The Myth of the De Se

Ofra Magidor

[Advanced Draft - July 2014. Final version to appear in *Philosophical Perspectives*]

Introduction

An increasingly popular line of thought in both philosophy and linguistics is one that I will call, unsympathetically, 'The myth of the de se'. Like most myths, it is not easy to articulate precisely what its content is, and different versions of the myth bear at best a family resemblance to each other. But at least one central line of thought which runs at the core of the myth consists of the following:

> There is a special class of propositional (or 'propositional-like') attitudes. These are *self-locating* or *de se* attitudes, ones that are typically expressed using indexical expressions such as 'I' and 'now' (call this claim '*Special Attitudes*'). Moreover, such attitudes pose a special challenge for our account of propositional attitudes. In other words, assume one starts with what might otherwise be considered an adequate account for standard (non de se) attitudes.[1] Once we take on board de se attitudes, this account ought to be fundamentally amended (call this second claim '*Special Challenge*').

No doubt the most prominent and influential defence of the myth appears in Lewis's seminal paper 'Attitudes De Dicto and De Se' (Lewis (1979), henceforth 'ADD'), where Lewis

---

[1] When no confusion arises, I follow Lewis (1979) in referring to these as 'de dicto' attitudes. It should be noted, however, that in this use the de dicto is intended to be contrasted with the de se, rather than the de re. When my focus will be on the de dicto/de re distinction, I talk of 'purely de dicto' attitudes.

argues that accommodating de se attitudes requires shifting from the possible-worlds account of attitudes to the centred-worlds account (more on this in §1). Indeed while Lewis and the many contemporary adherents of the centred-worlds framework for handling de se attitudes are the principal defenders of the myth[2], they are not the only ones. Consider for example, Perry's classic paper 'The problem of the essential indexical' (Perry (1979)). Perry maintains that beliefs involving indexicals pose a special challenge to propositional attitude reports (even though the way he chooses to address that challenge is very different than that of Lewis). [3],[4]

My aim in this paper is to debunk the myth of the de se. While the discussion will touch on various lines of thought that go along with the myth, my main target will be the central tenet

---

[2] For further theorists who adopt the centred-worlds framework for handling de se attitudes, see for example Cresswell (1985) ch. 14, Chalmers (2006), Egan (2007), Elga (2000), Hitchcock (2004), Liao (2012), Meacham (2008), Moss (*forthcoming*), Ninan (2009), Nolan (2006), Stephenson (2010), Turner (2010), and Torre (2010) (While not all of the authors in this list accept every aspect of Lewis's view or endorse every one of his arguments in favour of the shift to the centred-worlds framework, they do share the assumption that the phenomenon of de se attitudes requires such a shift).

[3] As he puts it: "I argue that the essential indexical poses a problem for various otherwise plausible accounts of belief" (Perry (1979), p.3). One somewhat confusing aspect of Perry's view in this context is that the way he ultimately chooses to address the challenge is one which is broadly consistent with some of what I argue for below (for example, although his view focuses in addressing problems with indexical beliefs in particular, the view he opts for arguably has the resources for addressing the Frege's puzzle more generally). But Perry should at least be considered to be a defender of the myth in so far as *he* construes the issue as a special challenge posed by de se beliefs (cf. the above quote). Moreover, Perry clearly supports Special Attitudes. Some worries concerning this claim (in particular as it is interpreted by Perry) are raised in §1.5, §1.6, §2.3 and in the conclusion (§4).

[4] For further support of the myth, not particularly in the centred-world framework, see also Castaneda (1966), Chierchia (1989), Perry (1977), and Schlenker (*forthcoming*). It should be noted that there have also been some critiques of the myth (see Boer & Lycan (1980), Devitt (*forthcoming*), Millikan (1990), Spencer (2007), Stanley (2011), ch.3, Tiffany (2000), and more implicitly Stalnaker (1981)). Some considerations these raise – especially the analogies between the problems concerning self-locating beliefs and Frege's puzzle – overlap to a certain extent with my remarks in §1 of this paper, but my criticisms are developed in a substantially different way. More notably, Cappelen and Dever's book (2013) appeared after completing this paper. Their book pursues many similar themes to the ones in this paper, and I am sympathetic to much of what they have to say. Their book does, however, differ from the current work in a range of matters of detail, as well as in some more central respects (it places far less emphasis on the role of possible-worlds semantics in defending the myth, and does not contain any material that parallels my discussion in §2 and §3 of this paper). I have not referenced the similarities and differences between our arguments throughout the text – hopefully, the reader can take both works to jointly provide support for the line we are pursuing.

of the myth outlined above, and in particular the Special Challenge aspect of the myth.[5] Moreover, I focus primarily (though not exclusively) on the Lewisian defence of the myth: this has the advantage of keeping the discussion reasonably concrete, and is also important because the Lewisian version of the myth is by far the most widely accepted one.

In §1, I present Lewis's view and his arguments in support of the move to a centred-worlds account of attitudes. I argue that Lewis's arguments are unpersuasive. In §2, I briefly highlight some additional problems that the shift to the centred-worlds account introduces. In §3, I present what I take to be the strongest argument in favour of the shift to the centred-worlds account ('the reducibility argument'). While this argument has not been presented by defenders of the myth in this precise form, it echoes various lines of thought brought up in this context, and I suspect that something like this argument motivates many who endorse the myth. I argue, however, that although initially compelling, the reducibility argument is ultimately unsuccessful. I conclude in §4, noting what implications my discussion has to the Special Attitudes aspect of the myth. The appendix contains some brief remarks on another aspect of the myth: the alleged connection between de se attitudes and indexical reports in general, and PRO reports in particular.

---

[5] However, in so far as this claim presupposes Special Attitudes, we need to consider both claims, and moreover in various places throughout the paper I also target Special Attitudes (see the conclusion for a brief summary of this issue).

§1 Lewis's centred-worlds account and its motivations

§1.1 Lewis's account

The most prominent and influential defence of the myth, one which will occupy a large part of this paper, appears in Lewis's seminal paper 'Attitudes de dicto and de se' (Lewis (1979)). Lewis's discussion begins with the assumption that, de se attitudes aside, an adequate account of propositional attitudes is the *possible-worlds account*. Consider for example the attitude of belief. According to the possible-worlds account, an agent's total belief state is taken to be a set of possible-worlds (intuitively: those worlds w such that it is compatible with the agent's beliefs that w obtains). Propositions are sets of possible-worlds (intuitively: the set worlds in which the proposition in question is true). Finally, an agent X whose total belief state is S believes the proposition p, if and only if, $S \subseteq p$ (or to put it otherwise: iff p is true in all the worlds in the agent's total belief state).

But crucially, Lewis argues that in light of de se attitudes, the account needs to be rejected in favour of the *centred-worlds* account. Let a *centred-world* be a pair $<w,x>$, such that w is an a possible world, and x is an individual.[6] Consider the attitude of belief. On the centred-worlds account, an agent's total (centred) belief state is taken to be a set of centred-worlds (intuitively: those centred-worlds $<w,x>$ such that is compatible with the agent's beliefs that w obtains, and she is x). Let a *centred-proposition* be a set of centred-worlds. An agent X whose total centred belief state is S believes the centred-proposition p, if and only if $S \subseteq p$. For example, suppose Jill believes (de se) that she is tall.[7] This will be construed as an

---

[6] Adherents of the centred worlds account often also add a time coordinate in order to account for beliefs such as my belief that the meeting starts *now*. For mere ease of readability, I leave out the time parameter throughout. Others restrict centred-worlds to those pairs $<w,x>$ such that x exists in w. This will not make any difference to my arguments below.

[7] The qualification in parentheses is intended to indicate that Jill's belief is taken to be an instance of the special kind of attitude posited by defenders of the myth. (The qualification is needed, because as noted in the appendix,

attitude towards the centred-proposition {<w,x>: x is tall in w}. Jill believes this centred-proposition just in case all the centred-worlds in her belief state are included in the proposition, i.e. just in case she believes that whichever individual she might be and whatever the world might be like, the world is such that she is tall.

Finally, note that Lewis (with good reason) does not wish to have bifurcated account with separate treatments for de se attitudes and for ordinary de dicto beliefs. To avoid this, Lewis maintains that the centred-worlds account can handle not only de se attitudes, but also ordinary de dicto attitudes. Belief – whether de se or de dicto – is always an attitude towards a centred-proposition. But while Jill's (de se) belief that she is tall is an attitude towards the 'interesting' centred-proposition: {<w,x>: x is tall in w}, her merely de dicto belief that someone is tall is an attitude towards the 'boring' centred-proposition: {<w,x>: such that someone is tall in w}.[8]

§1.2 The original sin

Before I turn to discuss the arguments that motivate Lewis's view, there is an important issue that needs to be highlighted concerning the general structure of his arguments. Taken in very general terms, Lewis's defence of the centred-worlds account proceeds via the following two claims:

Claim One: De se attitudes aside, the possible-worlds account is an adequate account of attitudes.

---

the mere use of the indexical in a belief report does not guarantee that we are reporting a distinctively de se attitude).

[8] The terminology of 'boring' and 'interesting' centred-propositions follows Egan (2006). I discuss the distinction between the kinds of centred-propositions in more detail in §2.2 below.

Claim Two: However, in order to accommodate de se attitudes, one needs to shift to the centred-worlds account.

Lewis's arguments focus on defending Claim Two. But it is important to realise that Claim One is just as essential for the arguments to go through. After all, if the possible-worlds account is *not* an adequate account for de dicto attitudes, one might worry that the problems Lewis raises concerning de se attitudes have nothing in particular to do with the de se - perhaps they are merely by-products of using an inadequate account as one's starting point, or to put it otherwise perhaps they are merely particular instances of the more general problems that plague the possible-worlds framework.

This is not merely a hypothetical concern: there are well-known reasons to worry that the possible-worlds account (at least without further amendments) is indeed inadequate. Recall the following related problems:

The problem of necessary truths: On the possible-worlds account, trivially, every agent believes every necessary truth (including highly complex logical truths, and all metaphysical necessities).

The problem of necessary falsehoods: On the possible-worlds account, it is practically impossible for an agent to believe any necessary falsehood[9] (including highly complex false logical claims, and metaphysical necessary falsehoods).

---

[9] Technically, the account allows for an agent to believe a necessary falsehood. But in that case, their total belief state must be an empty set of worlds, which in turn entails that for any proposition p, the agent believes p. Since presumably, no agent believes every proposition, it follows that no agent believes any necessary falsehoods.

The problem of entailment: On the possible-worlds account, if an agent believes *p*, and *q* is an (strict or necessary) entailment of *p*, then the agent believes *q* (even if *q* is entailed from *p* via some highly complex logic, or as a matter of a metaphysical rather than logical necessity).

Frege's puzzle: More generally, the possible worlds account assumes an extremely coarse-grained view of belief. For example, if an agent believes that Hesperus is bright, then on the possible-worlds account, she thereby also believes that Phosphorus is bright.

Generalisation to other attitudes: it is important to keep in mind that the possible worlds account is not intended just as an account of belief, but rather offers a framework for handling all propositional attitudes. This means that the above problems generalise to other attitudes: for example, all agents *know* all necessary truths; if an agent *wants* it to be the case that p, and p entails q, then the agent wants it to be the case that q; If an agent *hopes* that Superman might be in the neighbourhood, then he thereby also hopes that Clark Kent is in the neighbourhood; and so forth.

These problems are of course well-known and defenders of the possible-worlds account have proposed various strategies to try and handle these problems.[10] Interestingly, Lewis himself mentions this issue at the outset of ADD, where he makes the following remark:

> "*You may think that it goes without saying that the objects of attitudes are not sets of worlds, because, for instance, believing that 2+2=4 is not the same as believing that 123+456=579 though both these equations hold at exactly the same worlds – namely all. I know perfectly well that there is such a thing as ignorance of non-contingent matters. I do*

---

[10] See for example Lewis (1986), pp. 34-36 and Stalnaker (1984), ch. 5. I discuss one central strategy for dealing with these issues in detail in §3 below.

not know what the proper treatment of such ignorance is;… My hunch is that this problem cuts across the issues I want to discuss, so I shall ignore it". (ADD, p. 135)

I agree with Lewis that if these issues were to cut across the problems he discusses then we can leave them aside for current purpose (after all, not all philosophical problems need to be solved on a single occasion…). However, I argue below that Lewis's 'hunch' is entirely wrong: most of the problems concerning de se attitudes that motivate Lewis's shift to the centred-worlds account are merely instances of these more general problems.[11]

A final preliminary remark is in order. While I do not wish to develop or defend here any particular alternative account of attitudes, it will be helpful in what follows to recall the main alternatives to the possible-worlds account and how these address the above problems. In very schematic terms we have the following strategies available.

First, one can take an agent's total belief state to be a set of propositions rather than a set of worlds (where an agent believes the proposition p if and only if p is a member of the agent's total belief state). This in itself will help address at least some of the above problems, even if propositions are taken to be highly coarse-grained (i.e. sets of possible-worlds). Consider for example the problem of necessary truths: it is not required that the necessary proposition (the set of all worlds) be a member of the set of propositions constituting a certain agent's total

---

[11] To emphasize: my purpose here is not to argue that defenders of the possible-worlds account cannot ultimately provide a satisfactory response to these problems (as it turns out, I am sceptical that they can – but I won't settle that issue here…). Rather, the crucial question is whether, if such a defence were to be offered, it would not *ipso facto* address the problems that Lewis raises with respect to the *de se*. What I am objecting to is a common attitude among defenders of the myth: one that dismisses the general problems as an inconsequential idealisation, while taking the instances involving the de se very seriously.

belief state; or consider the problem of necessary falsehoods: there is nothing to preclude the necessarily false proposition (the empty set of worlds) to be a member of any agent's total belief state.

It is clear, however, that this strategy will only go so far. If propositions are as coarse-grained as sets of worlds, then on the current proposal, if an agent believes *one* necessary truth (e.g. that 2+2=4) it would follow that she believes all of them (including e.g. Fermat's Last Theorem); Similarly, if she believes that Hesperus is bright, she will thereby also believe that Phosphorus is bright.

There are two broad strategies for handling this last set of problems.[12] The first, (call it 'the Fregean strategy'), involves taking propositions to be highly fine-grained. Assume, as above, that an agent's belief state is represented by a set S of propositions and that an agent believes a proposition p if and only if $p \in S$. Now suppose that propositions are so fine-grained that the proposition that Hesperus is bright (call it 'p_1') is not identical to the proposition that Phosphorus is bright (call it 'p_2'). It would be perfectly straightforward to explain how an agent can believe that Hesperus is bright while failing to believe that Phosphorus is bright (this happens if p_1, but not p_2, is a member of the agent's total belief state).

The second strategy (call it 'The Third Factor' view), allows for propositions to be more coarse-grained[13], but slightly complicates the way in which an agent's total belief state is construed, as well as the truth-conditions of ordinary belief ascriptions. There are various

---

[12] For defenders of the first strategy see for example Frege (1948), Forbes (1990), and Pietrosky (1996). For defenders of the second, see for example Salmon (1986), Soames (1989), and Richard (1993). The difference between semantic and pragmatic versions of the second strategy need not concern us for current purposes.

[13] Defenders of this strategy typically opt for a Russelian notion of propositions (namely, propositions being as coarse-grained as n-tuples of objects and properties). But note that this construal of propositions is in no way essential to the strategy: there is nothing in principle to bar one from combining this strategy with the view that propositions are sets of possible worlds.

ways to implement the strategy, but – to focus on one – suppose we take an agent's total belief state to be a set of ordered-pairs, each consisting of a proposition and a mode of presentation (roughly: a way that proposition is represented to the agent). We will now assume that for each (token) ascription of the form 'X believes that s' there is some mode of presentation m, such that the ascription is true if and only if the pair <the proposition that s, m> is a member of X's total belief state.[14] With this machinery in place we can allow, for example, that the proposition that Hesperus is bright is identical to the proposition that Phosphorus is bright (call this proposition 'p'), but that the ascription of 'X believes that Hesperus is bright' (on a particular occasion) has different truth-conditions than the ascription 'X believes that Phosphorus is bright' (on a particular occasion). For assume that the former ascription involves the mode of presentation m_1, and the latter the mode m_2. If X's total belief state includes the pair <p,m_1>, but not the pair <p, m_2> we can account for how the agent believes that Hesperus is bright, while failing to believe that Phosphorus is bright. There are of course outstanding issues surrounding the question of what are modes of presentation and how each particular ascription gets to be associated with a particular mode of presentation m, but these issues need not concern us for the moment. The purpose of the above remarks is not to develop one particular account of propositional attitudes, but merely to remind the reader of the salient alternatives to the possible-worlds account.

With this background in place, we can now turn to examine Lewis's arguments.

§1.3 Lingens, the two gods, and book learning

The argument: Consider the following example, due to Perry: "*An amnesiac, Rudolf Lingens, is lost in the Stanford library. He reads a number of things in the library, including*

---

[14] It is important that we focus on token ascriptions: it may well be that the same ascription-sentence uttered in different contexts requires a different mode of presentation for its truth.

*a biography of himself, and a detailed account of the library in which he is lost.... He still won't know who he is, and where he is, no matter how much knowledge he piles up, until that moment when he is ready to say, 'This place is aisle five, floor six, of Main Library, Stanford. I am Rudolf Linge*ns'". (Perry (1977), 492).

Lewis makes the following remarks concerning the case: "*Book learning will help Lingens locate himself in logical space. The more he reads, the more he finds out about the world he lives in, so the fewer worlds are left where he may perhaps be living…But none of this, by itself, can guarantee where in the world he is living*" (ADD, 520).

Similarly, Lewis considers the example of the two Gods, who "*inhabit a certain possible world, and they know exactly which world it is. Therefore they know every proposition that is true at their world. Insofar as knowledge is a propositional attitude, they are omniscient. Still I can imagine them to suffer ignorance: neither one knows which of the two he is*". (ADD, 520).

The lesson Lewis draws from these examples is clear: Lingens or the two Gods have learnt all the (relevant) true propositions. Thus if knowledge or belief are propositional attitudes (as in the standard possible-worlds account) there would be no way of accounting for the fact that even after learning all the true propositions the agents in question still suffer from ignorance. To account for this ignorance, one needs to shift to the centred-worlds account which allows "learning" of all the relevant propositions[15] by updating one's belief state to include the corresponding (boring) centred-propositions, while remaining ignorant of the relevant

---

[15] I write 'learning' in quotes, because if learning involves acquiring knowledge rather than merely belief, then it is not clear how Lewis's framework can account for it. (After all, knowledge is factive, and it is not clear what it takes for a centred-proposition to be true. Cf. §2.1 below).

(interesting) centred-propositions, such as the proposition $\{<w,x>: x$ is Lingens and x is in the Stanford Library in $w\}$).

Response: Lewis raises a certain puzzle: how can we account for the lingering ignorance in the above cases? It is crucial to realise that the puzzle only arises in the context of the possible-worlds account. To see this, note that the puzzle does not arise with either of the two alternative classes of accounts I sketched in §1.2. First, consider the Fregean account. According to the Fregean account it is simply an incorrect description of the case that by reading the book in the Stanford library Lingens has learnt all the (relevant) true propositions: the proposition that is expressed (in the relevant context) using the sentence 'You are in the Stanford library' is not the same proposition as the one expressed by the sentence 'Lingens is the Stanford library'. And since the book presumably contains only the latter but not the former sentence, it does not after all inform Lingens of all the relevant true propositions, and thus there is no surprise that Lingens has some lingering ignorance.[16]

Alternatively, consider the Third Factor account. An initial subtlety that arises is whether on this account one can even talk of an agent being ignorant of or learning a proposition (rather than, e.g., a proposition/mode pair). Plausibly, though, the account should indeed allow for such talk: let us say that an agent stands in a propositional attitude R towards the proposition p, just in case there is some mode of presentation m, where $<p,m>$ is a member of the agent's total R-state. Admittedly, on this way of thinking Lingens has indeed learnt all the (relevant) true propositions. Nevertheless, on the Third Factor account, ordinary ascriptions of knowledge, belief, and ignorance have truth-conditions that are relativised to modes of presentation. Thus on this account, it is completely standard for ignorance to persist even after one learnt all the relevant propositions (under some mode or other): one can come to

---

[16] For a defence of a Fregean solution to Lewis's puzzle along these lines see Stanley (2011), ch. 3.

believe the proposition that Superman (/Clark Kent) can fly under a 'Supermany' mode of presentation, while failing to believe that Clark Kent can fly. The Lingens case thus poses no special puzzle to the account.[17]

A second, related point is this: in so far as these cases do raise a puzzle, the puzzle has nothing in particular to do with the de se. Suppose for example that the Stanford library has a book which contains all the true propositions concerning Hesperus. On the face of it, Lingens could read the whole book and still fail to know whether Phosphorus is a planet. (Note that Lewis also need to account for the lingering ignorance in this case, even though it has nothing in particular to do with de se attitudes, and the move to the centred-worlds account does nothing to help here).[18]

A Lewisian could try to address this particular version of the puzzle by using the standard machinery of the possible-worlds account. For example, one might maintain that if the book contains *all* the true propositions concerning Hesperus, it contains the fact that Hesperus is called 'Phosphorus'. This, however, will not do: after all, the book could report this fact using the sentence 'Phosphorus is called 'Phosphorus'', which will not help alleviate Lingens's ignorance. Nor will it help to insist that the book contains the fact that Hesperus is called both 'Hesperus' and 'Phosphorus': since propositions on the Lewisian view are highly coarse-grained, this fact can be reported using the sentence 'Hesperus is called 'Hesperus' and

---

[17] Similar remarks apply in the case of the two gods. Suppose one of the two gods does not know that he is the god on the tallest mountain. On the Fregean account, the god - contrary to the initial description of the case - is not omniscient (there is some true proposition he fails to know – namely one that she would express by saying 'I am the god on the tallest mountain'). Whether the god is omniscient on the Third Factor account depends on how precisely we define omniscience (does it require that the schema 'x knows that *s*' is true for any true sentence *s*, or does it merely require that for each proposition *p*, x knows the proposition *p* in the sense defined above?). On the first way of understanding 'omniscience', the case is again misdescribed and the gods are not after all omniscient; on the second way, the case as described is possible but there is no puzzle about how it arises.

[18] As above, a Fregean will maintain that if the book did not contain the sentence 'Phosphorus is a planet' then it did not after all contain all the facts about Hesperus, and a proponent of the Third Factor view will address the puzzle by appealing to modes of presentation.

Phosphorus is called 'Phosphorus''. Alternative proposals which involve more general claims face similar challenges. Consider for example the proposition that there is a single planet called both 'Hesperus' and 'Phosphorus'. A quick reflection shows that this proposition is necessarily equivalent to the claim that either 'Hesperus' is called Hesperus and Phosphorus is called 'Phosphorus' or there is a single planet called both 'Hesperus' and 'Phosphorus'. And since on the possible-worlds account any two necessarily equivalent propositions are identical, one can report the relevant content using the latter sentence – one which intuitively will fail to relieve Lingens's ignorance (even after reading the latter sentence, it is an open possibility for Lingens that the disjunction is true in virtue of the first disjunct).[19] The general point should be clear: if all one requires is that the book report the relevant coarse-grained propositions (irrespective of *how* they are reported), then it is seems that there could be a book which reports these propositions under a sufficiently unhelpful guise, one that will not relieve the relevant agents of their ignorance. (Or to put the same point differently: defenders of the possible-worlds need to be careful that their solution doesn't have intuitive appeal just because they have presented the relevant propositions under a favourable guise – after all according to such theorists, the guise under which the proposition is presented should play no role in accounting for knowledge or ignorance.) It is thus far from clear that the possible-worlds account has some simple solution to the Hesperus/Phosphorus puzzle, one that is not available for I/Lingens case.[20]

As a final note, it is worth pointing out that the 'book learning' metaphor which is often employed in this context to contrast standard propositional knowledge with so called self-locating or de se knowledge is not a particularly helpful one. After all, in the right

---

[19] Also, the proposal of using descriptions such as 'called 'Hesperus'' and 'called 'Phosphorus'' to relieve Lingens's ignorance is likely to run into trouble when we consider other variants of the case, e.g. ones where Lingens possesses two planet names 'Phosphorus' (cf. the 'Paderewski' case), and knows of one but not that other that it denotes Hesperus.

[20] More on this issue in §3 below.

circumstances, a book can provide one with what is standardly classified as de se knowledge (consider for example a book permanently chained to the Stanford library, which contains the sentence 'You are in the Stanford library': presumably, reading such book gives you de se knowledge that you are in the Stanford library). Conversely, there are forms of knowledge that are at least not paradigmatic cases of self-locating knowledge, but such that it would nevertheless be hard to learn from (ordinary) books (consider for example knowledge of what an A-flat sounds like or knowledge of how to ride a bike).

§1.4 believing the impossible

***The argument*:** Another case Lewis considers is that of Heimson, a madman who falsely believes that he is Hume. The problem, according to Lewis, is that Heimson "*couldn't be Hume. If he believes the proposition that holds at just those worlds where he is Hume, then he believes the empty proposition that holds at no worlds*" (ADD, 524). This, Lewis maintains, gives rise to the following puzzle: "*the proposition that Heimson is Hume…is the empty proposition, hence unfit to be believed…Yet Heimson does believe that he is Hume. How can that be?*" (ADD, 524-525).[21]

Here, as above, Lewis's proposed solution is to shift to the centred-worlds account, and to construe Heimson's belief as a belief in the centred proposition {<*w,x*>: *x* is Hume in *w*}, which presumably is fit to be believed.

Response: This is another case where the problem Lewis raises is just a product of the 'original sin' – namely the initial appeal to the possible-worlds account. As we have seen, the account suffers from the problem of necessary falsehoods. But contrary to what the possible-

---

[21] This is a version of the puzzle that assumes there can be strict cross-world identity. Lewis argues, however, that the puzzle can also be generated in the counterpart-theoretic framework (see ADD, 524).

worlds account predicts, it is perfectly possible for agents to believe necessary falsehoods (false mathematical claims, negations of non-trivial metaphysical necessities, and so forth), and alternative accounts of propositional attitudes have no problem accommodating this fact. Moreover, as above, this problem has nothing in particular to do with the de se: instead of believing that he is Hume, Heimson might believe that Aristotle is Hume, or that Clark Kent is Batman – and these too would be beliefs in impossible propositions which the possible-worlds framework has a hard time accounting for. But the shift to the centred-worlds account does nothing to help with these other versions of the problem, as presumably no agent can practically believe the centred-proposition $\{<w,x>$: Aristotle is Hume in $w\}$, given that this would require their belief state to be empty and hence for the agent to believe every centred-proposition.[22]

§1.5 The role of the de se in practical deliberation and explanation of action

The argument: Another argument which Lewis raises concerns the role of de se attitudes in practical deliberation.[23] Suppose that Lingens learns that there are actually two amnesiacs who are sitting in two very similar looking libraries – the Stanford amnesiac in the Stanford library, and the Harvard amnesiac in the Harvard library. In Lewis's terminology, Lingens knows all the propositional facts concerning each of the two amnesiacs ('he knows what the world is like'), but he doesn't know which of the two amnesiacs *he* is ('he doesn't know how to locate himself in the world').

---

[22] Working within a Fregean framework, Stanley (2011), pp. 96-97 discusses the related problem of explaining how I can imagine that I am Napoleon, and notes "This is a difficult case for any Fregean view of the content of PRO". Stanley attempts to resolve this problem by arguing that "the verb "imagine" can take a character of a sentence as its object" . I don't think such a solution is needed (assuming the Fregean view, it is not clear there is any puzzle about how agents imagine such impossible situations), and moreover, Stanley's solution has the disadvantage that it does not generalise to other cases of imagining the impossible, namely ones which involve concepts with a constant character (e.g. imagining that John is Napoleon).

[23] See especially §XI of ADD.

We have already discussed the issue of accounting for such ignorance. But Lewis emphasises the importance of the issue by noting that the question of what one believes de se is crucial for practical and rational deliberation. Suppose for example that Lingens learns that there is a fire in the Harvard library (but not in the Stanford library). He might reason as follows: 'If I am the Harvard amnesiac, I should jump out the window (otherwise, I will be caught in the fire). On the other hand, if I am the Stanford amnesiac, I should not jump out of the window (it is dangerous to jump out of windows and a bad idea to do so if the building I am in is not on fire)'. Thus the issue of where Lingens locates himself in the world is crucial to his practical deliberation and to how he chooses to act.

Response: It is true that as this specific case is laid out, the question whether *he himself* is the Stanford amnesiac (rather than, say, whether *Lingens* is the Stanford amnesiac) plays a central role in Lingens's process of practical deliberation. But yet again, this point has nothing in particular to do with the de se. Consider an alternative case: I am not an amnesiac. I know that I am in the Harvard library. I then learn that the Widener library is on fire (perhaps because it is announced on the radio, to which I am listening while working in the library). But I am not sure whether the Widener library is the Harvard library or the Stanford library. I might reason as follows: 'If the Harvard library is the Widener library, I should jump out of the window. On the other hand, if the Widener library is the Stanford library, I should not jump out of the window'. Thus in this second case, the crucial question to my process of practical deliberation is whether the Harvard library is the Widener library. But this question does not in any way concern any de se attitudes.

The upshot should be clear: like all other forms of thought, practical deliberation is importantly dependant on modes of presentation (or, if you prefer the Fregean way of

thinking, on a fine-grained notion of propositions). This is true whether or not one of the mode of presentations involved is a 'self-locating' or 'first personal' one.

A defender of the myth might at this stage propose a slightly different version of the argument. According to the revised argument, what is special about de se attitudes is that, as opposed to other kinds of attitudes, they are an essential component of any practical deliberation or intentional action. Perry, for example, expresses a thought along these lines when he discusses the case of the messy shopper. As the case goes, Perry notices that someone in the supermarket is making a mess, but only after realising that it is he himself making the mess, he stops and looks at his own cart. Highlighting the essential role of the 'indexical' belief in explaining action, Perry comments as follows: "*Suppose that I had said 'I came to believe that John Perry is making a mess'. I would no longer have explained why I stopped and looked at my own cart. To explain that, I would have to add, 'and I believe that I am John Perry', bringing in the indexical once again*". (Perry (1979): 4-5).

There are several things to say about this line of thought. First, let us grant for a moment that one cannot account for any intentional action without appealing to attitudes that that are, in some sense, self-locating or first-personal. This in itself would be insufficient to support the central tenet of the myth of the de se – in particular the Special Challenge claim. One could plausibly argue that any account of ordinary propositional attitudes would need to involve modes of presentation (or similar devices), due to Frege's puzzle. One could then concede that there is one (or one kind of) mode that is first-personal, and that this mode is particularly important for attitudes that play a role in intentional action. However, this in itself does not require any revision of our standard account of attitudes or attitude ascriptions. (Thus at best, this line of thought supports Special Attitudes, but not Special Challenge – though I will immediately go on to question this as well).

Second, even if it is true that some first-personal attitudes form an essential component of intentional actions, it is far from clear that this aspect distinguishes them from de dicto attitudes. Consider again the above case of the messy shopper. In order to fully explain Perry's action (stopping to look at his cart) we need to appeal not only to the claim that Perry believed that he himself was making a mess (the de se belief), but also to the fact that he believed, say, that making a mess is inappropriate (a de dicto belief). Thus in explaining Perry's action, an appeal to both de dicto and de se beliefs seems necessary and there is no asymmetry between the two sorts of belief.[24]

Finally, it is not even clear that first-personal beliefs are in fact essential to account for one's intentional actions - at least not in order to account for *any* action. The issue is quite complex, and it partially depends on what precisely we expect of an account or explanation of action in different contexts. But to see that first-personal beliefs are at least not obviously always essential consider the following case. Suppose that I believe that people in Africa are starving, that when money is donated to Oxfam this prevents people from starving, and that preventing people from starving is a good thing. Suppose that as a result of holding these beliefs, I go ahead and donate money to Oxfam. This seems like a clear case of an intentional

---

[24] One could try to argue that the action is explained by Perry's belief that he himself is making a mess and making a mess is inappropriate, and maintain that this belief counts as de se. But it would be a very cheap victory for the defenders of the myth if de se beliefs are taken to play a distinctive role in the explanation of action simply because conjunctions of de se and de dicto beliefs are automatically classified as de se. Another option is to explain Perry's action by appealing to his belief that he himself was doing something inappropriate. Certainly in some contexts this explanation will be sufficient, but in others it will seem too thin (indeed – this thin explanation is not the one that Perry provides in this case). Similarly, one might try to argue that although Perry's action in this case is not fully explained by a purely de se *belief*, it can be explained using another kind purely de se attitude - e.g. the *desire* to stop and look at his cart. But again, this explanation seems too thin in this particular case (after all, Perry had no antecedent desire to stop and look at his cart: he only formed this desire in so far as he realised that he was making a mess, and one ought not to make a mess).
To be clear, I am certainly not denying that in *some* cases actions can be explained by appeal to only de se attitudes (or, for that matter, that in some other cases they can be explained by appeal to only de dicto attitudes – see the discussion below). The point is merely that in the specific case as Perry describes it, an appeal to both kinds of attitudes seems important to the explanation, and there is no asymmetry between the two kinds of attitudes.

action. Note that at least as I have described the case, no first-personal beliefs were involved. One could, of course, insist that to *fully* explain my action, one must also appeal to other beliefs, ones which are first-personal. Suppose that in order to donate money to Oxfam I need to reach into my pocket ant pull out some money. One might insist that in order to explain my action, one must also appeal to my believing that in order to donate money *I* need to reach into *my* pocket. However, without further argument it is not obvious that an appeal to such further beliefs is indeed essential. After all, suppose in order to lift my hand, I must first shift my shoulder blade. It does not seem that I cannot intentionally lift my hand without also having the belief that in order to lift my hand I need to shift my shoulder blade: even intentional actions are ultimately achieved via other actions, ones that are more basic or direct, rather than being the result of some process of thought.[25] In a similar manner, reaching into my pocket in order to donate money might be such a more basic or direct action.

To recap: while I do not take these remarks to be conclusive, they do suggest that even if explaining one's actions often involves appeal to first-personal attitudes, such attitudes might not always constitute an essential component of such explanations. It is thus far from clear that so called de se attitudes play the unique or distinctive role with respect to action that defenders of the myth commonly attribute to them. More importantly, even if it turned out that de se attitudes do play such a role that would not be sufficient to justify the Special Challenge claim.

§1.6 Shared beliefs and internal duplication

The argument: Lewis's final line of thought concerns comparing the belief states of different agents. Recall that case of Heimson, the madman who believes that he is Hume. Now

---

[25] Similar remarks hold for other attitudes: it is not clear that in order to lift my hand I need to intend to or desire to shift my shoulder blade.

compare Heimson's belief state to that of Hume. According to Lewis, *"[T]here had better…be a central and important sense in which Heimson and Hume believe alike. For one thing, the predicate 'believes he is Hume' applies alike to both: Heimson believes he is Hume and Hume believes he is Hume*" (ADD, 525). The worry, then, is that the standard propositional model might not account for the sense in which Heimson and Hume believe the same thing: after all, on the standard propositional model, Heimson's belief is an attitude towards the proposition that *Heimson* is Hume, while Hume's belief is an attitude towards the proposition that *Hume* is Hume – and these are not the same proposition. The centred-worlds account, on the other hand, allows Lewis to account for the shared attitude: Heimson and Hume's centred belief states both include the centred-proposition {<w,x>: x is Hume in w}.

Another related argument which Lewis raises, concerns the comparison between two agents who are internal duplicates. Suppose that Heimson not only believes that he is Hume but is a full internal duplicate of Hume. If one's total belief state is construed on the standard propositional model this would entail that Hume and Heimson are not duplicates as far as their beliefs are concerned. But if internal duplicates can differ in their total belief states, this entails what Lewis takes to be an unacceptable conclusion: namely that "*beliefs ain't in the head!*" (ADD, 525). The upshot is that in order to preserve the intuition that internal duplicates share their belief states we must revert to modelling the agents total belief states using the centred-worlds framework.

Response: Lewis's arguments concerning shared beliefs differ somewhat from the previous arguments: while I think these arguments are unpersuasive, in this case the problems with the arguments do not arise specially because of Lewis's appeal to the possible-worlds framework.

Consider the first version of the argument: Hume and Heimson both believe that they are Hume, and the predicate 'believes that he is Hume' is true of both - we can accept that much. However, Lewis fails to take into account is that there are a range of strategies for accommodating this observation without abandoning standard picture of belief as a propositional attitude, or positing any special account of the de se. Probably the most obvious strategy is to appeal to the mechanism of lambda-abstraction. Heimson believes that he, Heimson, is Hume. Hume believes that he, Hume, is Hume. They each believe a different proposition. But each of them satisfies the predicate 'λx.x believes that x is Hume' (just as Lewis requires), and it is in this sense that they count as both believing that they are Hume.[26]

The centred-worlds account is not only inessential in order accommodate this 'shared-belief' intuition, but also it is not a particularly successful strategy for comparing the beliefs of Heimson and Hume. After all, Lewis himself acknowledges that in addition to the intuition that (in a sense) Hume and Heimson believe alike, one is also tempted to say that (in a sense) Hume and Heimson *differ* in their beliefs.[27] (This second intuition becomes particularly compelling when we look at agreement data: it seems wrong to say that Heimson and Hume agree that they are Hume).[28] The alternative standard accounts have the resources to accommodate both these intuitions. For example, on the lambda-abstraction proposal noted above, the predicate 'λx.x believes that x is Hume' is true of both Heimson and Hume, but the predicate ''λx.x believes that Heimson is Hume' is true only of Heimson. It is much less

---

[26] Note that this is not the only strategy available to a defender of the propositional model. Another strategy is to allow for ordinary relations to hold between types as well as tokens. Thus for example, we might both be wearing the same shirt in virtue of the fact that we are both wearing the same shirt-type, and similarly two agents might count as believing the same thing, by both believing the same type of proposition. For further discussion on various strategies for accommodating such data, see Cappelen & Hawthorne (2009).

[27] "Doubtless it is true in *some sense* that Heimson does not believe what Hume did." (ADD, p. 525)

[28] For the importance of agreement data as a test for shared content, see Cappelen & Hawthorne (2009), ch. 2.

clear, on the other hand, that Lewis's centred-worlds account is able to accommodate this second intuition.[29]

These points aside, there is a more general way to see that Lewis's argument on the wrong track here: the phenomenon Lewis is pointing to has nothing in particular to do with attitude ascriptions. Suppose for example that Jill rides her bike to school, and Jane does too. The following paraphrase of Lewis's quote seems just as right: 'There had better be a central and important sense in which Jill and Jane's travel arrangements are alike. For one thing, the predicate 'rides her bike to school' applies alike to both: Jill rides her bike to school and Jane rides her bike to school'. It is clear that in whatever way we choose to account for this observation, our account will have nothing to do with propositional attitudes in general, and with de se attitudes in particular.

Let us turn then to Lewis's second argument, concerning full internal duplicates. An obvious response is to endorse the externalist conclusion that internal duplicates may differ in their total belief states, i.e. simply accept that 'beliefs ain't in the head'. After all, even putting aside the de se, there are strong reasons to adopt this position: when Oscar and Twin-Oscar have beliefs that they each expresses by saying 'There is water in the lake' the contents of their beliefs differ (Oscar's belief is about H2O and Twin-Oscar's belief is about XYZ). More generally, on plausible assumptions, externalism about meaning entails externalism about belief.[30]

---

[29] A Lewisian could try to propose that Heimson believes the (boring) centred-proposition that Heimson is Hume, while Hume believes the (boring) centred-proposition that Hume is Hume. But given Lewis's other commitments he is bound to maintain that neither agent believes the first (necessarily false) proposition, and that both believe the latter (necessarily true) proposition.

[30] For a general defence of semantic externalism see Putnam 1975 and Burge 1979. In a brief remark on the topic Lewis seems happy to endorse semantic externalism but not its ramification to mental states ("The New Theory of Reference teaches that meanings ain't in the head…That may be right…[b]ut the proper moral isn't that beliefs ain't in the head. The proper moral is that beliefs are ill-characterized by the meanings of the

Interestingly, Lewis concedes that one might wish to have a notion of 'belief' and other attitudes that may be externally individuated, but maintains that for certain central purposes, it is crucial to appeal to another, internalist conception of attitudes.[31] As he puts it: *"...attributions of beliefs enter into systematic common-sense psychology, and...for this purpose beliefs had better be in the head"* (ADD, 536).[32]

This is not the place to settle the internalism versus externalism debate. But it is worth briefly recalling that there are some apparently compelling reasons to reject Lewis's contention that, at least as far as common-sense psychological attitudes are concerned, those attitudes must be internalist ones. First, assuming that ordinary propositional attitudes (i.e. those attitudes which are referred to in ordinary ascriptions such as 'X believes that p') are not internal, it is not clear that one can separate away the external factors from such ordinary attitudes, achieving some kind of underlying narrow or internal mental states of the sort that Lewis seeks.[33] Second, even if there are such narrow attitudes in the vicinity, it is far from clear that it is these attitudes, rather than the externalist ones, that play a role in 'systematic common-sense psychology'. No doubt *knowledge* is an attitude that plays an important role in common-sense psychology, but internal duplicates can surely differ in what they know.[34] Or consider the following: a natural common-sense psychological explanation for why Oscar ran away is that he believed there was a tiger in the room and he feared that tigers are dangerous. But these attitudes are ones that Oscar may not share with his internal duplicates: an internal duplicate who lives in a planet where there is another, tiger-resembling, natural kind (call

---

sentences that express them", ADD, 526). But see Williamson (2000) (especially ch.2) for arguments that semantic externalism entails externalism about beliefs and other mental states.

[31] ADD, 526.

[32] Lewis is here attributing this view to Perry, but it is clear from the context that this is a view he endorses as well.

[33] See Williamson's argument for primeness in Williamson (2000), ch.3.

[34] On the importance of knowledge for common sense causal and psychological explanation, see Williamson (2000), §2.4.

them 'schtigers') might instead believe that there is a schtiger in the room, and fear that schtigers are dangerous. Of course, an internalist could insist that it is not this belief that explains Oscar's behaviour, but rather another belief - one that Oscar does shares with his duplicates (e.g. the belief that there is a striped yellow animal in the room). But it seems a real stretch to claim that it is the latter sort of explanations that typically enter into common-sense psychological explanations rather than the former. (After all, typical every-day explanations would cite the tiger-beliefs, rather than the 'striped yellow animal' ones).

Needless to say, these brief remarks are not intended to convince the committed internalist. But they should serve to remind us that externalism is not some piece of higher-order philosophical sophistry: it is a position which attempts to correctly characterise and explain agents' attitudes, psychological states, and behaviour. It is thus far from clear that Lewis's internalism has the upper hand when such matters are concerned.

Finally, it is worth mentioning another argument which defenders of the myth sometimes raise in the context of comparing beliefs across agents.[35] First, the suggestion goes, shared *centred*-beliefs entail sameness of action: if Jack believes that *he* is in danger, and Jill believes that *she* is in danger then (all else being equal), they will both flee the scene. The centred-worlds account can presumably explain this sameness of behaviour, because Jack and Jill share a centred-belief state. Secondly, shared *propositional* (non-centred) beliefs do not lead to sameness of action. Suppose, for example that Jack and Jill both believe precisely the same propositions. In particular, Jack and Jill both believe that Jack is in danger but that Jill is not in danger. Presumably, even though they share the same propositional beliefs, they will act differently (Jack will flee the scene, and Jill will stay put). Taken together, these points

---

[35] See Perry (1977), p. 494 (though Perry obviously doesn't phrase this line thought using the centred-worlds framework, as I do above).

purport to show that it is *centred*-belief states, rather than ordinary propositional beliefs that are important for the explanation of action, in particular across agents.

We should not endorse this line of thought. With respect to the first point, note for a start that just as there is a sense in which Jack and Jill act the same way (they both flee the scene), there is also a sense in which the act in different ways (for example, Jack lifts Jack's legs to run, while Jill lifts Jill's legs to run). And that aside, one does not need to assume that both agents have precisely the same beliefs in order to explain shared actions: the fact that I believe the party is going to be good, and you believe that the party is going to be amazing can explain why we are both going to the party. It is sufficient in this case that our beliefs share some crucial property (thinking of the party in a positive way) to explain our shared action.[36] With respect to the second point, at least two considerations are worth noting. First, although beliefs certainly play an important role in explaining action, the connection between belief and action must also be mediated by some more hard-wired, physiological constraints. For example, suppose John's arms are paralyzed and Jane's are not. Even if John and Jane share all their beliefs (or if you want, all their centred-beliefs), and Jane's beliefs lead her to lift her arm, those same beliefs will not lead John to lift his arm – simply because he is unable to. Similarly, while Jack's beliefs can lead Jack to lift his legs and flee the scene, Jill's beliefs cannot lead her to lift Jack's legs (at least not in the same manner), simply because she is physiologically unable to do so. Finally, note that this argument again implicitly assumes a very coarse-grained account of attitudes. Alternative accounts, ones which allow for belief states to be dependent on modes of presentation, can maintain that although Jack and Jill

---

[36] One might try to argue that what explains our shared action in this case is some third, shared belief (e.g. we both believe that the party will be at least good). But first, even if we in fact have such a shared belief, it is far from obvious that the explanation *must* appeal to it – on the face of it, it seems that a shared property of different beliefs provides a sufficient explanation. And second, it is not clear that in all such cases we do in fact have a shared belief (e.g. consider the case where I believe the party will be intellectually stimulating and you believe it will be fun and neither of us makes the inference that the party will be good).

believe the same *propositions*, they are in different total belief *states*, because Jack's belief that Jack is in danger is presented to him in a different mode than Jill's belief that Jack is in danger. And since Jack and Jill are in different belief states, there is no problem in accounting for why they act differently.

§2. Trouble in centred-pluralverse

In the previous section I argued that Lewis's move to a centred-worlds account was unmotivated. In this section, I point out several problems that the shift to the centred-worlds framework brings with it. While these problems may not be entirely insurmountable, they serve to show that the centred-worlds view is much less attractive than it is often to taken to be.

§2.1 Assertion, communication, and truth

Consider Jill, an agent who was born in 1972, and believes that she was born in 1972. Suppose that Jill utters the following sentences:

(1) I believe I was born in 1972.

(2) What I believe is true.

(3) When I utter the sentence 'I was born in 1972', I assert what I believe.

(4) When I utter the above sentence, my audience believe what I asserted.

Given the right circumstances, all claims seem perfectly true, and the original possible-worlds account has no problem accommodating these mundane facts.

However, once one shifts to the centred-worlds account, things become much trickier. Claim (1) presumably attributes to Jill a belief in the centred proposition $\{<w,x>: x$ is born in 1972 in $w\}$. But now consider (2): while we have a perfectly straightforward understanding of what it is for an ordinary proposition to be true (on the possible worlds account: it's true just in case the actual world is a member of it), it is far from clear what it takes for a *centred*-proposition to be true. (For example, we cannot maintain it is true just in case for *some* individual x, $<@,x>$ is a member of the centred-proposition, or else if Jill was born in 1973, but believes that she was born in 1972, her belief would count as true).

Claim (3) can be accommodated if we assume that the objects of assertion are not propositions, but rather *centred*-propositions. However, this causes trouble with respect to (4): if by uttering 'I was born in 1972', Jill asserts the centred-proposition $\{<w,x>: x$ is born in 1972 in $w\}$, it is simply not true that when she utters the sentence, her audience comes to believe what she asserted. For that would amount to each of Jill's audience members, coming to believe that he or she was born in 1972.[37]

---

[37] The problem of accommodating (3) and (4) in the centred-worlds framework was probably first raised in Stalnaker (1981), 9. 146-7. It is worth noting that this may be a problem not only for the centred-worlds account, but also for some the alternative accounts I considered in §1.2. For example, taking the Fregean view, one might worry that in order to accommodate (3), we must assume that Jill is asserting a very fine-grained proposition (roughly, one that refers to herself in a first-personal mode of presentation), and that it cannot be this fine-grained proposition that her audience comes to believe, and hence (4) cannot be correct.

There are various ways to respond to this worry. First, one could try to argue that her audience can after all come to believe precisely this very fine-grained proposition (see Bermudez (2005) for a defence of this proposal). Second, on either the Third Factor or Fregean views, one could argue that there are different types of modes of presentation (/Fregean propositions): ones that are very 'thin' or 'minimal' as well as ones that are very rich. (This view is motivated independently by the need to accommodate shared belief reports). Thus in this case Jill might believe the same proposition in two different ways: one involving the very rich, first-personal mode of presentation, and one involving a more minimal or general one. It follows that (3) can be made true by Jill's asserting the more minimal proposition-mode pair (/Fregean proposition), which makes accepting (4) straightforward. Finally, the Third Factor view, also has the simpler option of claiming that the objects of assertion are simply propositions (rather than proposition-mode pairs), and hence (3) and (4) are made true by the fact that Jill believes, asserts, and her audience believes precisely the same proposition.

Several strategies for dealing with these problems have been proposed in the literature by defenders of the centred-worlds framework.[38] While, I do not have the space to discuss them here in detail, the crucial point is that the shift to the centred-worlds account comes at a hefty price: some very straight-forward ways in which the possible-worlds account analysed the connection between belief, assertion, communication, and truth become highly complicated.

§2.2 The de-dicto/de se distinction

The centred-world framework displays a certain tension. On the one hand, the move to the framework was motivated by the apparent observation that there is something highly distinctive about de se attitudes. On the other hand, as we have noted above, Lewis does not wish to have a bifurcated account of attitudes, and once the centred-worlds framework is introduced, it is used to account for both de se and de dicto attitudes. Doesn't this common framework threaten undermine the distinction between the de dicto and the de se?

Lewisians have a common response to this worry:[39] Call a centred-proposition p 'boring' just in case, for all worlds w, and individuals x, if $<w,x> \in p$, then for all individuals y, $<w, y> \in p$. Call a centred-proposition 'interesting' just in case it is not boring. The proposal is that distinctively de se attitudes are attitudes towards *interesting* centred-propositions, while ordinary de dicto attitudes are attitudes towards *boring* centred-propositions. For example, the proposition $\{<w,x>: x$ is tall in $w\}$ is an interesting centred-proposition (if Jill is tall but

---

[38] On the issue of trying to accommodate (2) see the sizable literature on relative truth (e.g. McFarlane (2005) as well as the extensive exposition and criticisms of relative truth in Cappelen and Hawthorne (2009)).
For attempts to address the problems concerning the connections between belief and assertion within the centred-worlds framework see Egan (2007), Moss (*forthcoming*), Ninan (2010), and Torre (2010). Without going into a detailed critique of these attempts, it is worth noting at least that they each come at the price of giving up some principles that may be thought of as central to the Lewisian framework and motivations. For example, Moss and Egan maintain that (except in some rather special cases) assertions do not express de se beliefs but rather de dicto ones - that is, even assertions of sentences such as 'I am Rudolph Lingens', 'It is now 5pm' or 'I expected to be saved', while Torre argues that the content of what is asserted (a 'multi-centred proposition' on his view), is not even the kind of object that can be believed.
[39] See Egan (2006).

Jack is not in w_0, then $\langle w_0, \text{Jill} \rangle$ will be a member of the proposition but $\langle w_0, \text{Jack} \rangle$ will not). Thus John's believing this centred proposition (i.e. believing that *he* is tall), is a de se attitude. On the other hand the proposition $\{\langle w,x \rangle: \text{someone is tall in } w\}$ is a boring centred-proposition. Thus John's believing this proposition (i.e. believing that someone is tall) is a de dicto attitude.

The problem is that this standard response does not work. It does not work because one can have attitudes that defenders of the myth would typically classify as de se towards properties that in each world either every individual possesses or no individual possesses. Suppose for example, that in each possible-world either there is a heaven, in which case everyone is such that if they are good they go to heaven or there is no heaven, in which case no one is such that if they are good they go to heaven.[40] Now suppose that I hope that I will go to heaven if I am good. This is the sort of attitude that would typically be classified as de se: the content of my hope concerns what might happen to *me*, and I can have this attitude even if I have suffered severe amnesia and do not remember who I am. Yet, this is an attitude towards the centred-proposition $\{\langle w,x \rangle: \text{in } w, \text{ if } x \text{ is good } x \text{ goes to heaven}\}$, which is a *boring* centred-proposition (for all worlds w and individuals x, y, if $\langle w,x \rangle$ is a member of the proposition, then so is $\langle w,y \rangle$). So by the above proposal it would be classified as a de dicto rather than a de se attitude. Defenders of the Lewisian line must thus find some other way to distinguish the two sorts of attitudes.[41]

---

[40] I am assuming here that the conditional is interpreted in a way that allows for this possibility (i.e. that the conditional is not interpreted so that if x is not good, the conditional is trivially true).

[41] This also raises additional related problems, which are due to the coarse-grainedness of the centred-worlds account: the account faces the unpalatable result that believing that I will go to heaven if I am good; believing that Jack will go to heaven if he is good; believing that everyone will go to heaven if they are good; or believing that someone will go to heaven if they are good – all amount to having precisely the same attitude.

§2.3 Ignorance of the purely de se

As we have seen, one of the Lewis's key motivations consisted of accounting for ignorance of the de se – for example, for the fact that Lingens can wonder whether he himself is Lingens. But what about ignorance of the *purely* de se?

Just as there can be scenarios where I might wonder whether I am Lingens, there could also be scenarios where I might wonder whether I am I, although it takes a bit more imagination to think of such scenarios. Consider for example the case of Jake. Jake suffers from some schizophrenia-like psychological disorder. Suppose that as part of that disorder, Jake's stream(s) of consciousness comes in two 'voices' (even if one is a proponent of a pure psychological criterion of personal identity, one can imagine that the thoughts expressed by the two voices are psychologically similar and connected enough that they would both count as belonging to the same person). Now suppose that Jake is occasionally confused about the source of the voices in his head: for example, thinking in voice-number-1 he might wonder whether voice-number-2 is also his, or rather involves some external agent talking to him (God, aliens, or what not…). Conversely, thinking in voice-number-2, he might have similar wonderings about the voice-number-1 thoughts. Jake could also have mixed-thoughts that switch between the voices. He might therefore ask himself: 'Am I [thought in voice-1] identical to I [thought in voice-2]?', and he may be ignorant of the fact that the correct answer is the positive one.

It is seems that according to the centred-worlds framework, Jake's wondering should be construed as an attitude towards the centred propositions $\{<w,x>: \text{in } w, x=x\}$. After all, if on Monday Jake has only thoughts in voice-number-1 we would naturally construe his first-personal thoughts on Monday in the usual manner: his thought that he (in voice-number-1) is happy is an attitude towards the centred proposition $\{<w,x>: x \text{ is happy in } w\}$. Similarly, if

on Tuesday, Jake has only thoughts using voice-number-2, we should construe his first-personal thoughts on Tuesday in the usual manner as well (his thought that he, in voice-number-2, is sad is an attitude towards the centred proposition $\{<w,x>: x$ is happy in $w\}$) . It is thus only natural that when on Wednesday, he has mixed thought in both voices, these should be construed so that either of the voices can be placed at the centre, and in particular his 'I (voice-1) am I (voice-2)' thought be construed as an attitude towards the centred-proposition $\{<w,x>: x=x$ in $w\}$. The problem, however, is that on the centred-worlds framework every agent trivially believes this centred-proposition (all centred attitude states are included in it). So while the framework can account for some forms of de se ignorance, it cannot account for cases such as these where ignorance of the purely de se is involved.[42]

§3 The reducibility argument

In this section, I present what I take to be the strongest argument in favour of the myth of the de se in general, and of the centred-worlds account in particular (I call it 'The reducibility argument'). I then go on to argue that despite its initial appeal the reducibility argument is not ultimately convincing.[43]

---

[42] Another case which can be used to demonstrate the same point (suggested to me by X) is the following. Suppose that Jane is a metaphysician who believes first, that minds are distinct from bodies; and second, that when ordinary speakers use the pronoun 'I', their use is ambiguous between two pronouns 'I_M' – which denotes a mind, and 'I_B' which denotes a body. (Let us suppose that both of Jane's beliefs are false: minds are identical to bodies, and in fact 'I' unambiguously refers to the agent in question). Jane might think to herself: 'I am not I', because she (falsely) construes this claim as 'I_M is not I_B'. (This case may have the advantage that it does not involve a psychologically abnormal agent, though more will need to be said about how to construe the beliefs of agents with false views about their own semantics).

For additional and much more vivid description of cases where one might be confused about whether 'I am I' or 'I am here' are true, see Dennett (1981).

[43] The argument as I phrase it has not been explicitly presented by defenders of the myth, but it echoes various lines of thought brought up in this context, and I suspect that something like this argument motivates many who endorse the myth. For a defence of related ideas see Lewis (1979), §VIII, Stalnaker (1978), Stalnaker (1981), Stalnaker (2008), and Chalmers (2006). Similar ideas are also discussed (though not defended) in Ninan (MS).

In §1, I argued that most of the concerns Lewis raises for handling de se attitudes within the possible-worlds framework, do not arise – as Lewis suggests - because of some distinctive features of de se attitudes, but are rather by-products of the general problems that plague that framework.

The reducibility argument aims to respond to this challenge by adopting the following line. First, while the simple version of the possible-worlds account indeed faces serious problems, these problems can be addressed by adopting *descriptivism about attitudes* (more about what this means below). Second, this solution is not quite adequate if implemented within the ordinary possible-worlds framework, but is successful if it is instead implemented within the centred-worlds framework (this will amount to adopting what I call '*centred-descriptivism'*). Putting the two points together, the thought is this: versions of Frege's (and other related) puzzles that pertain to the de se are not merely ordinary instances of the general puzzle. Rather, general versions of the puzzle can be resolved by reducing them to instances of the puzzle involving the de se, while the latter instances can be resolved by appeal to the centred-worlds framework. Thus, if successful, the reducibility argument would address both the charge that the possible-worlds account is not adequate in general, as well as the charge that no special account is needed to handle de se attitudes.

Let me turn to the details of the proposal. Leaving aside the de se for a moment, consider the standard possible-worlds account of attitudes (for convenience, I focus on belief). As I have sketched it, the account posits a very simple link between what is required for the truth of ordinary belief ascriptions ('X believes that p') and the relevant agent's total belief state. Descriptivism about de re belief maintains that this straightforward link can be preserved in the case of ascriptions where 'p' is a *purely de dicto* statement (roughly, a claim not

containing any singular terms), but needs to be modified when 'p' is a *de re* statement (roughly, a claim containing some singular terms). Put together we get the following view:

Descriptivism about belief

(1) If 'p' is a purely de dicto statement: X believes that p if and only if X's total belief state is included in the proposition that p.

(2) If 'p' is a de re statement (for simplicity we will focus only on the case where 'p' has the form '*a* is F'): X believes that p, if and only if there is some adequate, purely de dicto description 'G'[44] for which the following hold:

(a) *a* is the unique G.

(b) X believes that the unique G is F.

Two essential points to note are, first, that clause (2b) uses a belief ascription that involves a purely de dicto statement ('The unique G is F'),[45] and thus is covered by clause (1). Second, the description in question has to be 'adequate'. Now a defender of the reducibility argument would of course need to supplement the account with some further theory of what it takes for a description to count as adequate. For current purposes, an important constraint is such a theory will be sufficient to address the various versions of Frege's puzzle. Suppose for example that, at least in the relevant context, the description 'planet that shines in the evening' is adequate for the purpose of the ascription 'Jill believes that Hesperus is bright' (and no other description is adequate), and that, the description 'planet that shines in the morning' is adequate for the ascription 'Jill believes that Phosphorus is not bright' (and again, no other description is adequate). We can then appeal to the descriptivist picture to

---

[44]Again, I am not providing here an analysis of what 'purely de dicto description' means. But to illustrate the notion by some examples: 'is yellow and square-shaped' or 'is a tall human' will count as purely de dicto, while 'is identical to George Bush' will not.

[45] I am assuming for simplicity that 'F' is a purely de dicto predicate, though one would ultimately need to extend the account so as to handle cases where it is not.

explain how, although there is no possible-world in which Hesperus is bright and Phosphorus is not bright, the above ascriptions can both be correct: assume that Hesperus (/Phosphorus) is both the unique planet that shines in the evening, and the unique planet that shines in the morning, and also that in every world in Jill's total belief state the unique planet the shines in the evening is bright, and the unique planet that shines in the morning is not bright (all of this, of course, is consistent with Jill's total belief state being non-empty, because in each such world it may be a different planet that shines in the evening and in the morning). Thus given a successful theory of 'adequacy', the idea can be generalised to address Frege's puzzle across the board.

One crucial issue for descriptivism, however, is ensuring that there is always an adequate description which on the one hand succeeds in uniquely picking out the object in question and on the other hand is 'purely de dicto'. As it turns out, it is in most cases extremely hard to satisfy both these constraints at once. Consider for example my belief that Obama is tall. Keeping in mind the role of adequacy in solving Frege's puzzle, it is likely that an adequate description in this instance will be something of the form 'The president of the United States in 2012'. But the problem is that this not a purely de dicto description: it mentions the singular terms 'United States' and '2012'.

Suppose however, that we slightly relax the constraint that the description be purely de dicto, and instead require that it be 'de dicto or de se': that is, we will not allow the description to contain any proper names, but we will allow it to contain indexicals such as 'I' or 'now'. We can then plausibly find an adequate description which uniquely picks out Obama in relation

to the above ascription - for example: 'the *current* president of the country that *I* am visiting *now*'.[46]

All this brings us back to the de se, and to an amended version of descriptivism, which I call 'centred descriptivism'[47]:

Centred Descriptivism about belief

(1) If 'p' is a purely de dicto *or de se* statement: X believes that p if and only if X's total *centred* belief state is included in the *centred*-proposition that p.

(2) If 'p' is a de re statement (for simplicity we will focus only on the case where 'p' has the form '*a* is F'): X believes that p, if and only if there is some adequate, purely de dicto *or de se* description 'G' for which the following hold:

(a) *a* is the unique G.

(b) *x* believes that the unique G is F.

As above, it is important to note that clause (2b) only appeals to de dicto or de se beliefs, and these are already handled via clause (1).

Let us take stock: descriptivism offers a potential way for the possible-worlds account to handle Frege's puzzle by reducing de re belief ascriptions to de dicto ones, and handling the latter using the standard possible-worlds account. However, it appears that the view does not quite work as it stands, and one ultimately needs to allow a reduction to de dicto *or de se* beliefs. Centred descriptivism offers a way of doing that, by handling de dicto or de se

---

[46] Another challenge for descriptivism is that one might argue that there is no adequate description for indexical singular terms such as 'I' (see Perry (1979)). The shift to centreddescriptivism can be motivated by this problem as well.

[47] Note that Lewis's proposal for handling the de re in Lewis (1979), §VIII is essentially a version of centred descriptivism, with adequacy defined in terms of acquaintance relations.

ascriptions directly via the centred-worlds account. Thus centred descriptivism suggests both that the centred-worlds framework can successfully overcome the problems that plague the standard possible-worlds account, and more importantly, that handling these problems in the case of the de se really does require a special apparatus (the centred-worlds framework), on which the more general solution (i.e. reducing de re belief ascriptions to de dicto or de se ones) rests.

If successful, the reducibly argument provides strong support for the myth of the de se. But unfortunately for the supporters of the myth, the argument is not successful. The problem is that although centred descriptivism offers an improvement on ordinary descriptivism, it nevertheless fails to provide a satisfactory account of attitudes.

In *Naming and Necessity*, Kripke famously argued against descriptivism about de re *contents*: the view that statements of the form '*a* is F' have the same truth conditions as statements of the form 'The G is F'. Descriptivism (either simple or centred) about belief is a subtler view than descriptivism about content: it is not in any way revisionary about the content of statements of the form '*a* is F' (these can express ordinary singular propositions), but rather is revisionary about *ascriptions of belief* involving such statements. Nevertheless, it turns out that many of Kripke's objections to descriptivism about content have natural parallels for descriptivism about belief, and these are no less damaging. Here, in a brief, are some of the problems the account faces:

(1) *The problem of error:* on the face of it, an agent can have de re beliefs about an object even if they have massively false beliefs about that object. Suppose, for example, that the only relevant beliefs an agent has concerning Gödel are that he was a famous physicist who discovered the theory of relativity and was called 'Gödel'. It thus seems that any

description that might count as adequate for representing the agent's de re beliefs concerning Gödel, would be one that does not in fact apply to Gödel.[48]

(2) *The modal problem:* consider my belief that possibly, Obama is not a president. If the (only) adequate description for my de re beliefs concerning Obama is 'the president of the country I am visiting', then according to (centred) descriptivism, I believe that possibly, Obama is not a president, just in case I believe that possibly, the president of the country I am visiting is not a president. But clearly, I do not have the latter belief.[49] As with descriptivism about content, this problem can be circumvented by choosing a rigidified one, such as 'The *actual* president of the country I am visiting'. Plausibly, I do believe that possibly, the actual president of the country I am visiting is not a president. However, this solution does not generalise to other versions of the problem – ones that concern what we might call 'counterfactual attitudes' (attitudes that one can have towards propositions that one knows to be false, e.g. imagining, pretending, or wishing).[50] Thus for example, suppose that I am pretending that Obama is not actually the president.[51] Presumably, I can pretend this, while not pretending that the actual president of the country I am visiting is a not actually a president (the former would be very easy to pretend – for example, I might

---

[48] It will not help to use the description 'The person called 'Gödel'' because there may be another person called 'Gödel' (or the agent may believe there is another such person). Another proposal for attempting to solve both this and other problems below is to use descriptions such as 'The person this thought is about'. One problem with such descriptions is that descriptivism is intended to provide a theory of what it takes for a thought to be about a particular object so that there is something circular – possibly viciously circular - about such descriptions. But more importantly, such descriptions do not help because they are not purely de dicto or de se (as required by the theory), but rather de re. This is so because they contain the singular term 'this'. (Note that the centred-worlds framework as it stands does not help in directly treating the demonstrative 'this' - the objects at the centres are thinkers, not token thoughts. And it is far from easy to see how the theory could be amended so that the centres consist of token thoughts instead).

[49] Note that on the possible-worlds account *I could not* (practically) have this belief because that would require my total belief state to be empty. But we can leave this issue aside here: even if I can believe some impossible propositions, I simply do not believe this particular impossible propositions.

[50] See Ninan (MS) for a detailed discussion of this problem.

[51] Of course, the possible-worlds account will have a problem explaining how I can pretend this impossible proposition, but this is another problem the account would need to address.

pretend that Hilary Clinton is actually the president; The latter, on the other hand, involves a very odd pretence).

(3) *The uniqueness problem:* We have already noted that it is often difficult to find descriptions that are at the same time adequate, as well as pick out the relevant object uniquely. The shift to centred-descriptivism and the corresponding concession that the relevant descriptions may contain indexicals went some way to address this problem, but it is far from sufficient to eliminate it. Suppose for example that George and Jack are identical twins, and I briefly meet each of them at a party. Suppose that when meeting George, I think he is a nice guy, and form the de re belief that George is a nice guy. But suppose also that all I recall concerning this brief encounter are properties that George shares with his identical twin (he's one of the twins, I met him at the party, he has dark hair, he was nice, etc.).[52] It seems that any adequate description we might use to account for my de re belief about George (e.g. 'The dark-haired twin I met at the party yesterday'), will fail to pick him out uniquely, and this is so *even if the description is allowed to contain indexicals*. The introduction of indexicals and the move to centred-descriptivism is simply insufficient to solve the uniqueness problem.

(4) *Purely de dicto or de se versions of Frege's puzzle*: Centred descriptivism attempts to solve the de re versions of Frege's puzzle by reducing de re beliefs to de dicto or de se ones. However, this ignores the fact that Frege's puzzle arises not only with respect to de re beliefs. There are purely predicative, de dicto versions of the puzzle: for example, I may believe that all physicians are physicians, without believing that all doctors are physicians. Similarly, as we have seen in in §2.3, there can also be purely de se versions

---

[52] Note that allowing for implicit memories or beliefs will not help here – I may have identical implicit as well explicit memories and beliefs concerning each of the two twins.

of Frege's puzzle. Centred descriptivism (at least as it stands) has nothing to offer by way of a solution to these versions of the puzzle.

A final remark is in order before we dismiss the reducibility argument altogether. There is a slightly more nuanced version of centred-descriptivism (call this 'Complex Centred Descriptivism' or 'CCD'), which can potentially circumvent at least some of the above problems. Recall that while for the case of de re beliefs, centred descriptivism complicates the connection between ordinary belief ascriptions and an agent's centred belief state, at least for the case of de dicto or de se beliefs, the view keeps the straightforward connection (an agent believes that p just in case the agent's total centred-belief state is a subset of the centred-proposition that p).

CCD severs this simple connection even in the case of de dicto or de se belief. Suppose there is no simple algorithm which connects what an agent believes (in the ordinary sense of the term), and those centred-propositions that are part of the agent's total centred-belief state. Suppose also that we amend clause (2b) so that it is phrased directly in terms of the agent's total centred-belief state, rather than in terms of an ordinary belief ascription. That is, on the amended proposal clause 2b reads 'The centred-proposition that the unique G is F is part of the agent's total centred-belief state' (without committing to the claim that the agent believes that the unique G is F).

How might this help? At least as stated, this amended view places far less constraints on the relevant descriptions. Consider for example, the problem of uniqueness. Suppose that George, but not his identical twin, had a tiny spec of dust on his forehead in the party. Suppose that I had not noticed this spec or completely forgot about it, and that we would thus reject the claim I believe that one of the twins I met in the party had a spec of dust on his

forehead. Still, it is compatible with CCD that even though I do not *believe* that one of the twins I met at the party had a spec of dust on his forehead this proposition is nevertheless part of my total centred-belief state. We could thus allow the relevant description to be 'The twin I met at the party, and who had a spec of dust on his forehead' – and this description does uniquely pick out George. Similar remarks apply to the problem of error: CCD allows that while I do not *believe* that there is a unique person called 'Gödel' who proved the incompleteness theorems, this proposition is nevertheless part of my total centred-belief state, and hence the description 'person called 'Gödel' who proved the incompleteness theorems' might serve as the adequate description in this case.[53]

CCD could potentially avoid (at least some of) the problems I raised above, but the view faces obvious objections. At least without a substantial further theory which connects belief to an agent's total centred-belief state, we might completely lose grip on what an agent's total centred-belief state is supposed to represent, and which centred-worlds might be included in it. Furthermore, even if such a further theory can ultimately be supplied, it is worth realising that to what extent CCD has given up on the simplicity and elegance which motivated possible-worlds account of attitudes in the first place (this simplicity and elegance was already significantly undermined with the shift to centred-descriptivism, but the move to CCD exacerbates the problem). Thus if the way to save the possible-worlds account is to shift to CCD, it is far from clear that the account is worth saving. But most importantly for our current purposes, the move to CCD will not be of any service to defenders of the myth of the de se. After all, given the extreme flexibility which CCD allows in offering suitable

---

[53] It is not however clear that CCD will help with the problem of counterfactual attitudes: suppose the adequate description for accounting for my attitudes towards George is 'The twin with the tiny spec of dust on his forehead'. I could still pretend that George does not have a tiny spec of dust on his forehead, but presumably the proposition that the person with a tiny spec of dust on his forehead does not have a tiny speck of dust on his forehead does not belong to my total (centred) pretence state. And CDD also seems to be of no help when dealing with problem (4) above.

descriptions, one might go back and question the need for *indexicals* in the relevant descriptions. Consider a version of the view which does not appeal to indexicals, and hence to centred-worlds at all (call this 'Complex Non-centred Descriptivism' or 'CND'). CND might have just as much resources to handle the above problems as CCD. For example, we might maintain that I believe that Obama is tall, in virtue of the fact that my total (non-centred) belief state includes the proposition that the only person who had precisely 15 hairs on his head when he was precisely 100 minutes old is tall (again, it is crucial that I do not need to believe this claim). CCD is thus not sufficient to save the reducibility argument. Despite its initial attraction, the reducibility argument fails to support the myth.[54]

§4 Conclusion

In this paper I have argued against the myth of the de se. My main focus has been to show that wide variety of arguments that are used to motivate the myth in general and the Special Challenge claim in particular are not compelling.

A defender of the myth, might however, opt for a last retreat: perhaps Special Challenge should be abandoned, but one can still maintain Special Attitudes. That is, perhaps there is a distinctive class of propositional attitudes that might be thought of as 'de se', even if this class does not require any amendment of our general account of attitudes.

---

[54] A different modification of centred descriptivism would be to give up clause (2a), namely allow that one can believe that '*a* is F' in virtue of believing that 'The G is F' without requiring that *a* actually satisfy the description 'G'. (See Stalnaker (1981) for a proposal along these lines). Putting aside any independent challenges for this view (e.g. accounting for how the relevant belief gets to be about *a*), the same remarks hold here as above: this view will not help to defend the myth, because the view is sufficiently flexible that if it works at all, there is no reason to opt for a centred rather than non-centred version of the view. (Indeed, this is a point Stalnaker would presumably endorse, as his view is presented as part of an argument for the possible-worlds over the centred-worlds account).

The claim that amongst our propositional attitudes there is a natural subclass of attitudes that might be thought of as 'de se' is hard to refute without some concrete positive suggestion as to how to characterise the relevant class. But it is worth noting that while my discussion has not been primarily targeted at the Special Attitudes claim, it nevertheless casts doubt on various features which are commonly thought to distinguish such attitudes. For example, my remarks concerning individual action in §1.5 and shared action in §1.6 cast doubts on the thought that so called de se attitudes play an essential or unique role in the explanation of action; My discussion of purely de se ignorance in §2.3, questions the idea that each of us has a *unique* first-personal way of representing ourselves to ourselves; Finally, the cases of purely de se ignorance also challenge the thought called de se attitudes have a certain kind of epistemically privileged status. It is often assumed, for example, that de se beliefs are *immune to error through misidentification* (IEM).[55] Different authors construe this notion slightly differently but to use the definition from Stanley: "*A judgment Fa is immune to error through misidentification if and only if, when it is defeated, its grounds cannot survive as the sole grounds for the existential generalization that something is F*".[56] The thought is that while my belief that Sally is in pain is not IEM (because I may discover the screams from the other room are produced by Bill, not Sally – and hence my initial grounds would survive as grounds for the claim that *someone* is in pain), my de se belief that *I* am in pain is IEM. However, cases such as that of Jake in §2.3 cast doubt on the soundness of this criterion. After all, suppose that Jake thinks in voice-1 'I am tired', but then receives (misleading) evidence that voice-1 is not his but rather that of an alien speaking to him. Jake would presumably still have grounds to believe that someone (namely the alien) is tired.

---

[55] See for example Shoemaker (1968), Higginbotham (2003), Recanti (2007) book II, and Stanley (2011), pp. 89-94.

[56] Stanley (2011), p. 90.

Of course, all this is nevertheless consistent with the hypothesis that amongst our attitudes, there is some natural subclass of de se attitudes, which does not have any of the above characteristics. I leave it as an open question (one which should probably be addressed primarily within the philosophy of mind or psychology) whether it is ultimately possible to delineate any such subclass in a way which is philosophically interesting or significant. But either way I maintain that the category of de se attitudes (if there is one) does not play any important role in the semantics of attitude reports or require any special amendment of our general account of propositional attitudes. The myth of the de se remains just that.[57]

Appendix: de se attitudes and indexical reports

In this appendix I offer some remarks on an additional line that is often accepted by those sympathetic to the myth: the connection between de se attitudes and indexical reports, and more specifically PRO reports.

A common line of thought amongst defenders of the myth is that special, de se kind of attitude, is closely tied to the notion of *indexicality*. More specifically: we can identify de se attitudes as those that are reported using indexical terms (for example, by reporting my belief that *I* am in the Stanford library, I report a de se attitude).

As should be clear by now, I maintain that even if this claim were true, this does not in itself justify the Special Challenge claim, and thus the myth more generally. But this point aside, we might ask whether there is indeed such a close connection between indexical reports and de se attitudes.

Very few, I assume, will argue that using *any* indexical term in a belief report is sufficient for the report to describe a de se attitude: after all, many philosophers hold that terms such 'tall', 'rich', or even 'know' are indexical, but few would accept that 'Jill believes that Jack is tall' report a de se belief on Jill's part. But what if we focus on a more restricted set of indexicals such as 'I', 'now', or 'she'? Even restricting ourselves in this way, the use of indexicals is neither necessary nor sufficient for a belief report to concern the so called special de se attitude.

To see that it is not sufficient, consider for example the following case proposed by Moss: "Suppose Kaplan sees himself in a mirror, without realizing that he is seeing himself. Looking at the mirror, Kaplan sees that his pants are on fire, without realizing that his own pants are on fire. In recounting his experience, suppose Kaplan says:

(1) I expected that I would be rescued.

Kaplan can truly utter (1), even though he was not aware of being in danger when he looked at the mirror". (Moss (*forthcoming*), 3).

Thus even though Kaplan reports his earlier attitude using the indexical 'I', he is not reporting a de se belief.

To see that the use of indexicals is not necessary, consider the following: suppose that I coin the term 'ME' to refer to myself. I fully realise that 'ME' refers to myself, and use it just as I would use the term 'I'. But 'ME' is not an indexical: it is a private name which I use to refer to myself (and presumably, no one else uses to refer to anything). In so far as there are distinctively de se beliefs, it seems clear that typically, a belief I would report by saying 'I

believe that ME is tall' would be reporting a de se belief, and thus use of indexicals is not necessary for de se reports.[58]

Amongst those who acknowledge that reports involving ordinary indexicals such as 'I' may not be necessary or sufficient for reporting de se attitudes, a common fallback position is to instead tie de se attitudes to reports involving the special covert pronoun 'PRO'.[59]

Consider again Kaplan's report in (1) above, and contrast that with the following report:

(2) I expected to be saved.

According to a widely accepted syntactic hypothesis, sentences such as (2) contain the covert pronoun 'PRO', and their structure is of the following form[60]:

 (2') I expected PRO to be saved.

And indeed there is a contrast between (1) and (2): while in the case described above, (1) is an adequate report, (2) seems inadequate, because it seems to suggest that Kaplan recognised the agent he expected to be saved as himself.

While PRO-reports have indeed a much stronger claim to being the hallmark of de se attitude reports, the connection between the de se and PRO reports is often exaggerated. For a start, PRO-reports are certainly not *necessary* for de se reports: after all, on many occasions we are

---

[58] See Tiffany (2000), §IV for evidence that some actual proper names have this kind of cognitive role. Also relevant here is Millikan (1990).

[59] More specifically, to subject control PRO, when it appears in the environment of attitude verbs. Stephenson, for example, remarks that "it is well known that controlled PRO must be interpreted de se in the sense of Lewis (1979)—that is the attitudes expressed crucially involve the attitude holder's access to their own 'self'". (Stephenson (2010), 210). For further support of this claim see Chierchia (1989), Recanati (2007), Ninan (2010), and Schlenker (*forthcoming*).

[60] See Carnie (2002), ch.10.

happy to report so called de se attitudes using ordinary indexicals such as 'I' or 'he'.[61] Are PRO-reports at least *sufficient* for de se reports? Even here, the data is less clear than it is commonly presented in the literature. Consider the following scenario: Jack is in a hospital bed. There are altogether five beds in the hospital, marked one to five. Jack does not know which bed he occupies, but as it turns out he occupies bed number five. In front of Jack is a screen showing live footage from each of the five beds (including their numbers). Jack sees the footage of bed number five, and notices that the patient in bed number five (which, for all he knows may or may not be him), is looking extremely faint and urgently needs some water. He cries out to the nurses: 'Water for bed number five, please!'. One of the nurses, who is fully aware of the scenario, says to his colleague 'Jack asked to be given some water'. This report seems to be an entirely adequate one.[62] But note first, that this report uses the covert indexical 'PRO', and second, that it reports a non-de se attitude.

One could wonder whether 'ask' should count as an attitude verb in this context (recall that the claim I am arguing against is that subject control PRO always indicates a de se attitude in the environment of attitude verbs). It is worth noting, that Schlenker, at least, seems to treat the similar verb 'tell' as an attitude verb in the context of this discussion.[63] Moreover, 'ask'-reports do seem to exhibit the same sort of dichotomy between so called 'de se' and 'de re' readings that other attitude reports give rise to. (For example, there is an intuitive difference between the default readings of the report 'Jill asked that Jill be given some water' and 'Jill asked that she be given some water' or 'Jill asked to be given some water'). Finally, in the above scenario even the report 'Jack *wants* to be given some water' seems adequate (though intuitions may be less clear in this case). While the above remarks are far from being

---

[61] For a more nuanced argument that reports involving indexicals other than PRO sometimes receive a de se interpretation see Schlenker (*forthcoming*), §2.2.3.

[62] An anecdotal survey suggests that this is a widely shared intuition.

[63] See Schlenker (forthcoming), §2.1.

conclusive, they should I think give us reason to rethink whether the connection between

PRO-reports and the de se is as tight as clear as it is usually taken to be. [64]

---

[64] Note that the data is even clearer when one reconsiders the purported connection between second-person uses of PRO, and so called 'de te' reports (see e.g. Ninan (2010), pp. 554-555). Suppose John announces in the speaker of a loud party: 'Could Mary please move her car!'. Jane, not hearing very well asks Jill what John just said, and Jill responds by saying 'John asked Mary to move her car'. Again this second-person PRO report seems perfectly adequate, even though this is clearly not reporting any de te ('Mary: you should move your car!') asking.

References

Bermudez, J. L. (2005), 'Evans and the sense of "I"', in Jose Luis Bermudez (ed.), *Thought, reference and Experience: Themes from the Philosophy of Gareth Evans,* Oxford University Press, pp. 164-194.

Boer, S., and Lycan, W.G. (1980), 'Who me?', *Philosophical Review* 89: 130-157.

Burge, T. (1979): 'Individualism and the mental', *Midwest Studies in Philosophy 4:* 73-121.

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

Cappelen, H., and Hawthorne, J. (2009), *Relativism and Monadic Truth*, Oxford University Press.

Carnie, A. (2002), *Syntax: A Generative Introduction*, Blackwell Publishing.

Castaneda, H-N. (1966), 'He*: A study in the logic of self-consciousness', *Ratio* 8: 130-157.

Chierchia, G., (1989), 'Anaphora and attitude de se', in Renate Bartsch, Johan van Benthem and Peter van Emde Boas (eds.), *Semantics and Contextual Expression*, Foris Publications, pp. 1-31.

Chalmers, D. (2006), 'The foundations of two-dimensional semantics', in Manuel Garcia-Carpintero & Josep Macia (eds.), *Two-Dimensional Semantics: Foundations and Applications*, Oxford University Press, pp. 55-140.

Cresswell, M. J., (1985), *Structured Meanings: The Semantics of Propositional Attitudes,* MIT Press.

Dennett. D. (1981), 'Where am I?', in his *Brainstorms: Philosophical Essays on Mind and Psychology* , MIT Press, pp. 310-323.

Devitt, M. (*forthcoming*), 'The myth of the problematic de se', forthcoming in Alessandro Capone and Neil Feit (eds.), *Attitudes De se: Linguistics, Epistemology, Metaphysics*, CSLI.

Egan, A., (2006), 'Secondary qualities and self-location', *Philosophy and Phenomenological Research* 72: 97-119.

Egan, A., (2007), 'Epistemic modals, relativism, and assertion', *Philosophical Studies* 133:1-22.

Elga, A. (2000), 'Self-locating belief and the sleeping beauty problem', *Analysis* 60: 143-147.

Frege, G. (1948), 'Sense and reference' *Philosophical Review* 57: 209-230. (originally published in 1892).

Forbes, G. (1990), 'The indispensability of Sinn' *Philosophical Review* 99: 535-563.

Higginbotham, J. (2003), 'Remembering, imagining, and the first person', in Alex Barber (ed.), *Epistemology of Language*, Oxford University Press, pp. 496-533.

Hitchcock, C. (2004), 'Beauty and the bets', *Synthese* 139: 405-420.

Kripke (1980), *Naming and Necessity*, Harvard University Press.

Liao, S. (2012), 'What are centered worlds', *Philosophical Quarterly* 62: 294-316.

Lewis (1979), 'Attitudes de dicto and de se', *Philosophical Review* 88: 508-543.

Lewis, D. (1986), *On the Plurality of Worlds*, Blackwells.

McFarlane, J. (2005), 'Making sense of relative truth', *Proceedings of the Aristotelian Society* 105: 321-39.

Meacham, C. (2008), 'Sleeping beauty and the dynamics of de se belief', *Philosophical Studies* 138: 245-269.

Millikan, R. (1990), 'The myth of the essential indexical', *Nous* 24: 723-34.

Moss, S., (*forthcoming*), 'Updating as communication', forthcoming in *Philosophy and Phenomenological Research*.

Ninan, D. (2009), 'Persistence and the first-person perspective', *Philosophical Review* 118: 425-464.

Ninan, D. (2010), 'De se attitudes: ascription and communication', *Philosophy Compass* 5: 551-567.

Ninan, D. (MS), 'Counter-factual attitudes and multi-centred worlds', unpublished manuscript.

Nolan, D. (2006), 'Selfless desires', *Philosophy and Phenomenological Research* 73: 665-679.

Perry, J. (1977), 'Frege on demonstratives', *Philosophical Review* 86: 474-497.

Perry, J. (1979), 'The problem of the essential indexical', *Nous* 13: 3-21.

Pietroski, P. (1996), '*Fregean* innocence', *Mind and Language* 11: 338-70.

Putnam, H. (1975), 'The meaning of 'meaning'', *Minnesota Studies in the Philosophy of Science* 7: 131-193.

Recanati, F. (2007), *Prespectival Thought: A Plea For (Moderate) Relativism*, Oxford University, Press.

Richard, M. (1993), 'Attitudes in context', *Linguistics and Philosophy* 16: 123-148.

Salmon, N. (1986), *Frege's Puzzle*, MIT Press.

Schlenker, P. (*forthcoming*), 'Indexicality and de se reports', forthcoming in Klaus von Heusinger, Claudia Maienborn, and Paul Portner (eds.), Handbook of Semantics, Mouton de Gruyter.

Shoemaker, S. (1968), 'Self-reference and self-awareness', *Journal of Philosophy* 65: 555-567.

Soames, S. (1989), 'Direct reference, propositional attitudes and semantic content', *Philosophical Topics* 15: 44-87.

Spencer, C. (2007), 'Is there a problem of the essential indexical?', in Michael O'Rouke and Corey Washington (eds.), *Situating Semantics: Essays on the Philosophy of John Perry*, MIT Press, pp. 179-197

Stanley, J. (2011), *Know How*, Oxford University Press.

Stalnaker, R. (1978), 'Assertion', *Syntax and Semantics* 9: 315-32.

Stalnaker, R. (1981), 'Indexical belief', *Synthese* 49: 129-151.

Stalnaker, R. (1984), Inquiry, MIT Press.

Stalnaker, R. (2008), *Our Knowledge of the Internal World*, Oxford University Press.

Stephenson, T., (2010), 'Control in centered worlds', *Journal of Semantics* 27: 409-436.

Tiffany, E. (2000), 'What is essential about indexicals?', *Philosophical studies* 100: 35-50.

Torre, S. (2010), 'Centered assertion', *Philosophical Studies* 150: 97-114.

Turner, J. (2010), 'Fitting attitudes de dicto and de se', *Nous* 44: 1-9.

Williamson, T. (2000), *Knowledge and its Limits*, Oxford University Press.