 'Which 'key' to all mythologies about the self? A note on where illusions of transcendence come from and how to resist them', in Simon Prosser and François Recanati (eds.),

17. Cited in Cassam (1997), p. 2; see that chapter for a discussion of other expressions of the same idea.

18. Thanks to audiences at Stirling, La Trobe, Adelaide, Uppsala, KCL and Oxford for feedback on earlier versions of this work; to Daniel Morgan and Thomas Brouwer for comments and discussion; and to two anonymous reviewers for this journal.

Immunity to Error through Misidentification: New Essays, Cambridge University Press

Epstein, Brian, 2013: 'Social objects without intentions', in Anita Konzelmann Ziv and Hans Benhard Schmid (eds.), *Institutions, Emotions, and Group Agents: Contributions to Social Ontology*, Springer

Epstein, Brian, 2018: 'Social ontology', in *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.)

Evans, Gareth, 1982: *The Varieties of Reference*, Oxford University Press

Fernández, Jordi, 2014: 'Memory and immunity to error through misidentification', in *Review of Philosophy and Psychology* 5:3

Hamilton, Andy, 2009: 'Memory and self-consciousness: immunity to error through misidentification', in *Synthese* 171:409

Haslanger, Sally, 2005: 'What are we talking about? The semantics and politics of social kinds', in *Hypatia* 20:4

Haslanger, Sally and Jennifer Saul, 2006: 'Philosophical analysis and social kinds', in *Proceedings of the Aristotelian Society, Supplementary Volumes* vol. 80

Lewis, David, 1979: 'Attitudes *de dicto* and *de se*', in *The Philosophical Review* 88:4

Magidor, Ofra, 2015: 'The myth of the *de se*', in *Philosophical Perspectives* 29:1

Martin, M. G. F., 1995: 'Self-awareness: A sense of ownership' in José Luis Bermúdez, Anthony Marcel and Naomi Eilan (eds.), *The Body and the Self*, MIT Press

Martin, M. G. F., 1997: 'Self-observation', in *European Journal of Philosophy* 5:2

Morgan, Daniel, 2012: 'Immunity to error through misidentification: What does it tell us about the *de se*?', in Simon Prosser and François Recanati (eds.), *Immunity to Error through Misidentification: New Essays*, Cambridge University Press

Morgan, Daniel, 2015: 'The Demonstrative Model of first-person thought', in *Philosophical Studies* 172:7

O'Brien, Lucy, 1994: 'Anscombe and the self-reference rule', in *Analysis* 54:4

O'Brien, Lucy, 2007: *Self-Knowing Agents*, Oxford University Press

Peacocke, Chris, 2008: *Truly Understood*, Oxford University Press

Peacocke, Chris, 2014: *The Mirror of the World: Subjects, Consciousness and Self-Consciousness*, Oxford University Press

Prosser, Simon and François Recanati (eds.), 2012: *Immunity to Error through Misidentification: New Essays*, Cambridge University Press

Pryor, Jim, 1999: 'Immunity to error through misidentification', in *Philosophical Topics* 26:1–2

Recanati, François, 2012: 'Immunity to error through misidentification: What it is and where it comes from', in Simon Prosser and François Recanati (eds.), *Immunity to Error through Misidentification: New Essays*, Cambridge University Press

Salje, Léa, 2016: 'The Subjective Perspective in Introspection', in *Journal of Consciousness Studies* 23:3–4

Salje, Léa, 2017: 'Crossed wires about crossed wires: Somatosensation and immunity to error through misidentification', in *Dialectica* 71:1

Searle, John R., 1995: *The Construction of Social Reality*, Free Press

Searle, John R., 2006: 'Social ontology: Some basic principles', in *Anthropological Theory* 6:1

Searle, John 2010: *Making the Social World: The Structure of Human Civilization*, Oxford University Press

Shoemaker, Sydney, 1969: 'Self-reference and self-awareness', in *The Journal of Philosophy* 65:19

Shoemaker, Sydney, 1970: 'Persons and their pasts', in *American Philosophical Quarterly* 7:4

Thomasson, Amie L., 2003: 'Realism and Human Kinds', in *Philosophy and Phenomenological Research* 67:3

Wright, Crispin, 2012: 'Reflections on François Recanati's "Immunity to error through misidentification: What it is and where it comes from"', Simon Prosser and François Recanati (eds.), *Immunity to Error through Misidentification: New Essays*, Cambridge University Press