

RECENT WORK

First-Person Thought

DANIEL MORGAN AND LÉA SALJE

1. Introduction

Subjects have various ways of thinking about themselves. Here are three examples: a subject can think of herself under an appropriate description (*the hiker*), demonstratively (*that person*) or by mentally mouthing her name (*NN*). Whenever a subject thinks of herself in one of these ways, there is a further question whether she also realizes that it is she herself she is thereby thinking of. In contrast, there is another way of thinking of oneself deserving of a place on this list that seemingly differs in this respect – the way of thinking of oneself *first-personally*, or *de se* (we'll use these terms interchangeably). As a heuristic, we might say that this is the kind of thought about herself that a thinker would naturally express using a first-person pronoun in language ('I am F' or 'The F is *me*').

It's easy to feel that *de se* thought doesn't just deserve a place on our list; it deserves a top spot. After all, even if a subject *can* think of herself in the three initial ways mentioned above, surely her first-personal way of thinking of herself looms much larger in her mental life than the others, both with respect to how often she thinks of herself in that way and how important the capacity to do so is to her normal functioning. Historically, first-person thought has attracted a great deal of philosophical attention. As Peacocke writes, '[a]rticulating the phenomena distinctive of the first-person has drawn forth some of the most striking contributions from the greatest philosophers ...' (2008: 77). The aim of this article is to frame some of the more recent movements and insights in the area from the last decade or so.

First, what is *de se* thought? We've offered as a heuristic a relation to a bit of language: the first-person pronoun. But you might wonder whether this is really its most explanatorily important property. One might think that a deeper characterization will be available in terms of something more intrinsic – for example the story about how *de se* thoughts achieve their reference. Such a characterization would inevitably import controversy. The most popular contemporary view is that *de se* thoughts are governed by a rule that they refer to whoever is thinking them (see Campbell 2012b, Peacocke 2014a, O'Brien 2007). However, not everyone agrees. Some are tempted to deny that *de se* thought is referential at all (see Doyle 2018). Some think that different instances of *de se* thought might be associated with different

referential stories (see Longuenesse 2017, Johnston 2010). Finally, some think that *de se* thoughts might have similar referential stories to perceptual demonstrative thoughts (see Morgan 2015, Setiya 2015). But all these people can agree that there is some distinctive kind of thought associated with the first-person pronoun. Since we want to use the term *de se* in such a way as to make it as uncontentious as possible, nothing we say will depend on anything beyond the linguistic characterization (unless otherwise mentioned).

The domain of first-person thought is not characterized by a central organizing question. Instead, it makes up a somewhat disorderly collection of questions – sometimes interconnected, sometimes not – about the cognitive, metaphysical and normative significance of first-person thought, as well as questions about how best to model it, and how to reconcile our philosophical theories in this area with findings from the cognitive sciences. Just a few topics that we might have, but have not, discussed in this article include: the continued development of arguments in the metaphysics of personal identity that draw on premises involving claims about *de se* thought (see Hamilton 2013, Peacocke 2014a, Madden 2011, 2013); a mini-resurgence of interest in the *Cogito*, and related epistemic issues (see Peacocke 2012a, Campbell 2012a, O'Brien 2015); an increased interest in the 'sense of mine-ness' (and relatedly, in the role that some argue *de se* thought plays in a full account of conscious experience; see Guillot and García-Carpintero forthcoming); the relation of sophisticated forms of *de se* thought to more primitive, including non-conceptual precursors (see Musholt 2015, Bermúdez 2016, Peacocke 2014a); how much the *de se* has in common with the *de nunc* (see Morgan 2019a); and even the possibility of leveraging *de se* phenomena in support of absolute idealism (see Rödl 2018).

Instead we have chosen two areas to discuss in detail: scepticism about the distinctive philosophical importance of *de se* thought (§2) and immunity to error through misidentification (IEM) as it shows up in *de se* thought (§3). The reason for this selection is that these strike us as two of the most active areas involving the *de se*: in both cases, there has been notable movement in the state of the debate in the last 10 years, partly as a result of contributions from different authors rebounding off each other in a sustained way.

2. Scepticism about the importance of the *de se*

One apparently pressing philosophical task concerning the *de se* is to get a clear picture of the distinctive features it has that are relevant to various aspects of mind – e.g. intentional action, emotion, perception, reasoning, understanding. The liveliest, most joined up, recent debate has focused on intentional action, so we'll focus on that, while leaving open the extent to which the dialectic around action might mirror other areas. The discussion here is very example-driven, and the diet of examples has been around since the 1970s. What is newer in the last 10 years is an acknowledgement that it's

not clear what those familiar examples establish (see, e.g., Prosser 2015: 211, Ninan 2016: 86, Torre 2018: 172). This is stimulated by a resourcefully defended minority position according to which they do *not* establish anything of interest about the *de se*.¹ We'll call this position *de se scepticism* and opponents of the position *de se fans*. Relatively few have found *de se* scepticism wholly convincing, but the versions of *de se* fandom that have emerged in response to it are instructively different.

Here is one of the influential cases: Perry is following a trail of sugar along a supermarket floor, looking for the messy shopper who is creating the trail. At some point, Perry realizes, 'I am the messy shopper'. Having realized this, Perry stops following the trail and sorts out his own trolley (Perry 1979: 3).

De se sceptics have acknowledged that cases like this one show that going from thinking of oneself in a non-*de se* way to a *de se* way can make a difference to how one acts (as Perry noted). But they point out that this just seems to a special case of the more general point that how one acts depends not just on what things one is thinking about, but how one is thinking about them. We can create the same kind of effect using a non-*de se* example in which an agent moves from using one proper name to another co-referential one. For example, Lois hails a man she had previously ignored because she comes to realize that the man, whom she had already recognized as Clark Kent, is Superman, the very person she'd been wanting to talk to. In slogan form: *de se* cases are just opacity cases.

On these grounds, Cappelen and Dever (2013) and Magidor (2015) both argue that in order for there to be a distinctive connection between the *de se* and action, the *de se* would have to be involved in *every* explanation of action, thus raising the bar on what the fan of the *de se* seems to need to show. They then note that simply inspecting a few cases, such as Perry's case, provides no support for that universal generalization. Constructively, they offer a recipe for generating examples of action in whose explanation no *de se* attitude figures. The key ingredient is that the agent lacks any self-directed attitude at all – so the question of whether that attitude must be *de se* does not arise. The agent forms an intention that a certain state of affairs obtain – e.g. that a certain door is closed, that a charity is donated to – and that triggers their action. Granted, to bring such a result about, the agent will have to do something else – e.g. move their body in a certain way. But, the suggestion is, this need not involve the formation of a self-referential attitude (e.g. a *de se* intention the agent would express with 'I walk over to the door and turn my hand'). The agent may simply be wired up so that when they form intentions whose contents are 'actionable' states of affair, their sub-personal systems cause their bodies to move in such a way as to cause

1 See Cappelen and Dever 2013 and Magidor 2015. Other defenders of views in the same broad area – in that they also think exaggerated claims have been made about the *de se* – include Devitt (2013), Millikan (1990), Spencer (2007) and Stanley (2011).

those states of affairs to obtain. It's uncontroversial that there comes a point at which the selection of means to our ultimate ends does not involve having a further thought (e.g. moving my arm might involve a shift in my shoulder blade, even if I do not have a further thought about my shoulder blade). There is, the proposal is, no guarantee that the self will always be in view by the time this point is reached.

This 'action without self-reference' idea could be given in different strengths. It could be outlining an in-principle, metaphysical possibility for agency. Or it could be outlining a possibility that human agents instantiate as they go about their business. Cappelen and Dever (2013: 40) are explicit that they need only claim the in-principle possibility. Their idea is that a need for the *de se* that is local to human agents would be of little interest, and could not in any case be discovered without detailed scientific investigation, of a sort fans of the *de se* have never rested their claims on. So considering whether an agent that was also a God could act without the *de se*, for example, would be dialectically legitimate. Most fans of the *de se* are, on the face of it, not restricting their claims to any particular category of agents (e.g., human agents, embodied agents, agents who are not Gods). A notable exception here is García-Carpintero (2017) who argues that the need for the *de se* is a deep contingent fact about human agency and not less interesting for that. Prosser (2015) explores some of the conceptual issues around Godly activity and the *de se*.

A second choice-point for *de se* fans is whether they want to claim that all action-explanations involve the *de se* (and therefore that there must be something wrong with the 'action without self-reference' proposal). Many clearly do (see, e.g., Babb 2016, Bermúdez 2017, Gjelsvik 2017, Morgan 2019b). Gjelsvik appeals to an Anscombian view on which doing something intentionally involves having practical knowledge of *I am doing it*. Babb appeals to a Castañeda inspired view on which intentional action always involves intention and intentions are, as such, *de se* (the intention *that the door be closed* is really an intention *that I bring it about that the door is closed*). The involvement of the *de se* in every case of action falls rather directly out of either of these theories. On either view, it looks as though an account of the relationship between action and the *de se* is supported – indirectly – by the full range of evidential considerations that the overall theory of action (and, in Babb's case, theory of intention) is supported by. In that case, it's hardly surprising that simply inspecting a few *de se* cases and seeing what our intuitions about them are, is not enough to show what the relationship is.

Other fans of the *de se* have seemed not to commit to the maximally strong claim that all action-explanations involve the *de se*. But, as noted, the only alternative *de se* sceptics tend to consider – the claim that some action-explanations involve the *de se* – seems too weak to be interesting. So what other kind of claim might such a fan of the *de se* appeal to? A candidate is: 'Some action-explanations involve the *de se* in a way that otherwise attractive

theory T cannot deal with'. Here the role of the *de se* is not just as a phenomenon with special features, but also as a special problem or challenge – in this case, for theory T.² Lewis conceived of *de se* cases in this way, seeing them as producing a problem for the possible worlds theory of propositional attitudes he thought otherwise attractive. So did Perry (1979), with respect to any theory that endorses what he calls 'the doctrine of propositions'. In the last decade, *de se* sceptics have put a great deal of pressure on whether the 'otherwise attractive' condition is ever really met. Perhaps theories that can't deal with *de se* cases equally can't deal with non-*de se* cases, in which case the *de se* doesn't raise a *special* problem after all. One reason this issue matters is that what form a more adequate framework will take is plausibly sensitive to the range of cases that were making difficulty for the previous framework. Moving to a framework like Lewis's centred-worlds theory, whose chief innovation is that it has a special bit of machinery to deal with *de se* cases, looks like a bad bet, if the problem provoking the move isn't in fact special to the *de se*.³

This has led in turn to greater precision from those who think *de se* cases are a distinctive problem about how the distinctive problem is supposed to arise. For example, Ninan (2016; who presents his position as a rational reconstruction of ideas that were already present in Perry and Lewis) argues roughly as follows. Consider a case where one agent believes 'I am being chased by a bear' and another believes, of her, 'You are being chased by a bear'. It looks as though this is a case in which the two agents should count as having *different beliefs*, since their expected actions are different (even if they both desire the same thing, e.g. *that nobody be eaten*). However, it also looks as though they are in complete agreement on how things are. This puts pressure on any theory on which there is a single notion of content that plays a role in capturing what it is for two subjects to agree, and that is also fine-grained enough to individuate mental states. The second, equally essential, part of the argument is the claim that it *isn't* possible to come up with non-*de se* cases that show the same thing. Ninan's idea here is that if two subjects count as agreeing in virtue of both believing, of Superman, that he has a certain property then they'll express their belief with the same sentence (it won't be that one thinks 'Superman is being chased by a bear' and

2 Magidor (2015: 249) makes this helpful distinction between treating the *de se* as having special *features* vs. posing a special *problem*.

3 See Magidor 2015 for argument that Lewis is wrong to think that, with generous enough appeal to descriptivism, the possible worlds framework that can't deal with the *de se* can deal with all non *de se* cases of opacity. Cappelen and Dever (2013) explore a different complaint against Lewis's centred-worlds theory: that the special bit of machinery it involves – the notion of self-ascription – is unexplanatory of *de se* ness. For more on the centred-possible worlds framework, see many of the essays in García-Carpintero and Torre 2016.

the other thinks ‘Clark is being chased by a bear’). But in this case, Ninan thinks, there is no particular obstacle to saying the two subjects have the same belief, because we don’t in fact expect very different actions arising from the belief each has. This argument, if sound, would seem to be a vindication of the claim that the *de se* plays a distinctive role in relation to action.⁴ But notice that there’s no incompatibility between the soundness of the argument and the in-principle possibility, or even everyday occurrence, of action without self-reference. At least some *de se* fans may thus be tempted to regard the claim that all action-explanations involve the *de se* as a red herring.

Most of the action discussed in this literature has been of a fairly unsophisticated humdrum sort – spilling sugar, hailing people, avoiding bears etc. Several recent contributions move towards normatively richer territory. Setiya (2015) discusses what he takes to be a widely accepted view that we have a distinctive kind of reason to favour ourselves, conceived *de se*. He argues that we in fact do not and that our tendency to favour ourselves is the primordial case of love at first sight. Paul (2017a, 2017b) argues that there is a notion of empathy that is both essentially *de se* and crucial to explaining our moral assessment and decision-making – being able to imaginatively project oneself into the first-person perspective of another.⁵ She also argues for a notion of *self-understanding* that is essentially *de se*. Titelbaum (2017) argues that facts about the outcomes of one’s own reasoning processes may have a different evidential significance than facts about the outcomes of others. O’Brien (2011) argues that *de se* thought – or more specifically, the subjective perspective it entails – is one of the preconditions for the various shades of self-conscious emotion, for example, shame, guilt and pride. From the point of view of evaluating *de se* scepticism, the range of points at which the *de se* might appear important makes the sceptical position more challenging to defend. Anyone who accepts the various features has an explanatory task of explaining why they are all aspects of a single kind of thought.⁶

- 4 The argument involves an ‘I-‘you’ case. Leaving aside the question of whether such cases pose a special problem for otherwise well-motivated theories, there has been a recent mini-explosion of interest, especially from those working in the neo-Fregean tradition, of how to think about the *de te*, either in its own right or in its relationship with the *de se*. Longworth (2014) argues that some *de te* thoughts are *de se* thoughts. Salje (2017a) argues that there is an irreducible mode of presentation associated with *you*. Peacocke (2014b) and Martin (2014) debate the right way to think about interpersonal self-consciousness, in which one is aware of oneself as figuring as an object in someone else’s consciousness and the connection to the *de te*. For more on the relationship between the *de se* and *de te*, see many of the essays in the *Philosophical Explorations* (2014) special issue on ‘you’-thought.
- 5 Cappelen and Dever (2017) reply to Paul by denying that the notion of empathy that matters in moral psychology involves the *de se*.
- 6 For attempts to derive various *de se* phenomena from the reference rule he takes to govern *de se* thought, see Peacocke 2012b.

3. *Immunity to error through misidentification*

A second live sub-area in *de se* thought involves a pair of epistemic properties – *error through misidentification* (EM) and IEM – as they show up in *de se* thoughts. Take a thought of the form *Fa*. Here are two ways it is potentially vulnerable to error: the thinker might be right that *something* is *F* but be wrong in identifying that thing as *a*; or she might be right that *a* is *something*, but wrong about whether it is *F*. Roughly, IEM is immunity against the first kind of mistake. Most people think that a privileged class of *de se* thoughts is IEM. (If I judge in the normal way that *I feel nervous*, then it's hard to see how I could be wrong about *who* feels nervous.) Most people also deny that *de se* thoughts are the *only* ones associated with this property – it's normally also taken to be a property of (some) demonstrative, other indexical and mathematical thoughts, and thoughts involving descriptive names.⁷ As a matter of fact, however, most work on IEM has been work on IEM as it applies to *de se* thought. This means that debates about IEM are intermingled with the *de se* in a way that has substantially affected the moves and positions that have emerged. (As we will see in a moment, an especially pronounced version of this preferential treatment can be seen in recent views of IEM that explicitly draw on metasemantic features of the *de se*.) Moreover, some theorists have looked to IEM as a precisely definable notion that might help render more tractable questions concerning the *de se* that seem to be of maximal philosophical interest but whose existing formulations make hard to adjudicate (see, e.g. Morgan 2018a).

IEM is not a new concept. Its potted history normally begins with Wittgenstein's brief remarks in *The Blue Book* (1958), through Shoemaker's elaboration of the phenomenon and coining of the phrase (1968), to Evans's (1982) influential account that (among other things) made it standard to relativize IEM to the grounds on which a thought is formed. The debate since then has been characterized by a spirited, if sometimes less than fully co-ordinated, search for the correct formulation, explanation and account of the significance of IEM. This more recent literature can sometimes be difficult to piece together, partly because different views of IEM are often developed in the context of pursuing different broader projects. However, the last decade or so in this debate has seen a new and interesting division in party lines, which we will focus on here.

The first camp emerging from the split is the *identification-freedom* approach, characterized by the claim that the presence of IEM is always explained by the absence of an identificatory component in the thought's grounds. If the thought *I am F* isn't reached by putting together a distinct thought of the form *b is F* with an identity of the form $b = I$, then it looks as

7 See also Smith 2018 for a discussion about whether it also applies to *plural* first-person reference (*we*-thought).

though there is no point at which an error of *misidentification* could get introduced. After all, a misidentification is just an identification that has gone wrong, and grounds of this sort do not involve an identification. On the *identification-freedom* approach, then, a *de se* thought is IEM when, and because, it is formed in a way that involves directly apprehending oneself in an exclusively *de se* way. A merit of this approach to IEM is that it ‘liberates us from any need for metaphysical or metasemantic extravagance in trying to account for the phenomenon’ (Wright, 2012: 255).

There are a number of internal disagreements among contemporary identity-freedom theorists.⁸ One is how to understand the idea of *grounds*. Nearly no-one thinks that the possibility of EM requires the performance of a conscious inference (*b is F, I=b, so I am F*). But there are still varyingly demanding ways of understanding what *is* needed. Does the identity component need to be part of a rational reconstruction of the subject’s reasoning? Would it suffice for it to be part of a thinker’s pre-suppositional background to which she has default entitlement, if not ready access? Or does it need to be accessible to the subject, at least in principle? Different identity-freedom theorists answer these questions differently, resulting in formulations of IEM properties of different strengths that sort cases differently.⁹

The most urgent current issue for the identity-freedom camp, however, is an external challenge descending from Pryor’s influential 1999 distinction between two kinds of error (and immunity) through misidentification.¹⁰ The first, *de re misidentification*, fits neatly with the identity-freedom approach: where there is *de re* misidentification, the belief’s grounds involve a false identity belief. The second is the notion of *wh-misidentification* and is still usually introduced with Pryor’s original example:

I smell a skunky odor, and see several animals rummaging around in my garden. ... Approaching closer and sniffing I form the belief, of the smallest of 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. (1999: 281)

8 One prominent continuing issue is whether to admit different grades of IEM, such as absolute/logical/de jure/de facto, or to only treat one of these as genuine IEM (see, e.g., Coliva 2012, 2017, Recanati 2012, Prosser 2012, García-Carpintero 2013, 2018 and McGlynn 2016). Another is how to treat the presence of indubitably known identities (see, e.g., Smith 2006 and Coliva 2017).

9 Sometimes this issue is posed in terms of *justificational architecture*, *doxastic or epistemic structure*, *formative structure* or *explanatory/causal basis* or discussed in the context of a distinction between *liberal* and *conservative* views of epistemic justification; for more discussion see Coliva 2006, 2012, Wright 2012, McGlynn 2016, MS, Morgan 2018a and García-Carpintero 2013, 2018.

10 Pryor 1999; several recent writers have attempted to improve on Pryor’s original formulation (e.g. Hu 2017, Merlo 2017, Morgan 2018b, Recanati 2012 and McGlynn 2016, MS).

My grounds for the belief *That is a skunk* seems not to rest on a false identity of the form $a = b$. Instead, I have grounds for an existentially general belief about a property instantiation (*Something hereabouts is a skunk*, based on smell), but I have picked out the wrong witness to that existential. It is still appropriate to say that my thought is mistaken through a misidentification. But it is a ‘misidentification’ in the sense of being a faulty ‘singling out’ of the right witness, rather than a faulty identity of the form $a = b$. Such cases appear to show that the fact that a thought includes nothing of the form $a = b$ in its grounds is compatible with its *not* being IEM. That puts pressure on the claim that a thought’s including nothing of the form $a = b$ in its grounds is ever enough to explain its being IEM, even in case of thoughts that have both properties.¹¹

Coliva kicked off a lively response to this challenge in her (2006), by arguing that all seeming cases of *wh*-IEM can be reduced to *de re*-IEM. Her strategy is to show that even in these cases, the grounds for the subject’s thought still involve a (mis)identity. It’s just that one side of the identity will feature a non-*de re* – but still singular – concept. For example, she offers the following reconstruction of the false identity involved in the skunk example:

This animal (I can now see) = the animal (in my garden) which is actually responsible for this odor I can smell. (412)

Objectors have variously responded that this strategy illegitimately makes use of a uniqueness assumption (who says there was only *one* thing I was smelling in the original case?) relies on idiosyncrasies of the showcase example, or that the descriptive singular concepts invoked by this strategy will always be too trivial to provide genuinely informative identity beliefs.¹²

The second emerging camp from the recent split in these debates is arguably better suited to accommodating *wh*-IEM. Call it the *metasemantic* approach (and recall that this is the kind of ‘metasemantic extravagance’ Wright hoped the identification-freedom approach might help us avoid). The metasemantic approach is newer and more heterogeneous than the identification-freedom approach, but can be grouped by a unifying ambition to provide an explanation of the IEM of *de se* thoughts that springs from an account of how the reference of *de se* thoughts is fixed, rather than just from the grounds of particular *de se* thoughts. As Peacocke writes:

Whenever there are phenomena distinctive of a given concept, a good theory of that concept should explain those phenomena. This general point applies when the phenomena are those of immunity to error

11 See, for example, Guillot 2014 and García-Carpintero 2013, 2018 for this objection to the identification-freedom view.

12 For objectors to Coliva’s strategy, see Wright 2012, McGlynn 2016, MS, García-Carpintero 2018 and Palmira 2019; Recanati (2012) and Prosser (2012) both endorse versions of Coliva’s strategy.

through misidentification, and the concept is the first person concept *I*. (2012: 144)

Different contributors to this approach follow this injunction in different ways depending on their preferred metasemantic account of the *de se* – the view that introspective impressions play a role in fixing the reference of *de se* thoughts (Palmira MS, García-Carpintero 2018) or where those thoughts are otherwise fixed in the manner of perceptual demonstratives (Morgan 2015), will explain cases of *de se* IEM differently from a view on which it is a version of the token-reflexive reference rule that does the explanatory work (Peacocke 2012b, 2014a, Hu 2017, see also Prosser 2012). It is not possible to survey all the options here. We will instead present one prominent version of the approach from Recanati (2007, 2009, 2012).

Recanati posits a category he calls *implicit de se* thoughts, whose distinguishing feature is that although they systematically concern the thought's thinker, that thinker is not represented in the thought's content. The thought's content is *thetic* (or gappy), and has a truth value only when evaluated relative to a particular subject.¹³ When these thoughts are based directly on experience, the subject is delivered by the *mode* of the experience (approximately, the type of experiential state) – not because the mode fills in a representation of the subject, but because it constrains who the subject must be. When I have the thought I would express as 'My arms are outstretched' on the basis of proprioception, the content of my thought is something like the open sentence '___ arms are outstretched'. The proprioceptive mode on which it is based, however, delivers *me* as its relevant circumstance of evaluation, because proprioception can only ever deliver information about oneself. Now, according to Recanati such implicit thoughts based directly on experience will always be IEM. Why? Because given that I am not represented at all in either the content of the thought or experience on which it is based, the basic conditions for any error of *misidentification* (in either sense) are not met.¹⁴

Since nothing in this style of account depends on specific claims about the thought's grounds, it isn't per se committed to the claim that IEM is always explained by the absence of an identity component in those grounds. As a result, metasemantic approaches seem better positioned to accommodate the notion of wh-IEM. This advantage is unlikely to impress opponents following Coliva, who anyway think that the idea of an independent notion of

13 This is the treatment Lewis gives of all *de se* attitudes.

14 Cf. accounts of IEM by Ismael (2012) and Guillot (2013). Recanati also thinks that there is a second layer of *de se* thoughts with IEM – *explicit de se* thoughts – that *do* explicitly represent the thought's subject. These *de se* thoughts are reached by a non-inferential process called 'reflection' from implicit *de se* thoughts – since no inference is involved, there is no new opportunity for a misidentification to be introduced. This addition to the account is not important for our purposes.

wh-IEM is spurious.¹⁵ Other objections that have been raised against versions of the metasemantic approach include the comparative complexity of the account, the worry that explanations at this level cannot account for the basis relativity or non-universality of *de se* IEM, the fact that the appeal of any one of these accounts of IEM is hostage to the appeal of a specific metasemantic view of the *de se*, and suspicions about the scope on this approach for a unifying account of IEM beyond the case of the *de se*.¹⁶

The two approaches to *de se* IEM explanation described in this section are pitched at a high level of abstraction. There has also been plenty of ground-level work in the last 10 years about the extension of *de se* IEM. It is striking that these first-order debates are similarly characterized by a lack of agreement: most (but not all) consider the IEM of introspective *de se* judgements to be non-negotiable but other candidates – including bodily awareness, agentive and memory *de se* judgements – remain controversial.¹⁷

Indeed, the absence of fixed points in the debate about *de se* IEM might raise alarm bells for some.¹⁸ There is disagreement in this literature on, in summary, all of the following: what the extension of *de se* IEM is; what the best explanation of it is; whether it subdivides into importantly different sub-categories (and, if so, which these are); what significance it has for an understanding of *de se* thought; and whether the phenomenon of IEM has similar interest in non-*de se* cases as in *de se* cases. What's more, the discussion has been intricate and long-running enough that an author of a paper might easily, and quite reasonably, concentrate on a highly localized aspect of one of these questions, without engaging with the overall question of why their doing so matters. This would be unproblematic in the context of a shared, unvoiced understanding of what is up for grabs in the debate. But it isn't obvious that such an understanding exists. What does this mean for the future of these debates? As with post-Gettier attempts to rehabilitate the JTB account of knowledge, which eventually resulted in wholesale changes of approach to the relevant sub-discipline, there is the possibility that at a certain point *de se* theorists will just start playing a different game. Perhaps the

15 See also Smith 2006 for another argument against wh-IEM, and McGlynn MS for a response; see also Salje 2016 for an objection to a closely related formulation of IEM.

16 See, for example, Wright 2012, Morgan 2012, Coliva 2017 and Hu 2017. See also Campbell 2012b for an argument that *de se* thought is *unique* in resisting a metasemantic explanation of IEM.

17 For some recent discussions of the IEM of introspective judgements see García-Carpintero 2018, Hu 2017, Parrott 2017; of bodily awareness judgements De Vignemont 2011, 2012, Schwenkler 2013a, Morgan 2018b, Salje 2017b, Lane and Liang 2011, Gallagher 2012, and Kang 2016; of action judgements O'Brien 2007, 2011, Romdenh-Romluc 2013, and Lane and Liang 2011; and of memory judgements Fernández 2014, Bermúdez 2012, 2013, and Hamilton 2009.

18 For example, Schwenkler 2013b and Howell 2007. A more extreme reaction to the state of this literature is to deny the existence of IEM altogether (e.g. Cappelen and Dever 2013).

concept of IEM is a tool that initially appeared helpful in understanding the *de se* but turned out not to be. This would, of course, be compatible with thinking that many of the ostensibly IEM-focused papers contain important insights – for example, about the nature of various bodily illusions or the notion of thought authorship – that are achieved, as it were, on the way to a claim about IEM. An opposing view would be that it is simply a datum that certain *de se* judgements are IEM – however unclear it is exactly which these are, or how best to define the phenomenon in general terms – and eschewing the topic would just involve ignoring a question because it has proven hard to answer.

4. Conclusion

None of the debates about the *de se* – about its significance, distinctive features or even existence – show any sign of letting up. As a topic with a long history, the continuing vigour of these discussions is testament to the tenacity and depth of the philosophical puzzles that give rise to them. While there is a widely shared instinct that there is something special about the *de se* that is illustrated by *de se* cases, it has proven surprisingly difficult to distil that significance into a single, simple point. And while there is a persistent sense that *de se* IEM tracks something of real importance, decades of sustained effort haven't yet given us a clear shared idea of what that is. What is clear, however, is that there is a rich seam of philosophical material in this area that ramifies across mind, language, morality, theory of action, metaphysics of the self – indeed arguably across all areas of human-centred philosophy.¹⁹

*School of Philosophy, Religion and History of Science
University of Leeds
Leeds LS2 9JT, UK
l.c.salje@leeds.ac.uk*

*University of York
York YO10 5DD, UK
d.morgan@york.ac.uk*

References

Babb, M. 2016. The essential indexicality of intentional action. *Philosophical Quarterly* 66: 439–57.

19 For discussion and comments, we owe thanks to Jessica Begon, Aidan McGlynn, Dilip Ninan, Michele Palmira and John Schwenkler.

- Bermúdez, J.L. 2012. Memory judgments and immunity to error through misidentification. *Graxer Philosophische Studien* 84: 1.
- Bermúdez, J.L. 2013. Immunity to error through misidentification and past-tense memory judgments. *Analysis* 73: 211–20.
- Bermúdez, J.L. 2016. *Understanding 'I'*. Oxford: Oxford University Press.
- Bermudez, J.L. 2017. *Understanding "I": Language and Thought*. Oxford: Oxford University Press.
- Cappelen, H. and J. Dever. 2013. *The Inessential Indexical: On the Philosophical Insignificance of Perspective and the First Person*. Oxford: Oxford University Press.
- Cappelen, H. and J. Dever. 2017. Empathy and transformative experience without the first person point of view (a reply to L. A. Paul). *Inquiry* 60: 315–36.
- Campbell, J. 2012a. Lichtenberg and the Cogito. *Proceedings of the Aristotelian Society* 112: 361–78.
- Campbell, J. 2012b. On the thesis that 'I' is not a referring term. In *Immunity to Error through Misidentification: New Essays*, eds. S. Prosser and F. Recanati, 361–78. London: Cambridge University Press.
- Coliva, A. 2006. Error through misidentification: some varieties. *Journal of Philosophy* 103: 403.
- Coliva, A. 2012. 'Which 'Key to all the Mythologies' about the Self? A note on where illusions of transcendence come from and how to resist them. In *Immunity to Error through Misidentification: new essays*, eds S. Prosser and F. Recanati. Cambridge: Cambridge University Press.
- Coliva, A. 2017. Stopping points: 'I', immunity and the real guarantee. *Inquiry* 60: 233.
- De Vignemont, F. 2011. A self for the body. *Metaphilosophy* 42: 230–47.
- De Vignemont, F. 2012. Bodily immunity to error. In *Immunity to Error through Misidentification: New Essays*, eds. S. Prosser and F. Recanati, 224–46. Cambridge: Cambridge University Press.
- Devitt, M. 2013. The myth of the problematic de se. In *Attitudes De se: Linguistics, Epistemology, Metaphysics*, eds. A. Capone and N. Feit, 133–62. Stanford: CSLI.
- Doyle, J. 2018. *No Morality, No Self: Anscombe's Radical Skepticism*. Harvard: Harvard University Press.
- Evans, G. 1982. ed. *The Varieties of Reference*, ed. J. McDowell. Oxford: Oxford University Press.
- Fernández, J. 2014. Memory and immunity to error through misidentification. *Review of Philosophy and Psychology* 5: 373.
- Gallagher, S. 2012. First-person perspective and immunity to error through misidentification. In *Consciousness and Subjectivity*, eds. S. Miguens and G. Preyer, 47–245. Ontos.
- García-Carpintero, M. 2013. The self file and immunity to error through misidentification. *Disputatio* 5: 191–206.
- García-Carpintero, M. 2017. The philosophical significance of the De Se. *Inquiry* 60: 3.
- García-Carpintero, M. 2018. *De se* thoughts and immunity to error through misidentification. *Synthese* 195: 3311.
- García-Carpintero, M. and S. Torre, eds. 2016. *De Se Thought and Communication*. Oxford: Oxford University Press.
- Gjelsvik, O. 2017. Indexicals: what they are essential for. *Inquiry* 60: 295–314.

- Guillot, M. 2013. The limits of selflessness: semantic relativism and the epistemology of *de se* thoughts. *Synthese* 190: 1793.
- Guillot, M. 2014. Identification-free at last. Semantic relativism. Evans's legacy and a unified approach to immunity to error through misidentification. *Teorema: International Journal of Philosophy* (3): 7–30.
- Guillot, M. and M. García-Carpentiro. forthcoming. *The Sense of Mineness*. Oxford: Oxford University Press.
- Hamilton, A. 2009. Memory and self-consciousness: immunity to error through misidentification. *Synthese* 171: 409.
- Hamilton, A. 2013. *The Self in Question: Memory, the Body and Self-Consciousness*. London: Palgrave Macmillan.
- Howell, R. 2007. Immunity to error and subjectivity. *Canadian Journal of Philosophy* 37: 581–604.
- Hu, I. 2017. The epistemology of immunity to error through misidentification. *Journal of Philosophy* 114: 113.
- Ismael, J. 2012. Immunity to error as an artefact of transition between representational media. In *Immunity to Error through Misidentification: New Essays*, eds. S. Prosser and F. Recanati, 62–80. Cambridge: Cambridge University Press.
- Johnston, M. 2010. *Surviving Death*. Princeton: Princeton University Press.
- Kang, S.-P. 2016. The body swap illusion, and immunity to error through misidentification. *Journal of Philosophy* 113: 463.
- Lane, T. and C. Liang. 2011. Self-consciousness and immunity. *Journal of Philosophy* 108: 78.
- Lewis, D. 1979. Attitudes *De Dicto* and *De Se*. *The Philosophical Review* 88: 513–43.
- Longuenesse, B. 2017. *I, Me, Mine: Back to Kant, and Back Again*. Oxford: Oxford University Press.
- Longworth, G. 2014. You and me. *Philosophical Explorations* 17: 289–303.
- Madden, R. 2011. Intention and the self. *Proceedings of the Aristotelian Society* CXI: 325–51.
- Madden, R. 2013. Could a brain in a vat self-refer? *European Journal of Philosophy* 21: 1.
- Magidor, O. 2015. The myth of the *de se*. *Philosophical Perspectives* 29: 249–83.
- Martin, M.G.F. 2014. In the eye of another: comments on Christopher Peacocke's interpersonal self-consciousness. *Philosophical Studies* 170: 25–38.
- McGlynn, A. 2016. Immunity to error through misidentification and the epistemology of *de se* thoughts. In *About Oneself: De Se Thought and Communication*, eds. M. García-Carpintero and S. Torre, 25–55. Oxford: Oxford University Press.
- McGlynn, A. MS. Immunity to wh-misidentification. Unpublished manuscript.
- Merlo, G. 2017. Three questions about immunity to error through misidentification. *Erkenntnis* 82: 603.
- Millikan, R.G. 1990. The myth of the essential indexical. *Noûs* 24: 5.
- Morgan, D. 2012. Immunity to error through misidentification: what does it tell us about the *de se*?. In *Immunity to Error through Misidentification: New Essays*, eds. S. Prosser and F. Recanati, 104–23. Cambridge: Cambridge University Press.
- Morgan, D. 2015. The demonstrative model of first-person thought. *Philosophical Studies* 172: 1795.

- Morgan, D. 2018a. Thinking about the body as subject. *Canadian Journal of Philosophy* 49: 435–57.
- Morgan, D. 2018b. Impersonal intentions. *Philosophical Quarterly* 68: 376–84.
- Morgan, D. 2019a. Temporal indexicals are essential. *Analysis* 79: 452–61.
- Morgan, D. 2019b. Accidentally about me. *Mind* 128: 1085–115.
- Musholt, K. 2015. *Thinking about Oneself: From Nonconceptual Content to the Concept of a Self*. Cambridge, MA: MIT Press.
- Ninan, D. 2016. What is the problem of De Se attitudes? In *About Oneself: De Se Thought and Communication*, eds. S. Torre and M. García-Carpentiro. Oxford: Oxford University Press.
- O'Brien, L. 2011. Ordinary self-consciousness. In *Consciousness and the Self: New Essays*, eds. J.L. Liu and J. Perry, 101–22. Cambridge: Cambridge University Press.
- O'Brien, L. 2007. *Self-Knowing Agents*. Oxford: Oxford University Press.
- O'Brien, L. 2015. Ambulo ergo sum. *Royal Institute of Philosophy Supplement* 76: 57–75.
- Palmira, M. 2019. Arithmetic judgments, first-person judgments, and immunity to error through misidentification. *Review of Philosophy and Psychology* 10: 155.
- Palmira, M. MS. Immunity, thought insertion and the first person concept. Unpublished manuscript.
- Parrott, M. 2017. Subjective misidentification and thought insertion. *Mind and Language* 32: 1.
- Paul, L. 2017. De se preferences and empathy for future selves. *Philosophical Perspectives* 31: 7–39.
- Paul, L.A. 2017. First person modes of presentation and the structure of empathy. *Inquiry: An Interdisciplinary Journal of Philosophy* 60: 189–207.
- Peacocke, C. 2008. *Truly Understood*. Oxford: Oxford University Press.
- Peacocke, C. 2012a. Descartes defended. *Aristotelian Society Supplementary Volumes* 86: 1.
- Peacocke, C. 2012b. Explaining *de se* phenomena. In *Immunity to Error through Misidentification: New Essays*, eds. S. Prosser and F. Recanati, 144–57. Cambridge: Cambridge University Press.
- Peacocke, C. 2014a. *The Mirror of the World Subjects, Consciousness, and Self-Consciousness*. Oxford: Oxford University Press.
- Peacocke, C. 2014b. Interpersonal self-consciousness. *Philosophical Studies* 170: 1–24.
- Perry, J. 1979. The problem of the essential indexical. *Nous* 13: 3–21; reprinted in Perry 1993, 3–52.
- Perry, J. 1993. *The Problem of the Essential Indexical and Other Essays*. New York: Oxford University Press.
- Prosser, S. 2012. Sources of immunity to error through misidentification. In *Immunity to Error through Misidentification: New Essays*, eds. S. Prosser and F. Recanati, 158–79. Cambridge: Cambridge University Press.
- Prosser, S. 2015. Why are indexicals essential? *Proceedings of the Aristotelian Society* 115: 211.
- Pryor, J. 1999. Immunity to error through misidentification. *Philosophical Topics* 26: 271.

- Recanati, F. 2007. *Perspectival Thought: A Plea for Moderate Relativism*. Oxford: Oxford University Press.
- Recanati, F. 2009. *De re* and *de se*. *Dialectica* 63: 249.
- Recanati, F. 2012. Immunity to error through misidentification: what it is and where it comes from. In *Immunity to Error through Misidentification: New Essays*, eds. S. Prosser and F. Recanati, 180–201. Cambridge: Cambridge University Press.
- Rödl, S. 2018. *Self-Consciousness and Objectivity: An Introduction to Absolute Idealism*. Harvard: Harvard University Press.
- Romdenh-Romluc, K. 2013. First-person awareness of intentions and immunity to error through misidentification. *International Journal of Philosophical Studies* 21: 493.
- Salje, L. 2016. The subjective perspective in introspection. *Journal of Consciousness Studies* 23: 3–4.
- Salje, L. 2017a. Thinking about you. *Mind* 126: 817–40.
- Salje, L. 2017b. Crossed wires about crossed wires: somatosensation and immunity to error through misidentification. *Dialectica* 71: 35.
- Schwenkler, J. 2013a. The objects of bodily awareness. *Philosophical Studies* 162: 465.
- Schwenkler, J. 2013b. Review: ‘Immunity to Error through Misidentification’ edited by Simon Prosser and François Recanati. *Analysis* 73: 180.
- Setiya, K. 2015. Selfish reasons. *Ergo* 2: 18.
- Shoemaker, S. 1968. Self-reference and self-awareness. *Journal of Philosophy* 65: 555–67.
- Smith, J. 2006. Which immunity to error? *Philosophical Studies* 130: 273.
- Smith, J. 2018. The first-person plural and immunity to error. *Disputatio* 10: 141.
- Spencer, C. 2007. Is there a problem of the essential indexical? In *Situating Semantics: Essays on the Philosophy of John Perry*, eds. M. O’Rourke and C. Washington, 179–97. Cambridge, MA: MIT Press.
- Stanley, J. 2011. *Know How*. Oxford: Oxford University Press.
- Titelbaum, M. 2017. One’s own reasoning. *Inquiry: An Interdisciplinary Journal of Philosophy* 60: 208–32.
- Torre, S. 2018. In defense of De Se content. *Philosophy and Phenomenological Research* 97: 1.
- Wittgenstein, L. 1958. *The Blue and Brown Books: Preliminary Studies for the ‘Philosophical Investigations’*. Oxford: Blackwell Publishing.
- Wright, C. 2012. Reflections on François Recanati’s, ‘Immunity to error through misidentification: what it is and where it comes from’. In *Immunity to Error through Misidentification: New Essays*, eds. S. Prosser and F. Recanati. Cambridge: Cambridge University Press.