

PRIMITIVE SELF-ASCRPTION: LEWIS ON THE *DE SE*¹

RICHARD HOLTON, MIT

In a memorable children's story, Winnie-the-Pooh follows the tracks of what he thinks might be a woozle; until he realizes that he has been 'Foolish and Deluded' and that the tracks are his own.² Similar ideas recur in many stories from classical mythology to science fiction.³ Characters take various attitudes towards people, whilst not realizing that they are themselves the very people concerned.

Stories like these do not just appear in fiction; they appear in philosophy too. Let us start with two examples (we shall see more later). John Perry recalls following a supermarket shopper who was spilling sugar from his shopping cart, only to realize that he, like Winnie-the-Pooh, was following his own trail. More fancifully, David Kaplan imagines seeing the reflection in a window of someone whose pants are on fire, whilst failing to realize that it is his own pants that are on fire.⁴

We can see that such stories might be engaging; but why are they *philosophically* interesting? The reason is that they highlight a difference between knowledge about the world, impersonally conceived, and knowledge about our place in the world. As David Lewis graphically puts it, they highlight the difference between the information given by a standard map, and that given by a map that is erected in a public place with a 'You are here' arrow. One tells you about the nature of the world; the other tells you, in addition, how you fit into that world. Following Lewis, call the former *de dicto* knowledge, and the latter *de se*.

The challenge posed by such a distinction is to come up with an account of the *de se*. Standard theories find it hard to do so. Standard theories treat our knowledge (and our beliefs, desires etc.) as involving attitudes towards propositions. There is some disagreement as to quite what

¹ Thanks to audiences at MIT, Paris and Cambridge, and to Rae Langton, Steve Yablo and Seth Yalcin.

² A.A. Milne, *Winnie-the-Pooh* (London: Methuen, 1926) Ch. 3 'Pooh and Piglet Go Hunting and Nearly Catch a Woozle'.

³ In one version of the story of Jason and the Argonauts, Jason suggests that a would-be assassin should be sent to fetch the Golden Fleece without realizing that he is the would-be assassin. In a rather more recent story, Robert Heinlein has a time-travelling character consistently failing to recognize himself: 'By his Bootstraps' in *The Menace from Earth* (New York: Gnome Press, 1959).

⁴ 'Demonstratives' in Joseph Almog, John Perry & Howard Wettstein (eds.), *Themes From Kaplan* (New York: Oxford University Press, 1989). p. 533

propositions are, but on none of the main approaches is there an obvious way to understand *de se* knowledge.

One approach takes propositions to be structured entities. For Russell, the structured proposition corresponding to Kaplan's belief that that man's pants are on fire is the ordered pair of the man in question, and the property of being on fire. Crucially, given that the man in question is Kaplan, nothing changes when we turn to the proposition corresponding to Kaplan's belief that his own pants are on fire. The object of that belief is again the ordered pair of the man in question—i.e. Kaplan—and the property of having burning pants.

There is more opportunity to make a distinction on Frege's account of structured propositions, which involves not just the objects and properties concerned, but also the ways that these things are thought of: in Frege's term, 'modes of presentation'. Here then the proposition corresponding to Kaplan's belief that that man's pants are on fire might be different from that corresponding to his belief that his own pants are on fire, since they might involve different modes of presentation of the same object. But what is the mode of presentation that one has of oneself? It doesn't look as though it could correspond to any normal description, since the same worries would apply: one could think that someone met that description, without realizing that one met it oneself. Frege concluded that 'Every one is presented to himself in a particular and primitive way, in which he is presented to no one else.'⁵ But it is unclear quite what this primitive way is, and, as Frege acknowledged, it makes communication indirect: I cannot communicate my *de se* thought to you, since you cannot grasp it.

If Frege's approach leads to obscurity, the alternative approach that takes propositions to be unstructured is in even worse trouble. The central proposal here is Lewis's own, according to which propositions are classes of possible worlds. But if names and demonstratives are rigid, referring to the same thing in each world, then, if it is Kaplan's pants that are on fire, the class of worlds in which 'that man's pants are on fire' is just the same as that in which Kaplan's pants are on fire. We have nothing to distinguish the *de se* belief from the *de dicto*.

In response to these difficulties Lewis proposes an account that, characteristically, is at once simple and radical.⁶ Do not think of *de se* thoughts as attitudes to propositions at all. Think of them rather as self-ascriptions of properties. When Perry realizes that he is following his own trail, he self-ascribes the property of spilling the sugar; when Kaplan realizes that it is his own reflection that he can see, he self-ascribes the property of having flaming pants. Moreover, the *de dicto* can then be thought of as a special case of the *de se*: the case in which what is self-ascribed is the property of membership of a world.

Lewis's account thus has two parts. The first part, which is the focus of most of his discussion, and of most of the discussion that has followed, involves treating the objects of the attitudes,

⁵ Frege 'The Thought' in *Collected Papers* (Oxford: Blackwell, 1984)

⁶ Lewis, 'Attitudes De Dicto and De Se' *Philosophical Papers I* (New York: Oxford University Press, 1983) pp. 133–55. See also *On the Plurality of Worlds* (Oxford: Blackwell, 1986) and 'Individuation by Acquaintance and by Stipulation' in *Papers in Metaphysics and Epistemology*, (Cambridge: Cambridge University Press, 1999) pp. 373–402.

not as *propositions* but as *properties*. The second part, which is much less to the fore in Lewis's presentation and in the subsequent discussion, involves treating our attitude to these properties as that of *self-ascription*. I hyphenate the term to stress that this is a primitive notion for Lewis. To self-ascribe a property is not just to ascribe a property to oneself, as one might ascribe it to someone else. That would just give us back the question of what it is to think of oneself in the right way. After all, when Kaplan ascribes the property of having burning pants to the figure in the window, he does ascribe that property to someone who is indeed Kaplan; in one sense then, he ascribes the property to himself. What Lewis wants to insist is that he does not self-ascribe that property. The self and the ascribing do not detach.

I will argue that, as a result of underplaying its second part, the radical nature of Lewis's account has been largely missed. In particular, much recent literature has tried to incorporate his account simply by treating the objects of the attitudes as *centered worlds*, where a centered world is an ordered pair of a possible world together with a spatio-temporal location. But centered propositions just correspond to properties. So whilst they give us the right *objects* for the Lewis account, they go no way to providing the right *attitude*. (Moreover, I shall argue that they provide an unnecessarily baroque way of specifying contents.)

As a result of this neglect, many of the difficulties that Lewis's approach faces have been overlooked. The idea of primitive self-ascription is an obscure one. Our natural grasp on it is via the general idea of ascribing a property to an object, but this is exactly what we are not allowed to do here. Instead we have to treat the relation as fundamental, something that becomes increasingly hard to do when we consider first person plural ascriptions, and ascriptions where the first person pronoun is not in subject position. These difficulties do not render Lewis's account totally unworkable; but they do make it much more problematic than has been generally recognized.

LEWIS'S METHOD

Let us start by getting clear on exactly what Lewis takes himself to be doing, and on the techniques he uses. His main contention is that propositions cannot be the objects of the attitudes, since there are situations in which an agent has a great deal of propositional knowledge, and yet lacks some further *de se* knowledge. The idea then is that this further knowledge cannot be knowledge of a proposition, i.e. it cannot be *de dicto*. (I will follow Lewis and treat talk of propositional belief and of *de dicto* belief as equivalent for now. The obvious problem with doing so is that many have thought that there exists a further category, *de re* belief, which is propositional without being *de dicto*. As we shall see, Lewis denies that this is really a separate kind of belief; but the issues there are tangential.)

One way of arguing for this makes use of examples like those that we have seen from Perry and Kaplan. These examples certainly establish that you can have some propositional knowledge about a situation and yet lack further *de se* knowledge about it. And the Kaplan cases shows, beyond that, that you can have direct perceptual knowledge that a property obtains of a person who happens to be you, without knowing that you—you *yourself* as Castañeda would say—have that property. What the cases don't show, at least not without a number of further

assumptions, is that this *de se* knowledge cannot be understood as propositional knowledge. For it could be that there is some way of deducing the *de se* knowledge from additional *de dicto* knowledge, if only one had enough of it. And if the *de se* knowledge is deducible from the *de dicto* knowledge, then presumably it in turn is *de dicto*.

This is where Lewis's approach is innovative. He wants to argue that you can have *all* of the *de dicto* knowledge that there is, and still lack *de se* knowledge.⁷ For this he needs some new examples, since the Kaplan and Perry cases lack this feature. If you filled in the propositional knowledge of the subjects involved—gave them information about the angle of the reflection, say, or about the complete extent of the sugar trail—they would be able to deduce that it was their pants on fire, or their trail of sugar. That is not so *say* that they would have deduced such conclusions from entirely *de dicto* beliefs; that would still need to be investigated. But it does mean that, without further elaboration, the examples will not serve to show that one can have complete *de dicto* knowledge and lack *de se* knowledge.

The main example that Lewis gives to make this point is that of the two gods:

Consider the case of the two gods. They inhabit a certain possible world and they know exactly which world it is. Therefore they know every proposition that is true in 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. They are not exactly alike. One lives on top of the tallest mountain and throws down manna; the other lives on the top of the coldest mountain and throws down thunderbolts. Neither one knows whether he lives on the tallest mountain or on the coldest mountain; nor whether he throws down manna or thunderbolts.⁸

This example is somewhat problematic, since it is rather hard to fill in the details. If the gods are able to choose their actions and then implement them, can they not infer who they are from seeing the results?⁹ So let me give another of Lewis's examples in which the difficulties, if still present, are rather less to the fore. This example, originating from Perry, is that of Lingens, lost in the Stanford Library, who has such complete amnesia that he does not know who, or where, he is. As Lewis develops the story, Lingens has an alter ego—Lauben let us suppose—who is lost, in a similar condition, in the Widener Library at Harvard. Let us further suppose that the Stanford and Widener Libraries are even more comprehensive than they are now—their works include all *de dicto* knowledge—and that Lingens and Lauben have read and remembered every word. Let us also suppose that the libraries are qualitatively identical inside; not only are there no 'You are here' maps on the walls, there are no stamps in the books saying 'Property of Widener' or 'Property of Stanford'; no Windsor chairs bearing crests emblazoned 'Veritas' or 'Die Luft der Freiheit weht' and so on. Then despite their complete *de dicto* knowledge,

⁷ In discussing an example of Castañeda's, Lewis writes: 'To support this claim [that attribution of knowledge *de se* is not equivalent to any attribution of knowledge *de dicto*] we need only find a case in which the editor knows well enough which of the worlds is his without knowing whether he is among the millionaires of his world.' 'Attitudes' pp. 139–40.

⁸ 'Attitudes' p. 139

⁹ For a good attempt to make sense of the possibility, see Robert Stalnaker *Our Knowledge of the Internal World* (Oxford: Oxford University Press, 2008) p. 56, n.6.

including their knowledge of there being two amnesiacs lost in different libraries, neither Lingens nor Lauben will know which they are.

Do they lack all *de se* knowledge? Surely not. If Lingens sneezes he knows that he sneezes; likewise for Lauben. (If the libraries really contain all *de dicto* knowledge, including reports of all the sneezing that goes on, then if one sneezes the other had better sneeze too; or else Lingens and Lauben would be able to work out who was who.¹⁰) And they will each know that they are either Lauben or Lingens. What each lacks is any *de se* knowledge that distinguishes him from the other, and so allows him to integrate his *de se* knowledge with the *de dicto* knowledge that is given to him by the books.

This is a powerful argument to show that *de se* isn't *de dicto*. It is reminiscent of Frank Jackson's famous *knowledge argument* for the non-reducibility of phenomenal knowledge to physical knowledge: Mary knows all of the physical facts of colour science, but because she has never seen a red thing, she does not know what red looks like.¹¹ Of course in that case Lewis doesn't want to acknowledge the existence of objects of knowledge that are not physical; he wants to understand the apparent phenomenal knowledge as knowledge-how.¹² But in the case of the *de se* he does want to move from what intuitively look like different knowledge states to the conclusion that there must be different objects of knowledge.

Before we examine the details of that account, let me remark on a potential distorting feature of Lewis's argument. Reflecting on cases like these leads naturally to the thought that if the *de se* and the *de dicto* are two independent bodies of knowledge, what is often needed is what I'll term *breakthrough knowledge*: a piece of knowledge that enables the two to be connected. And give the nature of the cases, it can seem that the breakthrough knowledge must be *de se* knowledge of spatio-temporal location. One of the gods realizes that he is the god on the highest mountain; Lingens realizes that he is in Stanford library. Once they have this piece of knowledge, everything else follows: once the wall between the two bodies of knowledge is broken, each can flow into the other.

However, this focus on spatio-temporal location as distinctive of breakthrough knowledge is misleading. For a start it is not necessary. Any uniquely identifying piece of *de se* information would do the job. Lingens could discover that he is called 'Lingens'; one of the gods might discover that he throws down thunderbolts. Second, for some creatures—permeable creatures who can share locations with others—it is not sufficient. More broadly, the idea that a single piece of breakthrough knowledge is enough to unify the *de se* and the *de dicto* is only true in the highly artificial cases that Lewis presents: cases in which the agents have complete and unified *de dicto* knowledge, complete and unified *de se* knowledge, and the ability to make the necessary inferences. Remove these conditions and spatio-temporal location will no longer always do the job. Perry might have had a GPS device that told him exactly where (and when) he was; that still

¹⁰ Don't push this thought too far, or we end up with similar problems to that of the two gods.

¹¹ Frank Jackson, 'Epiphenomenal Qualia', *The Philosophical Quarterly* 32 (1982) 127–36.

¹² In not knowing what red looks like, Mary lacks the ability to identify the color of things just on the basis of their appearance. See 'What Experience Teaches' in *Papers in Metaphysics and Epistemology*. (Cambridge: Cambridge University Press, 1999) pp. 262–90.

wouldn't have allowed him to identify himself as the person making the mess, since he lacked the relevant *de dicto* knowledge that it was the person located where he was who was making the mess. In creatures who lack some *de dicto* knowledge then, coming to know one's spatio-temporal location provides no guarantee of breakthrough knowledge. This will be important when we examine centred worlds.

LEWIS'S ACCOUNT OF THE *DE SE*

Let us examine Lewis's account a little more closely. As we have seen, it involves two features:

- (i) the content of the attitude is a property;
- (ii) the attitude taken towards the content is that of self-ascription.

As I have already stressed, taking self-ascription as primitive is crucial to Lewis's account. We normally think of ascription as a two place relation: one ascribes a property to a thing. Self-ascription would then be the special case where the thing is the self. But that won't do the work here. If the self is just thought of extensionally, then we would have no way to distinguish the belief that one's pants were on fire from the belief that the pants of someone, who is you though you don't realize it, are on fire. So we would have, instead, to think that the self came with a certain mode of presentation. If we were to do that, then Lewis's account would collapse into a rather baroque version of Frege's.

So we have to think of self-ascription as a one-place relation: one simply self-ascribes a property. There is a real question whether we can make sense of this, but I'll postpone that till later. For now note how little discussion there is of the idea, either in Lewis's own discussion or in the subsequent literature.

It is perhaps not surprising that Lewis gave little fanfare to this second feature, for when he presented his account this was not the innovation. Implicitly at least, Lewis's account of our attitude to propositions had been much the same—to believe a proposition is to self-ascribe membership of a set of worlds¹³—so he understandably focussed on what was new. Indeed, as mentioned above, Lewis now explicitly treats the *de dicto* as a special case of the *de se*: a *de se* belief in which the property that is self-ascribed is membership of a certain set of worlds.¹⁴

There is a second reason why the crucial role of self-ascription has been largely overlooked. In much recent work, Lewis's account has been understood in terms of *centred worlds*, where a centred world is a possible world together with a centre, typically (though not in all accounts) a

¹³ See 'Attitudes' p. 137.

¹⁴ As a result it is unclear that Lewis can give a plausible account of beings that can have only *de dicto* beliefs: that lack the concept of the self. Yet mightn't there be some? (Perhaps some of those creatures that fail the mirror test.)

set of spatio-temporal coordinates.¹⁵ On this way of thinking, the objects of the attitudes are classes of centred worlds. There seems to be no appeal here to a primitive relation of self-ascription. Indeed, there seems to be no appeal to properties as the objects of the attitudes either. So how do we get from Lewis's account to this? Let us examine the proposal in some more detail.

CENTRED WORLDS

This explanation of the account in terms of centred worlds originates with Lewis himself, in his development of some comments from Quine. As Lewis puts it, the idea is that whenever we have a property we have a centred world: the centred world in which the object at the centre has that property. Conversely, whenever we have a centred world we have a property: the property that is had by the object at the centre. So instead of phrasing the account in terms of properties, we can phrase it in terms of centred worlds.

The first thing to note is that, as Lewis presents things, the role for centred worlds is to stand in for properties. So we still need the idea that they are self-ascribed. That is something that is obscured in much of the subsequent literature, where centred worlds are often taken to play the role of propositions, i.e. as things that are straightforward objects of belief, rather than of self-ascription. Introducing centered worlds does not change the fundamentals of the account; it just changes the way we describe it.

Why would we want to couch the description in terms of centred worlds rather than properties? A common idea is to view the centre as akin to the 'you are here' flag on Lewis's map. Or, less metaphorically, as David Chalmers puts it, 'We can think of the centre of the world as representing the perspective of the speaker within the world.'¹⁶ Still, if this is going to help, we must get clear on what such a perspective is.

We need to put one possible misunderstanding immediately out of the way. The centre cannot be the perspective that the thinker actually has on the world, the place where they are actually located. Of course in working out the denotation of indexical expressions we often need to know that: in working out what a speaker denotes with a use of the term 'Here' we need to know where they actually are. But this is not what we are doing when we use a centred world to represent the content of a belief state. After all, a thinker can make a mistake about where they are, in which case the centre of the possible world will not be their actual location.

So would it be better to think in terms of where the speaker *believes* her perspective is? That is better, but it is still not quite right. To see this, consider some cases.

¹⁵ For a summary of the work using centred worlds, and of the roles to which they have been put, see Shen-yi Liao, 'What are Centred Worlds?', *Philosophical Quarterly* 62 (2012) pp. 294–316.

¹⁶ 'The Foundations of Two-Dimensional Semantics' in Manuel García-Carpintero and Josep Macià (eds.) *Two-Dimensional Semantics* (Oxford: Oxford University Press, 2004) at p.60.

(i) *Lingens discovers that he is in Stanford Library*

This is easy to represent on the centred world approach. Since he knows all the *de dicto* facts, he will have narrowed the possible world that he is in down to one. When he discovers that he is in Harvard library this will give him his spatio-temporal location, which can serve as the centre of his centred world. Self-ascribing the property of being at that location will enable him to make all of the other self-ascriptions, he will know his name, date of birth and so on. The location thus serves as his breakthrough belief.

(ii) *Lingens is unsure whether he is in Stanford or Harvard*

Now we return to his state before the discovery. Here the case is like the previous one, except that there are two places in which Lingens thinks that he might be. So his belief is modelled by two centred worlds, one centred on Harvard, and one centred on Stanford. Each of the two provides a potential breakthrough point, in the sense that were he to self-ascribe that location, then that would serve to integrate his *de se* and his *de dicto* knowledge.

(iii) *The Consul knows only that he's in pain*

We leave the upright Lingens to focus on the dissolute Consul. After hard nights out, the Consul sometimes awakens with no knowledge of who, or where, he is. For the short time before he opens his eyes, all he knows is that he is in pain. On the centred world approach, the content of his thought amounts to the full set of possible locations; it is just that each location is paired with a world in which there is a person in pain at that location. This brings out how odd the centred world approach is. We started with the idea that the Consul simply self-ascribes a single property; he believes far less than Lingens. But that is cashed out using a far larger content: not just huge array of different worlds, but also a huge array of different centres. Is this a problem? I think not, but it does make clear what these centred worlds are doing in Lewis's proposal. They are not meant to be psychologically realistic, in the sense that agents would actually have an idea of each of the worlds and the centres. Rather, they simply work to describe, or index, the various thoughts that the agent might have.¹⁷

(iv) *Felicité thinks she may be the Holy Ghost*

Felicité has come to think that she may be the Holy Ghost. That is, although she thinks she might just be Felicité, an ordinary human being, she thinks that it is possible that she is actually the Holy Ghost, currently seeing the world from Felicité's perspective. She thinks that Felicité has a simple location but the Holy Ghost does not. Perhaps he is everywhere or perhaps he is outside time and space altogether. If he is everywhere, she thinks that it is possible that he is the only being that is everywhere (assuming that he is not different from the Father and the Son, something that Felicité finds it hard to get clear on); but being convinced of the reality of the beings of the Dark Side, she is far from confident of that. How then do we characterize her belief using centred worlds? The possibility that she is Felicité is easily handled: that is just the class of centred worlds whose centre is wherever she takes Felicité (and only Felicité) to be (note that on her view, only Felicité is at *just* that one place). But the possibility that she is the

¹⁷ For Lewis's description of this as his aim, see the Postscript C to 'Attitudes' pp. 158–9. Of course we might still want something that is more psychologically realistic, which would give us reason for avoiding the centred world approach.

Holy Ghost cannot be captured using a spatio-temporal centre, for she does not believe that there is any centre that uniquely locates the Holy Ghost.

In response to worries of this kind, Lewis suggests that we should take centres to be, not spatio-temporal locations, but, rather, individuals. Will that always do the job? Shen-Yi Liao has pointed out that, with beliefs involving time travel, it would not.¹⁸ He looks at various other possibilities for the role of centre that might do better. However, it seems to me that to try for something that is guaranteed to work is to raise the standard impossibly high. It is to seek a *characteristica universalis*: a universal conceptual framework that has the resources to express, or at least to accommodate, any possible thought. It is far from clear that we will be able to do that, or that we need to. All that we need is to provide for each centred world, a centre that will serve as a breakthrough point, in the following sense: if the thinker self-ascribed the property given by that center (whether location or some other property), then that would enable the thinker to self-ascribe the other properties that are true of that world.

If we generalize the notion of centre in this way, we might think that whenever a thinker gains *de se* knowledge, that is because they self-ascribe location of some such centre. But even that is not true. It is an impression generated, once again, by an over-restricted focus. To see this consider again one of the examples with which we started.

(v) *Perry comes to realize that he is spilling the sugar*

Assume, once more, that Perry knows exactly where and who he is. What then does he come to know when he realizes that he is spilling the sugar? It is not that he distinguishes between centres, accepting some and ruling others out. Rather, on the natural understanding at least, he distinguishes between worlds. That is, amongst the class of centred worlds with the same centre, he comes to reject those in which the person located at that centre—i.e. he, Perry—lacks the property of spilling the sugar.

So, in working with the centred world approach, we are left with a very complicated picture. When we try to identify what is distinctive about *de se* knowledge we cannot in general do so by pointing to the role of the centres; indeed, it is far from clear quite how we would mark out the class of the *de se*. In contrast, when working with Lewis's original formulation, the *de se* is clearly marked out: a *de se* belief is one that involves the ascription of a property that is not membership of a world. My suspicion is that the enthusiasm for the centred world formulation stems from a preoccupation with truly self locating beliefs: the kind that feature in the Two Gods case. But as we have seen, not all *de se* beliefs are like that. We would do much better to stick with Lewis's original formulation.

ATTITUDES DE RE

Traditionally philosophers have recognized at least three classes of belief. There is the pure *de dicto*, which we can think of as composed of knowledge—or, at least, belief—by description. There is the *de se*. And then there is the *de re*: belief that involves direct acquaintance with an

¹⁸ Liao, 'What are Centred Worlds?'

object, but which need not be *de se*. When Kaplan sees, unknowingly, that his pants are on fire, he has direct (or at least, merely reflection-mediated) *de re* perceptual knowledge of the flaming pants; but he lacks the *de se* knowledge that they are his.

Lewis's approach is designed to capture the *de se*. The *de dicto* come as the special case in which I self-ascribe membership of the relevant worlds. But what of the *de re*? A *de re* belief is a belief about some particular thing, for instance something that I can identify by seeing it or touching it. In that sense it goes beyond the *de dicto*, and so cannot be reduced to the *de se* by the device of merely having a world in common. But on the other hand, many *de re* beliefs are not directly *de se* beliefs. If I say, by looking at you, that you are tall, it looks as though I have made an ascription of a property to you. It doesn't look, on the face of it, as though self-ascription comes into it at all.

Here Lewis's approach is to divide and conquer. *De re* beliefs are not pure mental states he says. Indeed, on Lewis's view, nothing that involves a relation to an external object can be a belief, since beliefs, like all mental states, are in the head. Instead *de re* beliefs are an amalgam of a mental state and a relation to an object. The mental state is *de se*.

Let us here grant to Lewis this composite characterization of the *de re*; whilst much could be said against it, and against the idea that belief must be in the head, my focus will be on the mental state that Lewis thinks makes up part of the *de re*. How can this be understood as *de se*?

Lewis's idea is that to believe *de re* of an object Y that it has a property X is 'to ascribe X to Y under some suitable description of Y'. Most of Lewis's discussion is concerned with elucidating what a 'suitable description' is (it is either something that captures the essence of Y, or is a relation of acquaintance to Y). But prior to this comes Lewis's account of what it is for a subject to ascribe a property X to an individual Y under description Z. This obtains iff:

- (1) the subject bears the relation Z uniquely to Y; and
- (2) the subject self-ascribes the property of bearing relation Z uniquely to something which has property X.

This makes clear the composite nature of the *de re*. The first condition is not mental; the second is: it is a *de se* mental ascription. To make this more concrete: whilst the character in Kaplan's example does not self-ascribe the property of wearing the burning pants, he does self-ascribe the property of seeing them.

If this account is to work, every case of *de re* belief must be a case in which the subject self-ascribes a particular unique relation to the thing in question, that is, to the *res*. But why should we think that that is so? Couldn't a subject have a belief about the *res* without having any belief about the relation they bear to it? Or, equally worrying for Lewis's account, does the subject have to believe that the relation is *unique*: couldn't a subject have, and take themselves to have, two different *de re* beliefs whilst self-ascribing no difference in the relation that they bear to the two different things concerned? At the very least, to deny this is a substantial assumption. But I shall not pursue that worry here. Let me instead turn to examine some of the other potential costs of Lewis's approach, those that stem from the idea of primitive self-ascription.

SOME COSTS OF PRIMITIVE SELF-ASCRPTION

I doubt that there is anything like a knockdown argument against the very idea of primitive self-ascription. But it is certainly odd; at least as odd as anything that Frege proposed. Indeed, like Frege's account it makes the *de se* essentially incommunicable.¹⁹ My own take on it tends to run back, in a way that is hard to resist, to the idea that I am simply ascribing properties to one particular thing in the world, namely myself. That is why it makes sense to ascribe to myself the same sorts of properties that it makes sense to ascribe to other people; and why—a point we shall come to shortly—it makes sense to make first person plural ascriptions, where the same property is simultaneously ascribed to myself and to others.

In contrast Lewis has to think that self-ascription is not only different, but fundamental. Since everything is at base a self-ascription, that is where we must start. This takes him to a position that is almost Cartesian in flavour: we start with self-ascription, and then build our way out. Indeed, self-ascription may be the best way to understand Descartes' position. It is sometimes complained that Descartes is wrong to start the *cogito* with the premise 'I think'; that suggests an independent take on the self as a self-standing object, which mere thought cannot give. All he is entitled to, the complaint continues, is the weaker premise 'There is thinking'. But that seems too weak. Suppose the subject both thought and worried: then it should follow not just that there would be both thinking and worrying separately going on, but that they would be somehow co-instantiated.²⁰ But how can we make sense of that if we are not allowed to posit a self-standing subject? Here the idea of self-ascription might help us. Both thinking and worrying are self-ascribed, and so their conjunction can be self-ascribed, without the need to posit a subject.

Whilst such a starting point will be unattractive to those who reject Descartes' approach²¹, it does seem that we can get our minds around it. But when we look to more complex examples, that is less clear. To see this, let us start with the first person plural. If I say, of you and me:

We are hungry

it seems that I have ascribed the same property—hunger—to each of us, and that the ascriptions are of the same form. Equally, if I can hear that one of our stomachs is rumbling but know not which, I can wonder whether to ascribe the property to you or to me.

¹⁹ See Stalnaker, *Our Knowledge*, for some discussion of this. Note though that I am not suggesting that we should abandon Lewis's account in favour of Frege's. Each of them makes sense only in the context of a very different account of the nature of the content of thought.

²⁰ This co-instantiation concern is raised by Bernard Williams in *Descartes* (Harmondsworth: Penguin, 1978) pp. 95–100. He credits Georg Lichtenberg with the original worry.

²¹ The classic response here is Quine 'Epistemology Naturalized' in *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969).

On Lewis's account, none of this is straightforward. When I say that we are hungry, it might intuitively seem that I make two ascriptions, one to me and one to you. Translated into Lewis's account, these become two very different self-ascriptions. One of them is of the property of being hungry. The other (assuming this is a *de re* ascription) is of the property of inhabiting a world in which the person I'm acquainted with in a certain way is hungry. And likewise, when I wonder whose stomach it is that is rumbling, I wonder, not to whom to ascribe the property of having a rumbling stomach, but which of two very different properties to self-ascribe. As I say, to point this out is not to provide anything like a knockdown argument; and those sympathetic to the balance book method of doing philosophy might well judge that the economy of provided by having just one sort of belief overcomes the inelegance of apparently similar contents coming out differently. But I am wont to use Lewis's response to Quine against himself: 'I protest that the advantages of uniform objects are not to be lightly forsaken.'

Is there then another way that we might understand first person plural ascriptions? An obvious approach is to think that the 'we' picks out a group, in which the speaker is included. Then rather than a conjunction of self-ascriptions, we could have one: self-ascription of membership of a group with a certain property. We can see how that might work for truly collective cases in which the group really does possess a property that is not possessed by the members individually, such as:

We carried the wardrobe upstairs

It could also perhaps handle truly distributive cases, like our original 'We are hungry'. Here the idea is that one self-ascribes membership of a group each of whom has that property. What is less clear is how one would accommodate the so-called cumulative readings, like:

We ate seven pizzas between us.

This isn't collective; unlike wardrobe carrying, pizza eating, is typically something that is done individually, even if in the company of others. Equally though it is not distributive; it is not true that each member of the group ate seven pizzas.²² Such cases are rather hard to accommodate on any account though, so they should not count too heavily against Lewis.

The point at which the plural pronouns become really problematic for Lewis is when we combine them with another kind of difficulty: cases in which the first person pronoun is not in subject position. In general, to understand sentences with the first as self-ascriptions will require radically reconceptualizing the sentence in question. Sometimes this is not too complicated. A sentence like

Hilary chose me

will involve self-ascribing the property of being chosen by Hilary, and

²² See Fred Landman, 'Plurality' in S. Lappin (ed), *Handbook of Contemporary Semantics*, (Oxford: Blackwell, 1995) pp. 425-457. Thanks to Friederike Moltmann for discussion here.

Hilary chose Rose to assess me

will involve self-ascribing the property of being the person whom Hilary chose Rose to assess. But what of

I chose myself

which seems to involve self-ascribing the property of self-choosing? Or, worse still, what of

I chose myself to assess myself

a doubtless corrupt but nonetheless comprehensible action, which seems to involve self-ascribing the property of self-self-assessing-choosing?

When we combine these cases with singular and plural pronouns things become even more intractable:

I chose us to assess me

which I cannot manage to parse in the form of a self-ascription. If possessive pronouns also involve self-ascription, the same difficulties will arise in much more mundane cases:

I cleaned myself with my handkerchief

I defended our house with my friends

Perhaps there is a way to parse these, but the underlying worry should be clear. To understand these sentences in terms of self-ascription requires not simply rearranging the ideas, but coming up with novel properties, properties of self-choosing and so on. If self-ascription really is primitive, then it seems that it is with such properties that we start. That is not an incoherent idea; but it is very hard to believe.

CONCLUSION

I think that Lewis's account of the *de se* is both more radical, and more unintuitive, than has been appreciated. Lewis, of course, was always prepared to accept the unintuitive if it made for an elegant theory. What is less clear is how elegant the theory can be made to be.